






Comparative bioacoustics: a roadmap for quantifying and comparing animal sounds across diverse taxa

Karan J. Odom^{1,2*} , Marcelo Araya-Salas^{1,2,8}, Janelle L. Morano^{3,9} , Russell A. Ligon^{1,2}, Gavin M. Leighton^{1,2,10}, Conor C. Taff^{1,4}, Anastasia H. Dalziell^{1,2,5} , Alexis C. Billings^{6,11}, Ryan R. Germain^{1,2,12}, Michael Pardo^{1,2,13}, Luciana Guimarães de Andrade^{4,7}, Daniela Hedwig⁷, Sara C. Keen^{2,7,14}, Yu Shiu⁷, Russell A. Charif⁷ , Michael S. Webster^{2,3} and Aaron N. Rice⁷ 

¹*Cornell Lab of Ornithology, Cornell University, Ithaca, NY, 14850, U.S.A.*

²*Department of Neurobiology and Behavior, Cornell University, Ithaca, NY, 14853, U.S.A.*

³*Macaulay Library, Cornell Lab of Ornithology, Cornell University, Ithaca, NY, 14850, U.S.A.*

⁴*Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, 14853, U.S.A.*

⁵*Centre for Sustainable Ecosystem Solutions, University of Wollongong, Northfields Ave, Wollongong, NSW, 2522, Australia*

⁶*Division of Biological Sciences, University of Montana, Missoula, MT, 59812, U.S.A.*

⁷*Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, Ithaca, NY, 14850, U.S.A.*

⁸*Sede del Sur, Universidad de Costa Rica, Golfito, 60701, Costa Rica*

⁹*Department of Natural Resources and the Environment, Cornell University, Ithaca, NY, 14853, U.S.A.*

¹⁰*Department of Biology, SUNY Buffalo State, Buffalo, NY, 14222, U.S.A.*

¹¹*Department of Environmental, Science, Policy and Management, University of California, Berkeley, Berkeley, CA, 94709, U.S.A.*

¹²*Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, DK-2100, Denmark*

¹³*Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO, 80523, U.S.A.*

¹⁴*Department of Geological Sciences, Stanford University, Stanford, CA, 94305, U.S.A.*

ABSTRACT

Animals produce a wide array of sounds with highly variable acoustic structures. It is possible to understand the causes and consequences of this variation across taxa with phylogenetic comparative analyses. Acoustic and evolutionary analyses are rapidly increasing in sophistication such that choosing appropriate acoustic and evolutionary approaches is increasingly difficult. However, the correct choice of analysis can have profound effects on output and evolutionary inferences. Here, we identify and address some of the challenges for this growing field by providing a roadmap for quantifying and comparing sound in a phylogenetic context for researchers with a broad range of scientific backgrounds. Sound, as a continuous, multidimensional trait can be particularly challenging to measure because it can be hard to identify variables that can be compared across taxa and it is also no small feat to process and analyse the resulting high-dimensional acoustic data using approaches that are appropriate for subsequent evolutionary analysis. Additionally, terminological inconsistencies and the role of learning in the development of acoustic traits need to be considered. Phylogenetic comparative analyses also have their own sets of caveats to consider. We provide a set of recommendations for delimiting acoustic signals into discrete, comparable acoustic units. We also present a three-stage workflow for extracting relevant acoustic data, including options for multivariate analyses and dimensionality reduction that is compatible with phylogenetic comparative analysis. We then summarize available phylogenetic comparative approaches and how they have been used in comparative bioacoustics, and address the limitations of comparative analyses with behavioural data. Lastly, we recommend how to apply these methods to acoustic data across a range of study systems. In this way, we provide an integrated

* Author for correspondence at address 1 (E-mail: kjo43@cornell.edu)

framework to aid in quantitative analysis of cross-taxa variation in animal sounds for comparative phylogenetic analysis. In addition, we advocate the standardization of acoustic terminology across disciplines and taxa, adoption of automated methods for acoustic feature extraction, and establishment of strong data archival practices for acoustic recordings and data analyses. Combining such practices with our proposed workflow will greatly advance the reproducibility, biological interpretation, and longevity of comparative bioacoustic studies.

Key words: acoustic communication, phylogenetic comparative analysis, signal evolution, acoustic measurements, multivariate analysis

CONTENTS

I. Introduction: the need for comparative bioacoustics	2
II. Approaches for quantifying animal sounds	5
(1) Choosing which acoustic units to compare	5
(2) Choosing which acoustic metrics to collect	7
(a) <i>Signal analysis</i>	9
(b) <i>Derived metric analysis</i>	10
(c) <i>Multivariate analysis</i>	11
(3) Tools and software	12
(4) Caveats for measuring sound	13
(5) Best practices for measuring sounds	13
III. Comparative phylogenetic analyses with sound	14
(1) Types of phylogenetic analyses	14
(a) <i>Quantifying phylogenetic signal</i>	14
(b) <i>Evolutionary rate and diversification</i>	15
(c) <i>Ancestral state reconstruction</i>	15
(d) <i>Phylogenetic correlative analyses</i>	16
(2) Caveats for phylogenetic comparative analyses	17
(a) <i>The choice of phylogeny</i>	17
(b) <i>Assessing the underlying models of trait evolution</i>	17
(c) <i>Sample size and taxon sampling</i>	18
(d) <i>Trait evolution and behaviour</i>	18
(e) <i>Correlation is not causality</i>	18
(f) <i>Vocal learning and cultural evolution</i>	18
(3) Best practices for phylogenetic comparative analyses with sounds	19
IV. Conclusions	19
V. Acknowledgements and author contributions	20
VI. References	20
VII. Supporting information	25

I. INTRODUCTION: THE NEED FOR COMPARATIVE BIOACOUSTICS

Animals exhibit a bewildering diversity of complex and highly variable sounds used for diverse communicative functions. These sounds range from the long, low-frequency modulated rumbles of forest elephants to the short, rapid burst pulses of dolphins, or the highly variable mimicry and species-specific songs of superb lyrebirds (*Menura novaehollandiae*) (Fig. 1; Dalziell, 2012; Dalziell *et al.*, 2013; de Andrade *et al.*, 2017; Keen *et al.*, 2017; Hedwig, Verahrami & Wrege, 2019; Dalziell *et al.*, in press). Such variation exists among distantly related taxa as well as among close relatives. For example, within both the New World blackbirds (family: Icteridae) and birds-of-paradise (family: Paradisaeidae), songs vary from pure-tone whistles to intricate combinations

of broadband notes, clicks, and buzzes (Price & Lanyon, 2002; Ligon *et al.*, 2018). Such variation can reflect the underlying morphology and physiology of animal sound production mechanisms, adaptations to different transmission properties of the environment, as well as variation in sexual and other social selective pressures (Wiley, 1982; Devoogd *et al.*, 1993; Andersson, 1994; Podos, 1996; Blumstein & Armitage, 1997; Bradbury & Vehrencamp, 2011; Suthers *et al.*, 2016; Taylor, Charlton & Reby, 2016). Evaluating how these constraints and selective pressures relate to extreme variation of acoustic signals is fundamental to addressing how diversity in this widespread communication modality arose (Rendell *et al.*, 1999; Cardoso & Hu, 2011; Riede & Goller, 2014).

In the digital age, extensive access to acoustic media and complementary morphological, environmental, and life-

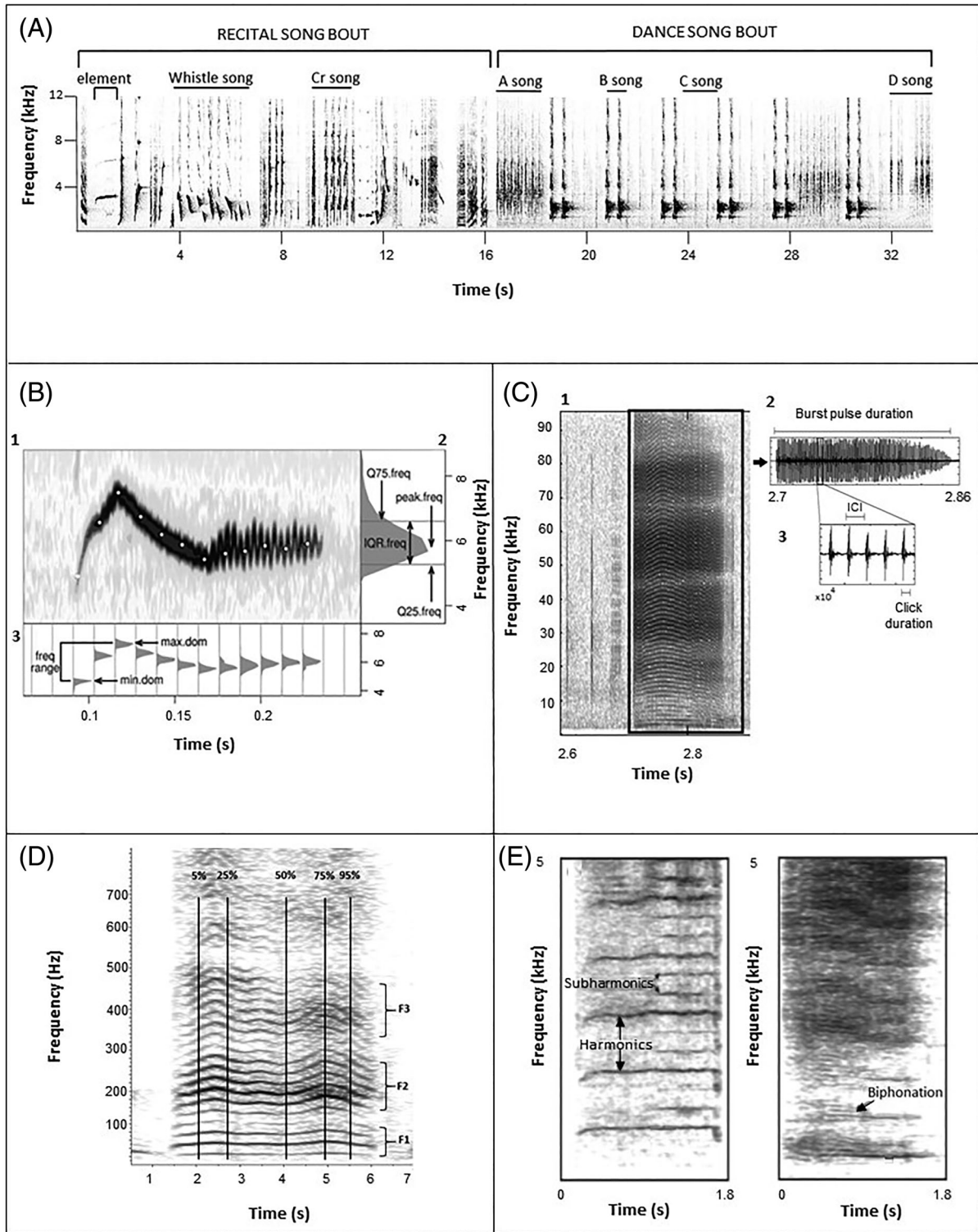


Fig. 1. Examples demonstrating structural diversity in animal signals and possible associated metrics. (A) Superb lyrebird (*Menura novaehollandiae*) song showing hierarchical structure, in which signals are organized into elements, songs (e.g. Whistle song, Cr song), and song bouts (Recital song bout, Dance song bout) with variation at each hierarchical level (modified from Dalziell, 2012; Dalziell *et al.*, 2013). (B) White-chinned sapphire (*Hylocharis cyanus*) song showing (1) robust (energy distribution based) point measures of frequency, including (2) frequency of the 1st and 3rd quartiles (Q25.freq & Q75.freq) and frequency bandwidth of the interquartile range (IQR.freq), and (3) frequency spectrum ‘slices’ calculated for separate time windows across the signal to (Figure legend continues on next page.)

history data sets gives us the data necessary to address questions about signal evolution at previously impossible scales (e.g. Wilman *et al.*, 2014; Dale *et al.*, 2015; Mason *et al.*, 2017b; Miller *et al.*, 2019). Combined with rapid advances in comparative phylogenetic analyses, statistical and bioacoustics software packages, and well-resolved phylogenies constructed from genomic data, we now have an unprecedented toolkit to address such questions with continuously improving precision (Charif, Waack & Strickman, 2010; Jetz *et al.*, 2012; Revell, 2012; Rabosky *et al.*, 2014). Recent large-scale comparative studies of signal evolution have made major advances in our thinking about the underlying evolutionary pressures shaping elaborate traits (Amézquita *et al.*, 2009; Farris & Ryan, 2011; Odom *et al.*, 2014; Riede & Goller, 2014; Tobias *et al.*, 2014; Dale *et al.*, 2015; Rabosky *et al.*, 2018). However, studies that evaluate sound as a continuously varying character in a phylogenetic framework have only started to become common recently (e.g. Goutte *et al.*, 2016; Mason *et al.*, 2017a; Billings, 2018; García & Tubaro, 2018; Ligon *et al.*, 2018; Charlton, Owen & Swaisgood, 2019). Yet such studies are important to elucidate nuanced evolutionary patterns shaping animal sounds. However, because bioacoustics and comparative biology are both specialized fields, researchers with the skills to quantify sound do not necessarily have the background for phylogenetic analyses and *vice versa*. With the current bioacoustics data and phylogenetic methods available, we are at a crossroads where assembling and evaluating the tools available for comparative analysis of animal sounds could enable a broader range of researchers to quantify and compare acoustic signals in a phylogenetic framework.

There are a number of challenges specific to phylogenetic analyses of animal acoustic signals, including identifying appropriate metrics for comparing diverse acoustic structures, inconsistent terminology across studies and taxa, and analytical challenges associated with the resulting high-dimensional data (Busnel, 1963; Hopp, Owren & Evans, 1998; Bradbury & Vehrencamp, 2011; Suthers *et al.*, 2016; Adams & Collyer, 2018). Sound, as a highly variable multidimensional trait can be difficult to quantify, especially in concise ways that capture overall acoustic structure and facilitate comparison across taxa. Nevertheless, for comparative acoustic analyses it is essential to identify and measure homologous or comparable units of sound so that various acoustic structures can be detected and processed consistently across taxa (Lauder, 1986; ten Cate, Lachlan & Zuidema, 2013; Russo, Ancillotto & Jones, 2018). However,

animal sounds can range from subtle structural variation to a seeming complete lack of similar acoustic features between species, not to mention variation in syntax or element spacing, which makes direct comparison of signals difficult (e.g. Goicoechea, De La Riva & Padiá, 2010; Dunn *et al.*, 2011; Matthews *et al.*, 2012; Katahira *et al.*, 2013; Ligon *et al.*, 2018). In addition, the terminology used to investigate animal sounds varies, including the use of multiple terms for the same metric or acoustic unit, single terms with multiple definitions, or non-mutually exclusive terms (Marler, 1961, 1967; Broughton, 1963; Thompson, LeDoux & Moody, 1994; Deecke & Janik, 2005; Cholewiak, Sousa-Lima & Cerchio, 2013; Köhler *et al.*, 2017). Thus, researchers risk terminological entanglements when trying to compare sounds across species, and especially across more diverse taxa. Lastly, the properties of sound itself present challenges for comparative bioacoustic analyses. Sound is a multidimensional signal with continuous variation, making it difficult to sample all aspects of an acoustic signal accurately (Hopp *et al.*, 1998; Deecke & Janik, 2005; Köhler *et al.*, 2017). Handling the resulting highly dimensional data in ways that are compatible with existing phylogenetic comparative analyses is a further challenge. The set of metrics collected and how they are prepared for analysis can also vary depending on the kind of phylogenetic comparative analysis conducted (Uyeda, Caetano & Pennell, 2015; Adams & Collyer, 2018). Therefore, careful consideration is needed at each of these stages of acoustic analysis to ensure that the resulting acoustic measurements accurately reflect the original animal signals and are comparable in a phylogenetic context.

Our goal is to collate the main bioacoustics approaches and comparative phylogenetic analyses useful for quantifying and comparing animal sounds in a phylogenetic context. This review is divided into two main sections: (i) approaches for quantifying animal sounds and (ii) approaches for phylogenetic comparative analysis of animal sounds. The section on quantifying animal sounds includes a synopsis of common terms and metrics used to compare animal sounds, discussion of techniques for handling and consolidating multiple acoustic variables, including data-reduction techniques, and pitfalls and best practices of acoustic analyses with diverse animal sounds. The section on phylogenetic comparative analysis summarizes types of phylogenetic comparative analyses, including examples of comparative acoustic studies that have used each type. The current limitations of phylogenetic comparative analyses are also discussed. At the end of

(Figure legend continued from previous page.)

illustrate frequency contour measurements based on the dominant (dom) frequency. Figure modified from Araya-Salas *et al.* (2017). (C) Spectrogram and waveform of a Guiana dolphin (*Sotalia guianensis*) burst pulse, emphasizing the importance of time window, including (1) the spectrogram of a complete burst pulse, (2) the waveform of the same burst pulse with the entire sequence of clicks for that duration, and (3) a zoomed-in waveform showing five clicks, the inter-click interval (ICI), and click duration (modified from de Andrade *et al.*, 2017). (D) Female forest elephant (*Loxodonta cyclotis*) rumble showing 5 to 95% percentiles of the temporal energy distribution and formants (F1–F3) (provided by the Elephant Listening Project, Cornell Lab of Ornithology). (E) Female koala (*Phascolarctos cinereus*) rejection call demonstrating two types of non-linear phenomena, modified from Charlton *et al.* (2017).

both sections, we recommend and outline best practices for acoustic and comparative phylogenetic analyses. We draw most heavily from literature pertaining to vertebrate vocalizations, especially birds, for which a bulk of literature exists; however, we expect the acoustic and comparative analyses described here to be broadly applicable to all animals. Our intention is to provide a roadmap for phylogenetic comparative analysis of animal sounds aimed at improving the accuracy and ease with which highly variable animal sounds can be quantified and compared across species.

We note that the term ‘signal’ has variable meanings within evolutionary biology and animal communication [e.g. signal evolution (Endler, 1992), phylogenetic signal (Blomberg, Garland & Ives, 2003), behavioural signals (Bradbury & Vehrencamp, 2011)], and a separate meaning in acoustics, with origins in signal processing (Morfeý, 2001). In the remainder of this review, we primarily restrict the use of the term ‘signal’ to meanings consistent with signal processing, as one major focus is on extraction and analysis of acoustic parameters. Unless specified, we remain agnostic about the evolutionary or animal communication interpretations of the sounds we discuss, although we acknowledge that species-specific experimental studies are often needed to verify whether animal sounds are signals in the animal communication sense of the term (e.g. Bradbury & Vehrencamp, 2011).

II. APPROACHES FOR QUANTIFYING ANIMAL SOUNDS

Because animal sounds vary in multiple dimensions, it is important to choose acoustic traits *a priori* that exist in the majority of species being examined, but that are variable and biologically relevant. We think of this as a two-step process. First, decide how consistently to delimit (separate) the acoustic units that will be compared. Second, choose metrics that accurately capture the variation that exists in the signals of interest. Both steps require becoming familiar with the acoustic variation in the model system and keeping the research question in mind when choosing appropriate acoustic units and metrics.

(1) Choosing which acoustic units to compare

At the most basic level, comparing acoustic structure across taxa requires identifying and defining the type of acoustic signals to be compared. This requires some familiarity with the organisms being studied, because most animals produce multiple sounds that are structurally and functionally distinct (Gerhardt & Huber 2002; Marler & Slabbekoorn, 2004; Catchpole & Slater, 2008; Cardoso, 2012; Cholewiak *et al.*, 2013; Russo & Voigt, 2016; Smotherman *et al.*, 2016). The broadest acoustic unit that will be compared is usually dictated by the research question (e.g. Do bat echolocation calls vary with habitat type? Which features of frog advertisement calls can be used to distinguish among species? Does the

syntax of wolf howls vary with context?). The first step in an acoustic analysis is deciding whether the research question, and thus the comparative analysis, is concerned with the entire acoustic repertoire of a given taxon, or a specific subset of those signals. For example, while megadermatid bats use echolocation calls to hunt prey, they also produce a large complex repertoire of mostly lower-frequency calls used during intraspecific interactions (Leippert, 1994; Hanrahan 2020; Hanrahan *et al.*, in press). Therefore, researchers interested in how echolocation calls vary with habitat type in this family will need to be able to distinguish between echolocation calls and non-echolocation acoustic social signals. Decisions about which acoustic units to compare should also take into account the function of the vocalizations, when known, because different selective pressures likely act on functionally different vocalizations (Greig, Price & Pruett-Jones, 2013; Greig & Webster, 2014). At the same time, if the ultimate comparative question is one of function or selective pressures, it is important to avoid circularity in the criteria used to select samples and the hypotheses being tested (i.e. evolutionary response to ambient noise measured on parameters affected by background noise). A broad enough range of vocalizations to test your hypotheses should be included. For example, if investigating whether a vocalization is sexually or naturally selected, analysing vocalizations exclusively used in mate choice might not provide much insight.

Once the research question and signals of interest have been identified, the next step is to partition the acoustic signals into component structures that can be analysed and compared across the species of interest (ten Cate *et al.*, 2013). Animal sounds are frequently hierarchically organized (e.g. elements structured into syllables, structured into multi-syllabic vocalizations, which in turn can be structured into vocal bouts). Within certain taxa, careful attention has been paid to develop terminology to describe acoustic signals and their component structures (e.g. Thompson *et al.*, 1994; Cholewiak *et al.*, 2013; Köhler *et al.*, 2017). However, even within taxa, acoustic terminology can be inconsistent (e.g. Thompson *et al.*, 1994; Catchpole & Slater, 2008; Bonnevie & Craig, 2018) and across taxa there is little consensus on terminology (but see Busnel, 1963). It is beyond the scope of this review to provide a comprehensive list of terminology or reconcile past inconsistencies. Luckily, more recent treatments have provided guidelines for delimiting acoustic units regardless of taxa or terminology (Kershenbaum *et al.*, 2016). Table 1 provides a glossary of broad terms that we have identified as useful for categorizing diverse animal sounds into acoustic units and hierarchical levels, some of which are illustrated in Fig. 1A.

Which acoustical units are compared and how to delimit them must be based on the research question. In some cases, measuring a broad acoustic unit is designated by the research question. For example, for research questions about total duration or frequency modulation over an entire multi-syllabic bird song, the entire song is the appropriate acoustic unit to measure (Podos, 1997; Dalziell & Cockburn, 2008). In other instances, researchers may want to measure multiple

Table 1. Some common terms used to describe animal acoustic signals. Note that there is little standardized agreement of terms used to describe animal sounds within and across taxa, so our terms and definitions may not be representative for all taxa. In addition, researchers specializing in certain taxa may use these terms differently. The definitions provided here are based on structural variation, but context is also likely important in differentiating among some of these vocalization types

Acoustic unit	Definition	Synonyms	References
Element	Smallest unit of sound; a discrete, continuous sound with defined start and end; discrete sound separated from other elements by silence or abrupt changes in the spectral energy distribution; related to the sound mechanism: muscle activation and firing	Note, call, syllable, pulse, acoustic unit, unit	Gerhardt (1998); Catchpole & Slater (2008); Cholewiak <i>et al.</i> (2013); Hedwig <i>et al.</i> (2014); Smotherman <i>et al.</i> (2016); Köhler <i>et al.</i> (2017)
Syllable	Sequence of one or more elements repeated rendered together in the same pattern sequence	Element group, note group	Catchpole & Slater (2008); Weir & Wheatcroft (2011); Lachlan <i>et al.</i> (2013)
Call	Single element or repeated sequence of one or more elements or syllables	Note, syllable	Cardoso (2012); Cholewiak <i>et al.</i> (2013); Marler & Slabbekoorn (2004)
Song	Stereotyped or otherwise distinctive pattern of elements or phrases. Songs are typically considered to have elaborate structure (multiple elements and/or element types)	Call, strophe, theme, motif, verse	Marler & Slabbekoorn (2004); Catchpole & Slater (2008); Cholewiak <i>et al.</i> (2013); Smotherman <i>et al.</i> (2016)
Vocal bout	A performance of songs/calls rendered discrete either by time (bouts separated by pauses) or distinctive acoustic features (i.e. ‘types of bouts’ or ‘singing modes’)	Song session, dawn recital, song sequence, singing mode	Catchpole & Slater (2008); Dalziell & Cockburn (2008); Cholewiak <i>et al.</i> (2013)
Repertoire	Set of acoustically distinct elements, syllables, calls, vocalizations, or songs produced by an individual, group or species	Song type, song repertoire, element repertoire, vocal repertoire (note that these are not direct synonyms, but rather refer to the class of vocalizations that may be grouped into a repertoire)	Catchpole & Slater (2008); Kershenbaum <i>et al.</i> (2016); Harris <i>et al.</i> (2016); Luttrell <i>et al.</i> (2016)

hierarchical levels of acoustic structure. For example, for a project investigating whether song syntax differs across whale populations, researchers may want to gather data at the element level (or the smallest acoustic unit) while keeping track of higher levels of organization, as that will offer flexibility to match when the same elements are or are not grouped together in a higher level.

Our general recommendation for delimiting animal sounds is to separate multi-component or hierarchically structured sounds into obvious, discrete acoustic units that are shared across the taxa of interest. There are multiple approaches for delimiting acoustic signals, including differentiating among acoustic units based on (i) silent intervals (breaks) in the acoustic signal, (ii) changes in the acoustic properties of the signal (e.g. a transition from a pure-tone to a broadband signal), or (iii) series of similar sounds that appear to be grouped together, such as grouping trilled notes, pulse trains, or syllables (see Fig. 2 in Kershenbaum *et al.*, 2016). Whichever approach is used, we recommend delimiting the component signals into the smallest possible unit (i.e. elements; Table 1), particularly for broad comparative

studies of fine acoustic structure when hierarchical patterning or perceptual capabilities of the included taxa are not known *a priori*, as this is often the least subjective approach. Also, in most cases, additional higher-level hierarchical units can be re-created from element-level measurements simply through concatenation (e.g. measurements for a call or song can be calculated from the component elements or syllables). In addition, packages and algorithms can often be used to quantify syntax or sequential patterning of elements or vocalizations within animal signals (Kershenbaum *et al.*, 2016) and automated methods for classifying the overall structure of acoustic signals removes observer bias or subjectivity associated with manual classification of acoustic signals (Deecke, Ford & Spong, 1999; Keen *et al.*, 2014; Wadewitz *et al.*, 2015; see Section II.2c). For an extensive review of approaches for syntax analysis with animal sounds, see Kershenbaum *et al.* (2016).

Nevertheless, we recommend that researchers also pay attention to and possibly delimit higher levels of acoustic structure (e.g. syllables, phrases, calls, songs) when commonly observed in their study system or designated by the research

question. Higher acoustic organization is often biologically meaningful and may be useful to quantify. For example, many songbirds perform trills (rapidly repeated series of elements) or combine elements into stereotyped syllables. Researchers that measured all silent intervals within serin (*Serinus serinus*) songs found a bimodal distribution representing the short intervals within repeated, stereotyped syllables and the longer intervals between non-syllable elements (Mota & Cardoso, 2001). They used this distribution to establish a threshold to distinguish among syllables *versus* less-stereotyped, independent elements within songs and they successfully applied this approach in a comparative framework (Mota & Cardoso, 2001; Cardoso & Mota, 2007). Also, we note that other disciplines recommend neurological, production, perception, or function-based approaches for delimiting acoustic structure (Spector, 1994; Gerhardt & Huber, 2002; Blumstein, 2010; Owren, Rendall & Ryan, 2010; Peshek & Blumstein, 2011; Bonnevie & Craig, 2018).

(2) Choosing which acoustic metrics to collect

Animal sounds exhibit a range of frequencies that vary through time. Even the smallest unit of animal sound (i.e. an element) can vary in multiple dimensions, ranging from a long pure tone that can be easily represented by frequency and time measurements to a short broadband harsh or ‘buzzy’ sound that may be best distinguished by its harmonic structure, and much more (Price, Earnshaw & Webster, 2006; Tyson, Nowacek & Miller, 2007; Charlton, 2015). Therefore, even within a set of clearly defined acoustic units, choosing a set of acoustic measurements that capture the range of ways the animal sounds vary while being well suited for the research question at hand is not always straightforward (e.g. Mason & Burns, 2015; Billings, 2018).

Here we provide a workflow for extracting and quantifying animal sounds across species in preparation for phylogenetic comparative analysis (Fig. 2). Quantitative assessments of animal sounds are typically based on measurements from sound spectrograms, which are visual representations of the relative power at different frequencies in a sound over time (Appendix S1). This framework involves three stages or classes of metrics ordered by the amount of processing involved: (i) signal analysis – the extraction of measurements of frequency, amplitude, time, or energy distributions from spectrograms of the acoustic signal; (ii) derived metric analysis – the calculation and extraction of measurements that capture overall structure, structural variation, or syntax of a vocalization or vocal bout, usually derived from measurements taken or acoustic software procedures performed during signal analysis; and (iii) multivariate analysis. This third class includes multivariate procedures used to create comparisons or composites of signal and derived metrics between two or more sounds (e.g. similarity or distance scores created by spectrogram cross-correlation, cluster analysis, or mapping features in multi-dimensional space). Multivariate analysis also includes data reduction and classification of acoustic units into distinct categories as a precursor for phylogenetic

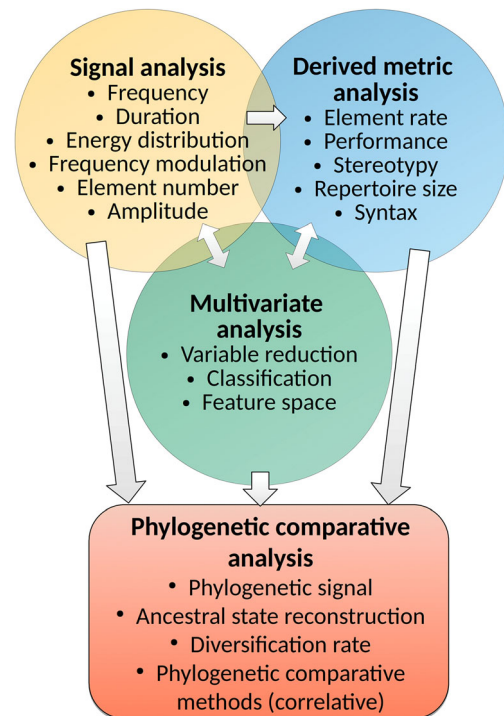


Fig. 2. Workflow for extracting and quantifying animal sounds across species in preparation for phylogenetic comparative analysis. Signal analysis (yellow) is the extraction of measurements directly from the acoustic signal, which can be compiled to compute derived metrics (blue), calculations from signal measurements that capture overall structure, structural variation, or syntax of a vocalization or vocal bout. Signal and derived metrics can subsequently be integrated into multivariate analysis (green) that combines multiple features into multiple dimensions. Both signal and derived metrics can be combined into multivariate analysis in preparation for phylogenetic comparative analyses (orange). Separation among the three stages of acoustic analysis is not always clear-cut (e.g. some derived metrics may be extracted from the signal or multivariate analyses may be used to calculate derived metrics), and is represented by the graded colours between these categories.

or additional acoustic analysis. Derived metrics can usually be differentiated from multivariate metrics in that derived metrics are calculated for specific acoustic units (individual elements, syllables, songs) in a study whereas multivariate metrics are computed as comparisons across pairs of acoustic units or by combining data from all acoustic units.

Depending on the research question, all three types of analyses can be prepared to evaluate a broad range of acoustic variables. Both signal and derived measurements can contribute to multivariate analysis, and then one, two, or all three of these metrics can be used in subsequent comparative analysis (Fig. 2). However, not all comparative analyses of sound require all three steps, nor is inclusion of all possible acoustic variables necessarily appropriate. For example, in some studies it is most appropriate to input the raw signal

variables directly into phylogenetic comparative analyses, either because the raw variables are closely tied to the research question (e.g. effects of habitat on frequency: Billings, 2018) or multivariate analyses may not be appropriate (Mason *et al.*, 2017a). In addition, the separation of acoustic metrics into one class of metric or the other is not always clear-cut. For example, element number is a quantity that can be counted directly on the spectrogram (signal) but may be considered a derived metric because it can also be calculated automatically from the number of selections made during signal analysis. Similarly, principal components created from multivariate analysis may replace signal analysis variables in phylogenetic comparative analyses. Therefore, we picture all three sets of acoustic metrics as somewhat

overlapping, but we think of them as distinct in the process by which they are computed (Fig. 2; Table 2; see online Supporting Information, Table S1).

We provide figures and tables to help navigate this framework for acoustic analysis, including additional explanation and examples of acoustic metrics (Figs 1–3; Tables 2 and S1). Fig. 1 provides example animal sounds that illustrate key acoustic metrics. Fig. 2 illustrates a schematic of the analytical workflow, while Fig. 3 is a decision tree with recommendations for data processing and reduction depending on the data structure and the comparative analyses to be performed. Table 2 provides a list of signal, derived, and multivariate metrics that can be collected and Table S1 provides more extensive explanations for each variable, including additional

Table 2. Commonly used quantitative acoustic variables that can be extracted or calculated from animal acoustic signals for phylogenetic comparative analysis. These quantitative variables may come from direct signal analysis (yellow) or calculations of derived (blue) or multivariate (green) metrics, and may be overlapping. See Table S1 for more extensive definitions, uses, and best practices for these metrics

	Metric (units)	Definition	Examples of use in comparative studies
signal metric	Frequency measurements	Quantitative measurements of frequency made in the spectrogram or power spectrum.	May-Collado <i>et al.</i> (2007); Gingras <i>et al.</i> (2013); Tobias <i>et al.</i> (2014)
	Time measurements	Point in time or difference between points in time that mark the start, end, median or other reference points of a unit of sound.	Gingras <i>et al.</i> (2013); Tobias <i>et al.</i> (2014)
	Magnitude measurements	Measurements that describe amplitude or pressure (in the waveform), or the power, energy, or intensity (in the spectrogram or power spectrum).	Calder (1990); Cardoso (2010)
	Frequency bandwidth	The difference between low and high frequency limits.	Podos (1997); Mason & Burns (2015); Cardoso & Hu (2011)
	Frequency contour measurements	Measures of frequency sampled as regular intervals over the duration of a signal.	Lachlan <i>et al.</i> (2013); Meliza <i>et al.</i> (2013)
Type of metric	Information entropy	Measure of diversity, uncertainty, or randomness of the signal.	Kershenbaum <i>et al.</i> (2016); Blumstein & Chi (2012)
	Harmonicity, tonality or harmonics-to-noise ratio (HNR)	The ratio of the energy from tonal/harmonic components to the energy from noise or non-harmonic components.	Lachlan <i>et al.</i> (2013)
	Element, call, or song-type diversity; repertoire size	Number of types of unique acoustic units in an individual's or species' repertoire.	Read & Weary (1992); Devoogd <i>et al.</i> (1993); Blumstein & Armitage (1997); Darling & Sousa-Lima (2005); Weir & Wheatcroft (2011); Greig <i>et al.</i> (2013); Singh & Price (2015)
Derived	Syntax	The particular order in which acoustic units are produced; or categorizing whether syllables are trilled or are repeating.	Dunn <i>et al.</i> (2011); Katahira <i>et al.</i> (2013)
	Vocal deviation or performance limits	The ability to produce vocalizations close to physiological limits. Usually related to a trade-off between the amount of frequency modulation and the time it takes to produce it.	Podos (1997); Podos <i>et al.</i> (2016); Derryberry <i>et al.</i> (2012); Cardoso & Hu (2011); Weir <i>et al.</i> (2012); Geberzahn & Aubin (2014); Cardoso (2019)
Multivariate	Stereotypy, vocal consistency, or measurements of acoustic variation	Composite measurement of the extent of variation among acoustic signals or pairwise comparisons of all acoustic units in a data set.	Pearse <i>et al.</i> (2018); but see Mikula <i>et al.</i> (2018) for a critical appraisal
	Measures of distance or area in feature space	Mapping of multiple acoustic variables in a multi-dimensional feature space.	Ligon <i>et al.</i> (2018); Tobias <i>et al.</i> (2014)

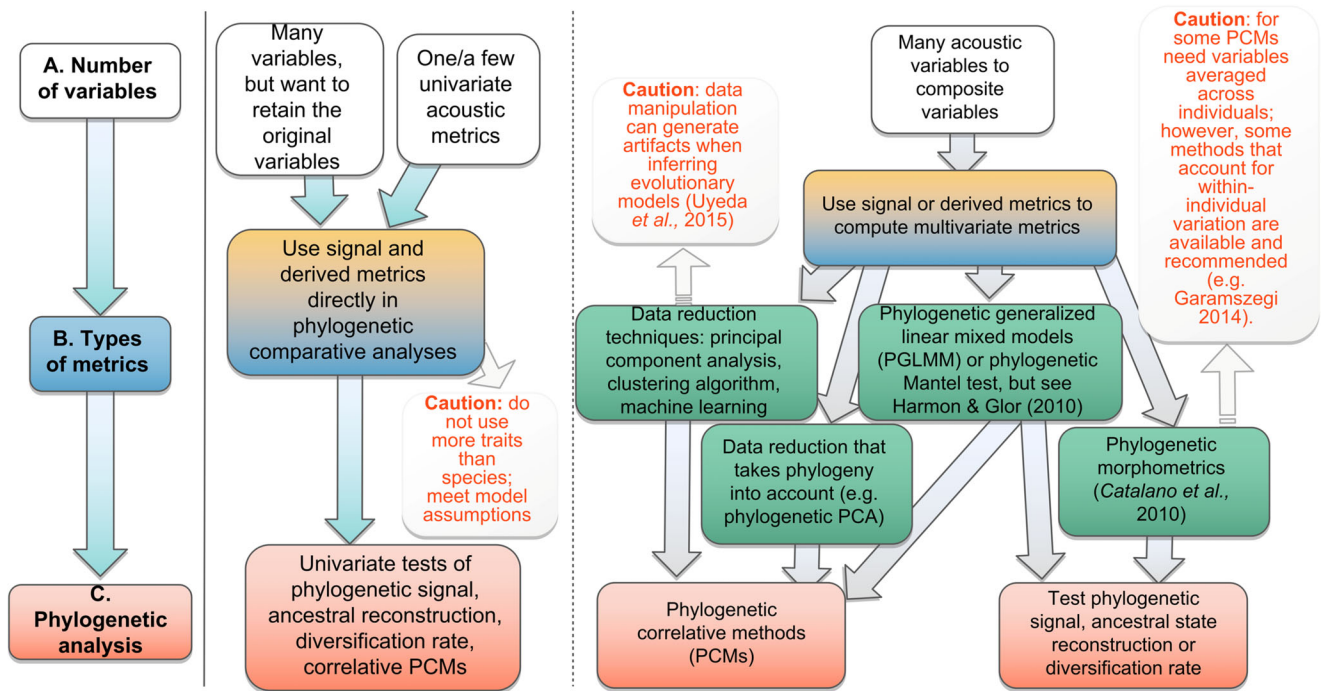


Fig. 3. Decision tree for treatment of acoustic data for phylogenetic analysis, depending on (A) the kinds of acoustic variables in your analysis (signal: yellow, derived: blue, or multivariate metrics: green), (B) the number of acoustic variables in your data set, and (C) the kinds of phylogenetic analysis (orange) to be carried out. We make recommendations on data handling, as well as indicate potential caveats of certain analyses in red.

references and recommendations for best practices. Below we describe each of the three classes of acoustic analysis in more detail.

If the analysis includes many acoustic variables which may contribute to differentiation of different taxa to varying extents, without *a priori* knowledge of which variables are most relevant, it may be more helpful to combine all metrics into multivariate analyses. Obvious caveats to such broad multivariate procedures include that interpreting how specific variables contribute to the analysis, and thus their biological interpretation, can be challenging (although parameters such as variable loading or importance can help with this; Ramasubramanian & Singh, 2016).

(a) Signal analysis

For both single-species studies and comparative analyses evaluating signal structure, the most common approach is to measure multiple acoustic features (signal metrics) per individual or species (Darling & Sousa-Lima, 2005; Rice & Bass, 2009; Cardoso & Atwell, 2011; Greig *et al.*, 2013; Tobias *et al.*, 2014; Mason *et al.*, 2017a). Typical features range from frequency or amplitude measured within each element or vocalization (i.e. measurements of frequency, time, or amplitude; e.g. minimum, maximum, or peak frequency/amplitude) to differences between single (point) measurements or calculations of energy distributions per element or vocalization (e.g. frequency bandwidth, duration, entropy,

or harmonicity; Table S1; Calder, 1990; Podos, 2001; Cardoso, 2010; Blumstein & Chi, 2012; Kershenbaum, 2014; Lachlan, Ratmann & Nowicki, 2018). Most acoustic software programs can extract these metrics from designated acoustic units, although the acoustic units to extract these features from often need to be specified by hand or in a semi-supervised manner. Researchers may also count discrete features, such as numbers of elements, syllables, or vocalizations (to calculate vocalization rates), number of frequency inflections, or instances of a particular feature, such as biphonation or trills (Fig. 1; Tables 2 and S1). For frequency measurements, Cardoso (2013) recommends analysing frequency on a logarithmic scale, as it provides a more accurate representation of animal sound perception and better reflects the relationship between animal body size and resonating frequency. Measuring frequency on a logarithmic scale is particularly important for comparative studies across terrestrial vertebrate species, as frequency analysis on a linear scale could bias results and overestimate differences in frequency (Cardoso, 2013). As with most analytical decisions, implementation will depend on the research question.

To capture more precisely how signals vary over the duration of an element or other acoustic unit, researchers also compare time series of signal measurements (i.e. vectors of acoustic features sampled throughout the signal, such as frequency contours; Fig. 1B; Table S1; Kogan & Margoliash, 1998; Tchernichovski *et al.*, 2000; McCowan, Hanser & Doyle, 2002; Lachlan *et al.*, 2013; Meliza, Keen &

Rubenstein, 2013; Wang *et al.*, 2013). While metrics like frequency contours provide a promising method to measure and reconstruct signal structure more precisely than point measurements, they have limitations. First, frequency contours are a vector of values, rather than a single value. Therefore, they need to be transformed in order to compare the overall similarity of each contour to all other contours [e.g. time series analysis, such as dynamic time warping (DTW); Wang *et al.*, 2013]. DTW is an algorithm that allows flexible comparison of disparate structures by stretching and aligning the signal over time (Kogan & Margoliash, 1998; Lachlan, 2007; Meliza *et al.*, 2013), the product of which is a proximity matrix (a pairwise matrix comparing each signal to all other signals). The resulting proximity matrix then usually needs to be further transformed to extract a set of values that meaningfully and independently represents each contour's shape. This is usually done with multidimensional scaling or other matrix vectorization procedures (see Section II.2c). In addition, the accuracy of frequency contours is vulnerable to signal quality. Noise or reverberation can lead to inaccurate contour tracing, especially at the ends of notes when reverberation masks the actual signal structure. Similarly, tight harmonic stacking coupled with frequency and/or amplitude modulation can lead to inaccurate frequency contour estimation, such that the trace of the contour 'jumps' among harmonics. Solutions include manually editing troublesome note structures (warbleR function 'seltailor'; Araya-Salas & Smith-Vidaurre, 2017), pitch-tracking algorithms, or measuring only the fundamental or dominant frequency (soundgen function 'analyze'; Anikin, 2019). While recent studies have successfully employed time series data (Lachlan *et al.*, 2013, 2018), classifying sounds using machine-learning algorithms on a wide variety of acoustic features can in some cases outperform DTW (Keen *et al.*, 2014). Nevertheless, combining frequency contours with spectrogram cross-correlation (pairwise comparison of amplitude matrices in the bi-dimensional time–frequency space, 'sliding' one spectrogram over the other and calculating a correlation at each step) and feature analysis (analysis of individual metrics) may improve overall classification beyond any one of these approaches on its own (M. Araya-Salas, unpublished data). For these and many other reasons, we generally recommend evaluating the performance of time series data, and consistency and accuracy of all automated acoustic metrics prior to statistical analysis.

Finally, certain irregular animal signals may require special attention or analysis, such as subharmonics or biphonation (Fig. 1E; Wilden *et al.*, 1998; Tokuda *et al.*, 2002; Tyson *et al.*, 2007; Charlton, Watchorn & Whisson, 2017). While rare across species, these naturally produced complex acoustic structures appear to be common in certain taxa [e.g. oscine birds (Zollinger, Riede & Suthers, 2003), cetaceans (Filatova *et al.*, 2007; Tyson *et al.*, 2007)]. One approach to measure such non-linear phenomena is to divide the signal into separate acoustic units where transitions to different acoustic structures occur. More sophisticated approaches are also available (see Tokuda *et al.*, 2002).

In addition, there may be benefits to measuring sound while taking into account perceptual capabilities of the animals being studied. For certain taxa, algorithms and tools have been developed to measure the amplitude of signals within predefined frequency bands that reflect species' auditory ranges (Lyon & Ordubadi, 1982; Fraile & Godino-Llorente, 2014). For example, researchers can apply a series of band-pass filters to a power spectrum to extract a series of spectral bands and the relative bandwidths of the filters can be adjusted to reflect the fact that many animals are better at discerning differences between certain frequencies than others. Typically, the modified spectrum is then mathematically transformed into a cepstrum, and a series of numbers that describe the cepstrum, known as cepstral coefficients, are used as acoustic features with which to compare sounds. The most commonly used type of cepstral coefficient, Mel-frequency cepstral coefficients, are derived from a filter bank that approximates the frequency response of the human auditory system, but have been used successfully to classify the vocalizations of other mammals as well as birds (Picone, 1993; Cowling & Sitte, 2003; Darch, Milner & Shao, 2004; Sandsten, Große Ruse & Jönsson, 2016). Coupling comparative analyses with taxon-specific acoustic perceptual models, especially for taxa that perceive sound in ways that are very different from how we perceive sound, could aid our understanding of signal evolution and ensure that we are comparing salient acoustic features (Clemins & Johnson, 2006; Ren *et al.*, 2009).

(b) Derived metric analysis

Derived metrics are acoustic measurements computed primarily from summary statistics or other calculations from signal measurements. Examples include estimates of repertoire size, element diversity, coefficients of variation, or syntactical patterns. For some derived metrics, the researcher may need to conduct calculations on a species-by-species basis or design a specific analysis of interest (Cardoso & Hu, 2011; Geberzahn & Aubin, 2014; Podos *et al.*, 2016; García & Tubaro, 2018; Cardoso, 2019). In other instances, general sound-analysis packages are available in R that are useful both for feature extraction and computing basic derived metrics of sounds (e.g. warbleR calculates vocalization-level measurements of vocalization duration, element rate, and averaged frequency parameters from the component element selections; Araya-Salas, Smith-Vidaurre & Webster, 2017). Automated methods and R packages also exist for calculating more complicated parameters such as repertoire size estimation and syntax (Kershenbaum, Freeberg & Gammon, 2015; Wadewitz *et al.*, 2015; Harris *et al.*, 2016; Kershenbaum *et al.*, 2016; Luttrell, Gallagher & Lohr, 2016). Recent application of more advanced techniques borrowed from ecology and economics, including rarefaction, offer particularly promising methods for estimating repertoire size (Peshek & Blumstein, 2011; Kershenbaum *et al.*, 2015). Multivariate methods can also be applied to automate the process of classifying acoustic units into categories for syntax analysis or

determining element diversity. For example, in some instances cluster analysis, random forest, and other classification algorithms can efficiently and reliably classify vocalizations into discrete types (Greig *et al.*, 2013; Keen *et al.*, 2014; Sandsten *et al.*, 2016). The steps involved in these approaches are discussed in more detail in Section II.2c. We recommend automated procedures for categorizing animal sounds whenever possible because this streamlines acoustic analysis, reduces observer bias, and improves reproducibility (Salisbury & Kim, 2001; Botero *et al.*, 2008; Wadewitz *et al.*, 2015; Pearse *et al.*, 2018; but see Mikula, Petrusková & Albrecht, 2018). Nevertheless, we recognize this depends on first adequately delimiting sounds into the appropriate acoustic units. While methods like band-limited energy detection exist to detect and delimit breaks in signals, they often require a high signal-to-noise ratio and work better on some signals than others. Thus, separating sounds into acoustic units is a required first step which currently often has to be done by hand for field recordings. Furthermore, in certain circumstances human classification may be the best available method (e.g. if sample sizes are small or vocal units contain higher hierarchical-levels of acoustic organization). In cases in which human classification is used, we recommend calculating a measure of inter-observer reliability (Kaufman & Rosenthal, 2009). We provide a catalogue of some of the most commonly investigated derived metrics in Tables 2 and S1.

(c) Multivariate analysis

Multivariate metrics can be computed from signal and/or derived metrics to produce reduced sets of combined variables or to classify or map acoustic features in multidimensional space. The multivariate metrics we will discuss include creating reduced sets of vectors produced from data-reduction techniques such as principal component analysis (PCA) or multidimensional scaling (MDS), which can then be input directly into subsequent analyses. Useful multivariate approaches for comparative bioacoustics also include creation and analysis of matrices of similarity scores produced by spectrogram cross-correlation, random forest classification, or DTW of frequency contours (e.g. Keen *et al.*, 2014; Table 2). We will also discuss the use of multivariate analyses to create derived metrics from high-dimensional measurements. This can include classification of acoustic units into discrete vocalization types to calculate repertoire size or categorize vocalizations prior to syntax analysis, and also includes data reduction to create fewer signal-measurement vectors (Hedwig *et al.*, 2014; Wadewitz *et al.*, 2015; Mason *et al.*, 2017a). Lastly, multivariate analyses can be used to compute derived metrics in multidimensional space, such as calculations of the area surrounding sets of sounds plotted in feature space – a useful proxy for signal diversity (Nelson & Marler, 1990; Tobias *et al.*, 2014; Ligon *et al.*, 2018). Accordingly, the multivariate methods explained here include both data reduction and classification procedures.

Here we briefly summarize some popular multivariate approaches for data reduction and classification. Common data-reduction procedures include PCA, factor analysis (FA), t-distributed stochastic neighbor embedding (t-SNE), MDS (also called principal coordinates analysis, PCoA), and random forest (RF), whereas common classification procedures also include RF, as well as linear discriminant analysis (LDA), neural networks, and cluster analysis (Johnson & Wichern, 1982; Izenman, 2008; Legendre & Legendre, 2012; Ramasubramanian & Singh, 2016). In data reduction, the goal is to produce vectors representing a reduced set of variables from a larger, potentially correlated set of variables. Classification analyses, on the other hand, are used to classify a set of signals into categories based on sets of acoustic variables. Both approaches can also be used to map acoustic signals in feature space (PCA, FA, t-SNE, MDS, RF, cluster analysis). An additional important distinction among these analyses is whether they are supervised or unsupervised (Ramasubramanian & Singh, 2016). Supervised approaches are those in which the classes of the response variable are known *a priori* and, with the explanatory variables (in this case, features of the signal), are used to predict the class of the signal [e.g. LDA, supervised RF (Armitage & Ober, 2010; Ramasubramanian & Singh, 2016; Tharwat *et al.*, 2017)]. Conversely, unsupervised methods are used when the classes of the input signals are not known *a priori*, but may be an aim of the analysis (e.g. PCA, FA, t-SNE, MDS, unsupervised RF, neural networks, cluster analysis). Unsupervised approaches, such as t-SNE and MDS, can also be used to visualize and explore relationships among variables in the data in a space with fewer dimensions than present in the input data (Ramasubramanian & Singh, 2016). Moreover, data visualization procedures, such as Uniform Manifold Approximation and Projection (UMAP) may also prove useful for assessing separation of variables in acoustic space (Parra-Hernández *et al.*, 2020). When using such approaches for dimensionality reduction, however, careful attention should be paid that the method preserves between-object distance, as data visualization methods such as t-SNE and UMAP may sacrifice global structure in order to preserve local variance. The specific applications of each of these analyses differs and certain analyses may be better designed for certain data structures. For example, for behavioural questions, PCA is usually more appropriate than FA (Budaev, 2010; Wadewitz *et al.*, 2015). Additionally, each of these methods has their own underlying assumptions and applications, which should be understood before implementing the above analyses (Johnson & Wichern, 1982; Izenman, 2008).

Some analyses, such as combining acoustic metrics that have multiple data structures (e.g. combining raw point measures with vectors of frequency contours or proximity matrix output from spectrogram cross-correlation) require multiple multivariate procedures. Two multi-step multivariate procedures we see as particularly valuable to comparative bioacoustics are (i) conversion of proximity matrix data into tabular vectors (or new data columns representing the

“coordinates” of each signal in multidimensional space) and (ii) using acoustic feature space to estimate repertoire size or element diversity. Common procedures that require the conversion of proximity matrices to vectors, include DTW, RF analysis, or spectrogram cross-correlation analysis – as their output is pairwise comparisons (proximity matrices) of all acoustic units included in the analysis. Once these matrix data are converted to vectors, they can then be used as independent variables for further analysis or to plot sounds in acoustic feature space. For DTW, this process first involves interpolating the frequency contours to be the same number of points per signal before conducting the DTW, RF, or spectrogram cross-correlation analyses. MDS or t-SNE can be used to convert the resulting proximity matrix data to a set of tabular vectors. By plotting any tabular vector data, acoustic feature spaces can be created, from which acoustic areas, overlap, or distances can be calculated.

If the analysis goal is to classify signals into categories or calculate element diversity, unsupervised RF or clustering methods are effective. To calculate element diversity or acoustic variability from acoustic area, first perform one of the procedures described in the above paragraph. Then take the area surrounding the points for the acoustic signals or taxon of interest in feature space (i.e. acoustic area). Similarly, repertoire size or element diversity could be estimated from cluster analysis. Here, the optimal number of categories defined in the clustering procedure is interpreted to be the repertoire size or number of signal types for the taxon. However, we advise researchers to use caution, as discrete clustering may not be possible for species with large repertoires or continuously varying acoustic signals. In such instances, silhouette coefficients can help to quantify how similar signals in their own cluster are compared to signals in other clusters to evaluate the discreteness of vocal repertoires across taxa (e.g. Hedwig *et al.*, 2014). Among the most commonly used methods for data clustering are: k-means clustering, fuzzy k-means, and hierarchical clustering. Clustering analyses to determine repertoire size will likely need to be repeated on a species-by-species basis, whereas calculating element diversity from acoustic area requires including all species of interest in the feature space so that the magnitudes are comparable.

Above we outline a range of multivariate techniques to produce reduced or combined variables for phylogenetic comparative analysis. A very important caveat is that multivariate procedures often rotate and recombine data such that the original data structure, and therefore underlying evolutionary signal, can be lost (Uyeda *et al.*, 2015). Especially when the planned phylogenetic analyses will involve evaluating underlying evolutionary structure (i.e. phylogenetic signal or diversification rate analyses) or make implicit assumptions about it, then multivariate approaches that take the underlying evolutionary structure into account should be used (Adams & Collyer, 2018). Phylogenetic PCA (pPCA) is used to replace regular PCA while incorporating underlying evolutionary structure. A variety of other multivariate approaches that incorporate phylogeny have also been created and tested; however, their performance varies, including

pPCA (Harmon & Glor, 2010; Uyeda *et al.*, 2015; reviewed in Adams & Collyer, 2018). We strongly recommend understanding the weaknesses of these analyses before employing them, and evaluating comparative analyses using both multivariate and the raw component variables until these methods are improved (Mason *et al.*, 2017a; Adams & Collyer, 2018). We think that phylogenetically controlled morphometric analyses show great potential to allow sound to be analysed in a multidimensional space, as described above (Catalano, Goloboff & Giannini, 2010). Lastly, additional useful phylogenetically controlled statistical procedures that exist include calculations of phylogenetic signal for high-dimensional data (Adams, 2014a), phylogenetic Mantel tests (Harmon & Glor, 2010), and phylogenetic MANOVA, and ANCOVA (Revell, 2012; Goolsby, 2015; Fuentes-G *et al.*, 2016).

We also ask researchers to be cautious when choosing multivariate approaches and input variables in general (Björklund, 2019). As with any analysis, multivariate approaches can be sensitive to the quality and coverage of the input data. Specifically, researchers should be conscientious about the inclusion of variables with high collinearity into multivariate approaches; while certain approaches, especially PCA and RF, are robust to this (Afanador *et al.*, 2016), PCA output can be weighted towards similar variables that are more heavily represented in the analysis, especially when correlated. For example, if many moderately correlated frequency measurements are included in a PCA with a few time or rate variables, the first principal component will often be a composite of the frequency measurements because of their prevalence in the analysis. For researchers interested in determining the relative importance or weighting of component variables in acoustic classification, RF and implementing variable importance rankings could be a useful alternative to PCA. These caveats are important to keep in mind when considering research questions and what subsequent information can be extracted from the analysis, and we encourage researchers to conduct exploratory analyses to evaluate the quality and contribution of input variables to such multivariate approaches. Nevertheless, as classification and data-reduction methods diversify and improve, we envision that combining wide varieties of acoustic metrics using multiple multivariate techniques or mapping acoustic features in multi-dimensional spaces will become powerful ways to quantify variation among acoustic signals prior to comparative analysis.

(3) Tools and software

Dedicated software for measuring fine-scale features of acoustic signals include Raven Pro, Avisoft-SASLab Pro, Praat, Luscinia, SoundRuler, Syrinx, and KOE. MATLAB also has a signal-processing toolbox and a variety of acoustics packages exist in R (e.g. seewave, tuneR, warbleR, Rraven, Rpraat, soundgen; reviewed in Sueur, 2018). Sound-editing software can also be used to measure certain aspects of sound manually, as well as to annotate recordings. Popular software for sound editing includes Audacity and Adobe Audition.

For data reduction and downstream analysis, again a large number of packages are available in R (e.g. stats, vegan, MASS, Rtsne, fpc, pvclust, mclust, randomForest, ranger, caret). Some of these methods can also be run on graphical user interface (GUI)-based software (e.g. JMP, SPSS), although creating similarity matrices from tabular matrices usually requires command line software packages (i.e. programming in Python or R).

(4) Caveats for measuring sound

Appropriate acoustic analysis begins with understanding sound transmission and the recording process. We provide a detailed explanation of important methodological considerations about recording parameters and standardizing recordings before acoustic analysis in Appendix S1. One important point to recognize is that field recordings collected across a wide range of habitats and by different recordists will likely vary greatly in recording quality and format. Such variation can cause artifacts during acoustic analysis, but given proper attention before sound analysis, they can usually be overcome. We recommend standardizing sample rate and bit depth across all recordings prior to analysis, especially if recordings come from multiple sources. Recordings with low signal-to-noise ratio (SNR; e.g. recordings with a high degree of background noise or faint signals) can significantly affect the precision of certain acoustic parameters, and so such recordings should typically be avoided (Araya-Salas *et al.*, 2017; Table S1, Appendix S1). Recordings that appear to contain distortion (e.g. over-amplification, aliasing; see Appendix S1) should also be avoided. In sound collections, choosing recordings ranked of a certain quality or higher can help avoid such issues (e.g. Billings, 2018). Recordings collected or stored with ‘lossy’ compression (compression with irreversible information loss; e.g. mp3) can distort acoustic measures of single extreme values (e.g. peak frequency) and affect the precision of DTW analysis (Araya-Salas *et al.*, 2017), although the resulting measurement error could be less problematic for across-species comparisons. Nonetheless, proper parameter selection can avoid some of these issues, but using uncompressed recordings (e.g. WAV file format) or recordings with lossless compression, such as FLAC file format, is preferable.

Researchers should also adopt practices for standardizing collection of acoustic metrics. First and foremost, spectrogram parameters (window settings) should be standardized so that frequency and time are measured at consistent resolutions across all recordings. Traditional ‘by-eye’ measurements measured directly on the spectrogram should be replaced with automated, energy-based or threshold measurements whenever possible, as signal strength and quality can greatly affect measurements taken directly on the spectrogram (Charif *et al.*, 2010; Zollinger *et al.*, 2012; Ríos-Chelén *et al.*, 2017). Additionally, measurements of absolute sound source level (i.e. amplitude, power, energy, pressure) require calibrated equipment and in some cases calculations of sound transmission. Therefore, unless recordings were

collected in standardized or controlled conditions, measures of absolute amplitude should be avoided. Lastly, researchers should be aware that spectrograms reflect how humans perceive sound (Lyon & Ordubadi, 1982; Dooling & Prior, 2017). Therefore, not all acoustic features that can be measured may be detected by or relevant to the species being studied (e.g. centre frequency, a computational signal measurement which the vertebrate auditory system does not encode, is not biologically relevant, unlike peak frequency). Similarly, acoustic measures of similarity in signal structures represent a proxy for perceptually important differences among distinct taxa, but may not reflect what is perceived by those taxa. Unfortunately, we seldom know *a priori* which variables are biologically relevant. One approach to address this is to test if acoustic variables identified as statistically significant or distinct elicit a response in the study species (e.g. Rand & Ryan, 1987). An interesting future direction for bioacoustics research would be to analyse sound within taxon-specific acoustic perceptual models (Clemins & Johnson, 2006; Stoddard & Prum, 2008). This is especially relevant for taxa with auditory ranges narrower than the signals they produce or that perceive sound in specialized ways that are not easily measured or encoded in spectrograms (e.g. Narins & Capranica, 1976). Such models could also take ‘just noticeable differences’ in animal sound perception into account to reflect accurately the scale or components of a signal that taxa are capable of perceiving (Kuhl, 1981). However, this is not known for many species.

(5) Best practices for measuring sounds

When the goal of a study is to classify or quantify a wide variety of acoustic signals, and there is little *a priori* knowledge of which acoustic variables may be biologically relevant, the best option is usually to measure a wide variety of acoustic features that capture the range of structural variation in the signals being studied. Incorporating measurements that best capture the spectral and temporal properties of a study system will improve the likelihood that meaningful variation is detected and categorized in subsequent phylogenetic comparative analyses. Therefore, choosing appropriate measurements should be done on a study-by-study basis, after some preliminary examination of signal variation and keeping the goals of the study in mind. Overall, of greatest importance is to ensure sufficient measurement precision and inclusion of appropriate variables to capture the natural variation within and among acoustic signals for the species of interest. We recommend automated procedures for feature extraction (e.g. robust measurements based on energy distributions and threshold metrics) and classification (e.g. cluster analysis, RF) whenever possible to limit extraneous error and human bias. In addition, using algorithms for classification and analyses that remove or downplay non-informative parameters (e.g. RF) may be especially beneficial for detecting meaningful variation.

Regardless of which metrics are collected, all measurements, metrics, and other associated acoustic terminology

used should be defined in the publication, preferably with figures (e.g. Fig. 1). Clear definitions are, of course, necessary for the analyses to be repeated but may easily be overlooked in large, multistage projects. More broadly, clearly defining metrics can help researchers to standardize analyses across studies and improve reproducibility. With the increasing movement towards data archiving, standardizing the basic metrics collected and procedures for sound analysis will allow hard-earned acoustic measurements to be combined into larger comparative analyses in the future. Along these lines, we encourage archiving not only of raw acoustic measurement data, but also of the associated recordings and annotations. Such practices will best enable the output and results from current comparative studies to be built upon as knowledge, media archives, and analytical tools improve (Caetano & Aisenberg, 2014).

III. COMPARATIVE PHYLOGENETIC ANALYSES WITH SOUND

Phylogenetic comparative analyses have recently been used to make considerable advances in our understanding of the evolutionary processes responsible for diversity on Earth (May-Collado, Agnarsson & Wartzok, 2007; Amézquita *et al.*, 2009; Jetz *et al.*, 2012; Rabosky *et al.*, 2013; Dale *et al.*, 2015; Sauquet *et al.*, 2017). Such studies are becoming increasingly popular to address questions about signal evolution, including comparative studies of detailed acoustic structure (Derryberry *et al.*, 2012; Mason *et al.*, 2017a; Ligon *et al.*, 2018). For these kinds of studies, phylogenetically controlled analyses are necessary because they transform evolutionary questions into statistical models that enable researchers to control for the statistical non-independence caused by shared evolutionary history (Felsenstein, 1985; Martins & Hansen, 1997). Also, incorporating phylogeny into comparative studies allows researchers to evaluate hypotheses for evolutionary mechanisms beyond phylogeny, examine evolutionary patterns leading to trait divergence, and study the role of traits in diversification (Hernández *et al.*, 2013; Garamszegi, 2014; Rabosky *et al.*, 2014). Even poorly resolved phylogenies provide improved accuracy compared to no phylogeny at all (Boettiger, Coop & Ralph, 2012; but see Davies *et al.*, 2012; Paradis, 2014). For these reasons, it is important to incorporate phylogenetic information in all statistical analyses with comparative data (Felsenstein, 1985; Martins & Hansen, 1997). Altogether, phylogenetically informed comparative analyses provide an important perspective on patterns and processes of trait macroevolution than cannot be earned by other means (Freckleton, Cooper & Jetz, 2011).

Several recent good reviews and books have been written on phylogenetic comparative analyses, which we suggest researchers consult for more in-depth coverage of these methods (Harvey & Pagel, 1991; Gingerich, 2009; Nunn, 2011; Garamszegi, 2014; Joy *et al.*, 2016). Here, we briefly

summarize the range of available phylogenetic comparative approaches, discuss some relevant bioacoustics case studies, and outline general limitations of comparative approaches to behavioural data. We finish by providing guidance for applying these methods to acoustic data given particular data structures and phylogenetic analysis of interest (Fig. 3).

(1) Types of phylogenetic analyses

There are several broad classes of phylogenetic comparative analysis and questions that can be addressed with phylogenetic comparative approaches. Below, we summarize current methods employed within each broad class. A major distinction among the methods within each class is whether they are appropriate for discrete (e.g. presence/absence or small/medium/large) *versus* continuous trait data (Harvey & Pagel, 1991; Garamszegi, 2014; e.g. Odom, Omland & Price, 2015). In some instances, certain approaches have been created for multi-state *versus* binary traits or response variables. Therefore, as with any statistical analysis, thinking in advance about data structure and the types of variables present in your data set will enable you to determine the appropriate analysis. In the case studies, we emphasize methods for continuous data, as the acoustic analyses described above are primarily aimed at extraction of continuous variables, but we list a range of analyses for a variety of data types, as major bioacoustics questions can be addressed with discrete data types (Odom *et al.*, 2014; Riede *et al.*, 2016; Tobias *et al.*, 2016; Snyder & Creanza, 2019).

(a) Quantifying phylogenetic signal

Phylogenetic signal refers to the extent to which closely related species resemble one another based on a trait of interest (Pagel, 1999a; Blomberg *et al.*, 2003; Davies *et al.*, 2012; Münkemüller *et al.*, 2012). For most phylogenetic comparative studies, it is important to quantify and understand the extent to which the trait of interest remains similar across close taxa. Phylogenetic signal can be measured using several parameters. Popular methods include Pagel's lambda (λ), Blomberg's K , Grafen's ρ , Ornstein–Uhlenbeck (OU) model parameter α , and Fritz & Purvis's D (Revell, Harmon & Collar, 2008; Kamlar & Cooper, 2013; Symonds & Blomberg, 2014). Recent methods also allow for tests of phylogenetic signal in multi-dimensional traits (Adams, 2014a). Note that Pagel's lambda is often provided with the output of phylogenetic comparative analyses (Hadfield, 2010a,b). However, when given as output with regression analyses, lambda applies to the residual errors of the regression model, not the response variable. Therefore, these estimates of phylogenetic signal do not represent the phylogenetic signal of the response variable, but rather for the variance explained by phylogeny for the regression of the traits' response and predictor variables compared to one another (Symonds & Blomberg, 2014). In most instances, it is important to directly calculate phylogenetic signal for the input variables and, depending on the question, it may be

important to investigate both. See Revell *et al.* (2008) for discussion of appropriate and inappropriate interpretation of phylogenetic signal.

Few comparative bioacoustics studies have directly focused on estimating phylogenetic signal; however, phylogenetic signal is usually reported for the acoustic traits within the study, which we encourage. Among bioacoustics studies that have focused on phylogenetic signal, Gingras *et al.* (2013) measured the strength of phylogenetic signal in five acoustic parameters from advertisement calls among 90 species of anurans. Despite strong selection on calls *via* mating preferences, the authors found that there was a strong phylogenetic signal in all five acoustic traits. They argued that this might indicate constraints on signal diversification among these anurans. It is important to note, however, that simulations demonstrate that phylogenetic signal alone is not a clear indicator of the rate of evolution or the evolutionary processes leading to diversification (Revell *et al.*, 2008). In a study of phylogenetic signal in the territorial songs of crests and kinglets (Aves: *Regulus*), Päckert *et al.* (2003) measured several different features of song including syntax, subunit measures, and abundance of certain components. Interestingly, the measured traits also differed in the extent to which they were learned or innate. The authors found that some measures, such as the syntax of whole-song structure, showed a strong phylogenetic signal while other learned song components did not. This study reveals interesting use of phylogenetic signal to test hypotheses about expectations for which aspects of a signal may be more closely tied to phylogeny, an important and valuable use of this metric (Päckert *et al.*, 2003).

(b) Evolutionary rate and diversification

Evolutionary rate analysis allows researchers to examine the extent to which traits change over evolutionary time (Gingerich, 2009), to compare evolutionary rates among traits (Adams, 2014b) and to evaluate variation in evolutionary rates between phenotypic sequence subunits, which can be applied to multi-syllable vocalizations (Caetano & Beaulieu, 2020). Additional methods have been developed to evaluate whether the evolutionary rate of a continuous trait is affected by another trait, either discrete (O'Meara *et al.*, 2006) or continuous (Weir & Lawson, 2015), and the following analyses have been extended to allow for or to test correlation between speciation and discrete traits (binary traits biSSE; Maddison, Midford & Otto, 2007), multi-state traits (muSSE; FitzJohn, 2012), continuous traits (quaSSE; FitzJohn, 2010) and trait evolutionary rates (Adams *et al.*, 2009; Rabosky *et al.*, 2014). A more recently developed method allows for simultaneously testing of the effect of a discrete trait on trait evolutionary rate and the association between traits (Fuentes-G *et al.*, 2016). These methods have been used to compare the rate of evolution of morphological and acoustic traits (Medina-García, Araya-Salas & Wright, 2015) and to test the association between evolutionary rate and signal function (Weir, Wheatcroft & Price, 2012), as well

as developmental mechanisms of vocal signals (Mason *et al.*, 2017a).

Generally, analyses of evolutionary rate and diversification require well-resolved sets of trees with dated nodes (Tarver & Donoghue, 2011; Paradis, 2013). An important consideration when using these models with acoustic data is that they often require assessing the appropriate underlying model of evolution (e.g. Brownian Motion, OU). Therefore, it is important that any data-reduction steps prior to evolutionary rate and diversification analyses should take phylogeny into account. Nevertheless, these analyses are highly subject to model misspecification, and improved models are still being developed (e.g. Harmon & Glor, 2010; Uyeda *et al.*, 2015; Adams & Collyer, 2018). Until improved methods are developed, we encourage researchers to compare comparative tests of diversification and rate for acoustic data using raw acoustic variables in addition to any planned multivariate data (Mason *et al.*, 2017a). Another consideration is whether the phylogenetic hypothesis contains sufficient information to accurately reconstruct diversification dynamics. The ability to infer diversification parameters from extant species phylogenies has been questioned and fossil calibration is advised [Quental & Marshall, 2010; Louca & Pennell, 2020; but see Dos Reis & Yang (2013) for possible caveats of fossil calibration].

Despite these present limitations, several comparative studies have incorporated evolutionary rate and diversification into analyses of acoustic signals. For instance, Mason *et al.* (2017a) used raw acoustic data (i.e. signal metrics *sensu* Table S1), as well as pPCA, to quantify shifts in the rate of vocal evolution and speciation across two major radiations of passerine birds. They found evidence for coincident evolutionary bursts in rates of speciation and song evolution among both groups. Further, several studies have investigated the potential role of acoustic signals in speciation across closely related taxa. For instance, Delmore *et al.* (2015) quantified divergence in song, plumage, and morphology among sister pairs of North American migratory birds with different migratory strategies to evaluate diversification of each set of traits as a measure of reproductive isolation. Such studies highlight the potential for acoustic signals to play a large role in diversification and speciation among diverse taxa, and thus when evolutionary models can be properly incorporated, represent a promising avenue for future research.

(c) Ancestral state reconstruction

Ancestral state reconstruction is used to infer the most likely ancestral states for a trait of interest and evaluate how that trait changed over evolutionary time. Three main methods exist: maximum parsimony, maximum likelihood, and Bayesian estimation (Pagel, 1999b; Pagel, Meade & Barker, 2004; Maddison & Colu, 2015). Parsimony minimizes the number of changes required to explain the distribution of characters in extant taxa (Harvey & Pagel, 1991; Joy *et al.*, 2016). While parsimony may give accurate reconstructions, especially when evolutionary rates are expected to be slow

(Cunningham, Omland & Oakley, 1998), maximum likelihood (ML) and Bayesian approaches are more sophisticated and have more realistic assumptions (Royer-Carenzi, Pontarotti & Didier, 2013; Joy *et al.*, 2016). ML attempts to find the parameter values that maximize the probability of the data given the underlying phylogeny, thus taking branch lengths into account as estimates of evolutionary time (Joy *et al.*, 2016). In addition, different forward and reverse rates of evolution can be specified in one- *versus* two-parameter models (Mooers & Schluter, 1999). Nevertheless, ML requires *a priori* specification of such parameters, which are often unknown. Furthermore, rates of change may not be stable over evolutionary time and are sensitive to variation in tree topology (Schultz & Churchill, 1999; Joy *et al.*, 2016). Bayesian approaches can integrate uncertainty in tree topology, branch lengths, and parameter estimates into the ancestral state reconstruction using Markov Chain Monte Carlo techniques. Such techniques account for sources of uncertainty as distributions or ‘liability’ terms calculated from the data and incorporated into the model (Huelsenbeck & Bollback, 2001; Pagel *et al.*, 2004; Revell, 2014). Both empirical and hierarchical Bayesian approaches can be employed, but hierarchical Bayesian methods are especially useful for averaging probabilities over a set of possible trees (Joy *et al.*, 2016). Still, users of ancestral state reconstruction approaches should be aware of underlying assumptions and limitations due to taxon sampling and tree topology (Losos, 1999; Omland, 1999; Salisbury & Kim, 2001; Li, Steel & Zhang, 2008; Revell *et al.*, 2008; Marshall, 2017). Specifically, ancestral state reconstruction is always an inference of how and when traits evolved in the past, as demonstrated by the fact that confidence intervals surrounding ancestral states are usually quite large (Cunningham *et al.*, 1998; Garland & Ives, 2000; Oakley & Cunningham, 2000). To partially deal with this uncertainty it is advisable to compare results using various underlying evolutionary models and reconstruction methods (e.g. compare ML and Bayesian approaches), to conduct sensitivity analyses (Cunningham *et al.*, 1998), in addition to report confidence intervals (Garland & Ives, 2000; Revell, 2013).

Ancestral state reconstruction of acoustic signals has provided valuable insights into when certain major signalling strategies and behaviours evolved (Shelley & Blumstein, 2005; Odom *et al.*, 2014; Riede *et al.*, 2016; Tobias *et al.*, 2016; Forti *et al.*, 2018). Most of these studies, however, have primarily used discrete traits to conclude when broad categories of vocalizations came to exist, whereas surprisingly few studies have applied ancestral state reconstruction methods to continuous or multiple acoustic features (exceptions include Rand & Ryan, 1987; Price & Lanyon, 2002; Price, Friedman & Omland, 2007; Goutte *et al.*, 2016). In one notable exception, Price *et al.* (2007) scored song as presence/absence of 26 vocal characters to investigate how songs have changed over time in the New World blackbirds, with the ultimate goal of comparing evolutionary patterns of song to plumage evolution. They found that New World blackbird song is fairly evolutionarily labile, which paralleled patterns of convergent evolution in

plumage. This study demonstrates an interesting approach for evaluating vocal evolution with discrete traits and analyses, while still extracting a considerable amount of vocal variation. With the current availability of continuous comparative approaches, these kinds of analyses can also be conducted with continuous acoustic variables. For example, a recent study with torrent-dwelling frogs compared several measured vocal variables to reconstructed features of calling-site habitat (Goutte *et al.*, 2016). They were able to show that vocalizations of torrent-dwelling frogs have likely been constrained by the noisy environments in which they evolved. In another creative use of ancestral state reconstruction, Rand & Ryan (1987) reconstructed ancestral male Túngara frog (*Engystomops pustulosus*) calls to assess the response of females to ancestral vocal features. Similarly, we think an exciting future use of ancestral state reconstruction will be to reconstruct composites of acoustic features to investigate the origins of complex acoustic signals (e.g. Adams, 2014*b*; Sauquet *et al.*, 2017).

(d) Phylogenetic correlative analyses

Phylogenetic correlative analyses are used to assess correlated evolution of two or more traits, including both discrete and continuous variables (Martins & Hansen, 1997; Paradis, 2014). Pagel’s discrete test of correlated change is a popular approach for analysing correlated evolution and transition rates among discrete character states (Pagel, 1994). For continuous data, Felsenstein’s (1985) independent contrasts (PIC; Garland, Bennett & Rezende, 2005), phylogenetic least squares (PGLS; Grafen, 1989; Paradis, 2011), and Markov Chain Monte Carlo simulations in a mixed-model framework [phylogenetic mixed models (PGLMMs); Martins & Garland, 1991; Housworth, Martins & Lynch, 2004; Hadfield & Nakagawa, 2010] all allow for phylogenetically controlled correlations that take topology and branch lengths into account. A key strength of the PGLS approach is that it is a straightforward regression model akin to ordinary least squares (OLS; in fact, PGLS approaches OLS output when the phylogenetic signal is weak; Symonds & Blomberg, 2014). Plus, this approach allows for alternative underlying evolutionary models to be incorporated and evaluated (e.g. Brownian motion, early burst, OU; see Section III.2*d*). If the data set contains repeated measures per species, however, the data are likely better suited to PGLMMs, as they can account for within-species variation (Martins & Hansen, 1997; Garamszegi, 2014). In addition, the Bayesian framework of some PGLMMs (e.g. the R package MCMCglmm) enables uncertainty about the phylogeny to be incorporated as a set of alternative tree topologies (Hadfield, 2010*b*; Hadfield & Nakagawa, 2010; Garamszegi & Gonzalez-Voyer, 2014). Phylogenetically controlled ANOVA, MANOVA, and ANCOVA procedures also exist and can be useful (Revell, 2012; Goolsby, 2015; Fuentes-G *et al.*, 2016). Similar to the transition rate analyses of Pagel’s discrete test, path analysis is a promising and appropriate procedure for examining the evolutionary order of events using continuous data sets (Gonzalez-Voyer & von Hardenberg, 2014). Recently

developed phylogenetic variable-rate regression models can be used to measure branch-wise rates of trait evolution in order to infer positive phenotypic selection and its link to other traits (Baker *et al.*, 2016).

Evolutionary biologists interested in acoustic evolution have answered a range of interesting and diverse questions using phylogenetic correlative analyses on signal metrics. For example, several phylogenetic comparative approaches were incorporated into the analyses of Gonzalez-Voyer *et al.* (2013) who investigated evolutionary relationships among a number of traits (including vocalizations) in barbet bird species. These authors lay out the logic behind their analytical decisions and use a number of current phylogenetic approaches required to facilitate confidence in their findings that larger barbet species produce longer, lower-frequency notes, that species living at higher altitudes produce longer songs, and that different elements of barbet vocalizations show distinct evolutionary rates. Likewise, an investigation into the influence of vocal learning on acoustic diversification of parrot vocalizations by Medina-Garcia *et al.* (2015) analysed several signal metrics separately (e.g. peak frequency, entropy, duration) to determine if these parameters exhibit different evolutionary trajectories than morphological traits. This powerful study combined ‘traditional’ correlative analyses (e.g. phylogenetic regression) with those focused on analyses of evolutionary rates (Adams, 2013) and found that the acoustic elements studied showed similar patterns and rates of evolutionary change to morphological traits, despite the potential for learning to increase the evolutionary rate of vocalizations. By contrast, Ligon *et al.* (2018) used multivariate metrics (see Section II.2c) to input signal measurements into a multidimensional feature space, which facilitated the classification of notes into distinct types. After identifying all the notes in an individual sequence, the authors analysed ‘acoustic richness’ and ‘acoustic diversity’ using sliding-window analyses that identified the most complex vocal sequence for each individual from all species. It was these computational measures (acoustic richness, diversity) that were then analysed using phylogenetic correlative analyses. Using these approaches, Ligon *et al.* (2018) found that vocal complexity is positively correlated with both behavioural and chromatic complexity at an evolutionary scale in the birds-of-paradise.

(2) Caveats for phylogenetic comparative analyses

Although powerful, phylogenetic comparative analyses are still developing and require important caveats. These analyses reflect our current best estimates of the evolutionary process and are dependent on the current resolution of the input trees. In addition, each approach has its own limitations and assumptions and combining these approaches with highly dimensional acoustic data can be complicated (Cunningham, 1999; Losos, 2011; Uyeda *et al.*, 2015; Title & Rabosky, 2019). Therefore, analyses should be regarded with appropriate skepticism, and inferences about deeper evolutionary processes should be made with caution (Losos, 2011; Marshall, 2017). Below we identify important limitations to phylogenetic comparative analyses of animal sounds.

(a) The choice of phylogeny

Any phylogenetic comparative study requires an underlying phylogenetic hypothesis, usually a phylogenetic tree or network, ideally with proposed branch lengths. All phylogenies are estimates (Rosenberg & Kumar, 2001) and while some phylogenetic comparative analyses may be somewhat robust to the underlying phylogeny, a more resolved phylogeny leads to higher confidence in the results (Harvey & Pagel, 1991; Garamszegi & Gonzalez-Voyer, 2014). The influence of accuracy of the underlying phylogeny is reviewed in Heath, Hedtke & Hillis (2008). For discussion of the methods for used to construct the phylogeny see Rosenberg & Kumar (2001) and Berlin, Tomaras & Charlesworth (2007).

One can compensate for phylogenetic uncertainty by including a posterior sample of phylogenetic trees within the analysis. This accounts for potential error in phylogenetic estimation and sampling. It is becoming standard practice for large sets of trees to be made available (e.g. birdtree.org) and Bayesian approaches make it possible to incorporate randomly selected sets from a posterior distribution of trees representing the range of phylogenetic uncertainty. In birds, one can sample and analyse over 500 trees (Leighton, 2017), although performing the analysis on several dozen trees (~50) may be sufficient (Griesser *et al.*, 2017). Rubolini *et al.* (2015) offer good advice when using widespread phylogenetic data, such as birdtree.org.

(b) Assessing the underlying models of trait evolution

Rate and phylogenetic comparative analyses require specification of an underlying model of evolution (Freckleton *et al.*, 2011). These analyses assume that the chosen model is a realistic estimation of the evolutionary process for that data set. For behavioural or life-history data, however, the appropriate underlying model is usually unknown or may not be intuitive. Some of the more common models of evolution used are Brownian motion, OU, lambda and ACDC. A Brownian motion model represents trait evolution as responding to many small evolutionary forces and is commonly used as a null model expected in the absence of strong stabilizing selection. The OU process models the strength of the evolution towards one or more theoretical optima, which can be used to represent adaptive evolution (Butler & King, 2004). Lambda models the contribution of phylogeny to trait values in which the lambda parameter ranges between no phylogenetic effect and pure Brownian motion (Pagel, 1999a), while ACDC (Blomberg *et al.*, 2003; also called the Early-Burst model *sensu* Harmon *et al.*, 2010) represents a trend in evolutionary rate across the phylogeny, exponentially increasing or decreasing. PGLS procedures recommend using model selection criteria (e.g. Akaike information criterion, AIC) to assess the fit of multiple models of evolution to the data set before continuing with analyses (Paradis, 2011; Symonds & Blomberg, 2014) and similar procedures apply to other analyses (Sullivan & Joyce, 2005).

(c) Sample size and taxon sampling

Taxon sampling is important in resolving the underlying tree, and is also important for making broad inferences across lineages. While Blomberg *et al.* (2003) concluded that trees with greater than 20 species have more robust results, researchers need to take care to also consider whether species are clustered on a tree, and the topology (symmetry) of the tree. Consequently, researchers should balance increasing the total number of sampled taxa while also identifying taxa and clades that will provide the most clarity given the hypotheses the researcher is testing. Arnold & Nunn (2010) provide methods for determining the proper taxonomic sampling and data collection (i.e. ‘phylogenetic targeting’).

Sample size is also relevant for estimates of trait evolution. In most cases, measures within a species are variable. Previously, many comparative analyses simply used single or mean trait values to represent a species. Recent methodological developments for analyses of trait evolution recommend taking within-species variation into account (Garamszegi, 2014). Repeated sampling within and among species is likely especially important for vocal behaviour, which can vary dramatically within and among individuals and populations.

(d) Trait evolution and behaviour

Traits differ in their phylogenetic signal. Blomberg *et al.* (2003) showed that behavioural traits exhibit a lower phylogenetic signal than morphological, life-history, or physiological traits, and concluded that “behavior is relatively labile evolutionarily” (p. 730). Consequently, researchers have raised concerns about the validity of applying phylogeny to behavioural data when phylogenetic signal is low, especially ancestral state reconstruction, because of the potential loss of evolutionary signal in highly labile traits (e.g. Losos, 1999). Specifically, recurring trait shifts within reconstructions with large confidence intervals suggest that two traits could occur together by chance. However, Revell *et al.* (2008) conclude that phylogenetic signal does not predict evolutionary process or rate, restoring confidence that behavioural traits are evolving in similar ways as other traits. Thus, phylogenetic methods should be useful for examining the evolution of behaviour. A potential concern for high-dimensional acoustic data is that transition rates are sometimes unable unambiguously to identify correlated trait shifts among multiple traits. Statistical models that allow different traits to evolve according to different models of evolution (Losos, 2011; Garamszegi, 2014, pp. 11–12) and path analysis both offer promising ways to look at evolutionary transitions within continuous data (Gonzalez-Voyer & von Hardenberg, 2014).

(e) Correlation is not causality

An association between traits at the tips of a phylogeny can often be explained *via* multiple hypotheses. Although experiments provide true tests of causality, testing evolutionary causation is usually limited to species with extremely short generation times. For species that do not have short generation times, certain analyses, such as transition rates in Pagel’s discrete test

(Pagel, 1994) and path analysis (Gonzalez-Voyer & von Hardenberg, 2014) allow some conjecture as to the probability that one event follows another, or whether or not a particular relationship among traits is meaningful (Garamszegi, 2014).

(f) Vocal learning and cultural evolution

Whether or not an acoustic signal develops *via* vocal learning can profoundly affect how it evolves. In some species, vocal signals are socially transmitted such that the development of an individual’s vocal signal depends on experience with the vocal signals of other individuals (Janik & Slater, 2000). To date, vocal learning is known to occur in some avian (parrots, hummingbirds, oscine passerines and some suboscines) and mammalian (cetaceans, elephants, phocid pinnipeds, humans, and bats) taxa. Unlike genetically determined signals that in vertebrates are inherited vertically (from parent to offspring), socially transmitted traits can also be transmitted horizontally (between individuals within the same generation) and obliquely (across generations of unrelated individuals; Danchin & Wagner, 2010). This flexibility in transmission modes poses a potential difficulty for phylogenetic approaches because such methods are built on the premise that traits are inherited vertically (Mace & Holden, 2005; Gray & Watts, 2017; Mesoudi, 2017). A second challenge is that learned vocalizations could evolve at faster tempos than genetically determined signals, because cultural innovations or learning errors (‘cultural mutations’) may occur at a faster rate than genetic mutations (Danchin *et al.*, 2004). The potentially high speeds of cultural ‘mutation’ could increase the chance of convergence (Price & Lanyon, 2002; Delsuc, Brinkmann & Philippe, 2005) or the loss of phylogenetic signal due to rapid divergence (analogous to ‘saturation’ *sensu* Delsuc *et al.*, 2005; Filatova *et al.*, 2015). Nonetheless, studies have shown that, for some taxa at least, vocal characters can be highly conserved across species (Price & Lanyon, 2002; Podos, Huber & Taft, 2004; Sakata & Vehrencamp, 2012; Medina-García *et al.*, 2015). In addition, studies of human language found some evidence that culturally evolved traits can diversify in a tree-like manner with limited transfer across ‘lineages’, so that comparative approaches can be informative within as well as between species (reviewed in Pagel, 2009; Currie, Greenhill & Mace, 2010; Hoppitt & Laland, 2013; Mesoudi, 2017). While there are modelling approaches that explicitly test for and incorporate ‘horizontal’ transfer of inherited information into analyses (reviewed in Mesoudi, 2016), in cases of frequent horizontal or oblique inheritance, or very rapid diversification, comparative approaches are expected to be unhelpful (Filatova *et al.*, 2015; Mesoudi, 2016). In general, however, phylogenetic correlative analyses provide useful methods to investigate the evolution across clades of both learned and non-learned vocalizations and, when used with care, offer an exciting frontier for understanding the tempo and modes of evolution in cultural traits (Mason *et al.*, 2017a). In addition, the cultural evolution process itself can be modelled using phylogenetic methods and the resulting phylogenetic hypotheses can be used to infer the dynamics of the cultural micro-evolutionary process from a comparative framework (Mace & Holden, 2005; Mesoudi,

2017). This approach will add a historical perspective to studies of animal cultures and a more detailed view of mechanisms involved in the micro-evolution of socially acquired (vocal) traits.

(3) Best practices for phylogenetic comparative analyses with sounds

In conclusion, phylogenetic comparative analyses are powerful tools that can be leveraged to answer important questions about signal evolution, especially involving acoustics. With the proper understanding of their limitations and underlying assumptions, these methods can begin to shed light on how acoustic signals have evolved and diversified across numerous taxa. When planning and conducting phylogenetic comparative analyses, it is important to understand the broad classes of analyses available and how they can be used to address your research question. It is also important to think about the acoustic data structure (are the variables discrete or continuous? How many variables are there? Are any data reduction steps planned?). The data structure determines the specific comparative analyses within these broad classes that are appropriate, and, in turn, there may be important caveats for how to treat the data, particularly during data reduction. For example, multivariate procedures can remove underlying phylogenetic structure, therefore, taking phylogeny into account within these procedures is necessary if subsequent analyses rely on determining an appropriate underlying evolutionary model (Catalano *et al.*, 2010; Uyeda *et al.*, 2015; Adams & Collyer, 2018). We provide guidelines for these considerations in Fig. 3.

Within phylogenetic comparative analyses it is also best practice to consider and evaluate multiple underlying models of evolution and/or algorithms underlying the comparative analyses. For example, diversification rate and certain correlative analyses (e.g. PGLS) require evaluating whether Brownian motion, OU, or other underlying evolutionary models best explain the overall patterns of the data. This is often done by comparing the fit of models constructed with different underlying evolutionary models *via* an information criterion, such as AIC. Similarly, for ancestral state reconstruction, it is important to evaluate the underlying assumptions of competing models and whether the outcome of the analysis is influenced by the particular reconstruction method chosen. For analyses that attempt to infer ancestral trait values using phylogenetic reconstruction, it is customary to compare multiple reconstruction methods (e.g. Bayesian and ML) to assess that they generally provide similar results (Omland, 1999).

Lastly, as with all statistical models, know the inference power and associated uncertainty. Remember that these are powerful tools but are still correlative and not causative. It is important to draw conclusions within the scope of the analysis and project.

IV. CONCLUSIONS

- (1) Combining continuous acoustic metrics and advanced phylogenetic comparative analyses has great potential

to advance our understanding of signal evolution; however, implementing both approaches can be challenging.

- (2) For acoustic analysis, it is most important to identify, define, and partition the acoustic signals into component structures that can be analysed and compared across the species of interest. In addition, we encourage researchers to give careful consideration to the metrics chosen for acoustic analysis. When beginning acoustic analysis, the metrics used should be aligned with the research question, the kinds of variation observed in the signals of interest, and the nature of the downstream phylogenetic comparative analyses to be employed. We also remind researchers to standardize recording specifications, quality, spectrogram parameters, and metrics as best they can prior to acoustic analysis. Standardizing recording specifications and choosing metrics can now be done in R (e.g. warbleR, seewave), so this largely comes down to good data-management practices.
- (3) For data reduction of acoustic data prior to comparative analysis, we ask researchers to be aware of important caveats about data-reduction procedures that are appropriate for specific phylogenetic comparative analyses. For example, phylogenetic analyses for which the output depends on assessing underlying models of evolution, any data-reduction procedures used should take underlying models of evolution into account. For more details, we provide guidelines to help researchers determine the best steps and procedures for incorporating highly dimensional acoustic data with these analyses (Fig. 3).
- (4) For phylogenetic comparative analyses, we encourage researchers to explore the vast amount of resources and powerful computational tools available. Recent advances in continuous approaches for phylogenetic comparative analyses, including Bayesian ancestral state reconstruction and phylogenetically controlled mixed models enable deep dives into exploring evolutionary patterns, while taking uncertainty of tree topology, model structure, and even complex data structure into account. However, it is important to remember that these are only models of our best estimates of the evolutionary process. Therefore, we encourage researchers to run phylogenetic analyses with multiple different approaches to look for overall similarity and robustness of results given alternative underlying models of evolution, input parameters, and phylogenetic hypotheses (different phylogenetic trees).
- (5) With recent advances in bioacoustic practices, data availability, and phylogenetic comparative approaches, we see a bright future for analyses of acoustic signal evolution. We also see opportunities for advancing specific areas at the interface of these fields. For example, there is a need to standardize terminology. This requires increasing our knowledge of the functional diversity of these signals, as well as agreeing upon appropriate

terminology for investigating this diversity. In addition, we advocate the adoption of automated methods for acoustic data collection and feature extraction wherever possible. This includes extracting robust feature measurements, but also using automated procedures for classifying signals and evaluating their syntax. We also envision more streamlined, automated processes for the detection of and separation of signals into comparable acoustic units, such that the entire procedure of acoustic signal analysis is more efficient, standardized, and reproducible. Lastly, we strongly encourage improved data-archival practices for both recordings and associated annotation data to facilitate data sharing. With such improvements, we expect the field of comparative bioacoustics to make great strides in our understanding of signal evolution and the drivers of animal diversity at large.

V. ACKNOWLEDGEMENTS AND AUTHOR CONTRIBUTIONS

We thank Daniel Blumstein, William Foster, and one anonymous reviewer for helpful comments. A.H.D., K.J.O., and A.N.R. were supported in part by a U.S. National Science Foundation IOS Grant No.: 1755132. A.H.D. was also supported by a University of Wollongong Vice-Chancellor's Research Fellowship. K.J.O., R.A.L., and G.M.L. were supported by U.S. National Science Foundation Postdoctoral Research Fellowships in Biology (Grant No. 1612861, 1523895, and 1523748, respectively). R.R.G. was supported by a Natural Sciences and Engineering Research Council of Canada Postdoctoral Fellowship. **Authorship statement:** All authors contributed to the design, planning, organization, research, and writing of this manuscript. K.J.O., M.A.-S., J.L.M. and A.N.R. contributed especially to conceptualization and writing.

VI. REFERENCES

- ADAMS, D. C. (2013). Comparing evolutionary rates for different phenotypic traits on a phylogeny using likelihood. *Systematic Biology* **62**, 181–192.
- ADAMS, D. C. (2014a). A generalized k statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. *Systematic Biology* **63**, 685–697.
- ADAMS, D. C. (2014b). Quantifying and comparing phylogenetic evolutionary rates for shape and other high-dimensional phenotypic data. *Systematic Biology* **63**, 166–177.
- ADAMS, D. C., BERNS, C. M., KOZAK, K. H. & WIENS, J. J. (2009). Are rates of species diversification correlated with rates of morphological evolution? *Proceedings of the Royal Society B: Biological Sciences* **276**, 2729–2738.
- ADAMS, D. C. & COLLYER, M. L. (2018). Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. *Systematic Biology* **67**, 14–31.
- AFANADOR, N. L., SMOLINSKA, A., TRAN, T. N. & BLANCHET, L. (2016). Unsupervised random forest: a tutorial with case studies. *Journal of Chemometrics* **30**, 232–241.
- AMÉZQUITA, A., LIMA, A. P., JEHLER, R., CASTELLANOS, L., RAMOS, Ó., CRAWFORD, A. J., GASSER, H. & HÖDL, W. (2009). Calls, colours, shape, and genes: a multi-trait approach to the study of geographic variation in the Amazonian frog *Allobates femoralis*. *Biological Journal of the Linnean Society* **98**, 826–838.
- ANDERSSON, M. B. (1994). *Sexual Selection*. Princeton University Press, Princeton.
- DE ANDRADE, L. G., BISI, T. L. & AZEVEDO, A. D. F. (2017). Burst pulses of Guiana dolphin (*Sotalia guianensis*) in southeastern Brazil. *Journal of the Acoustical Society of America* **141**, 2947–2956.
- ANIKIN, A. (2019). Soundgen: an open-source tool for synthesizing nonverbal vocalizations. *Behavior Research Methods* **51**, 778–792.
- ARAYA-SALAS, M. & SMITH-VIDAURRE, G. (2017). warbleR: an r package to streamline analysis of animal acoustic signals. *Methods in Ecology and Evolution* **8**, 184–191.
- ARAYA-SALAS, M., SMITH-VIDAURRE, G. & WEBSTER, M. (2017). Assessing the effect of sound file compression and background noise on measures of acoustic signal structure. *Bioacoustics* **28**, 1–17.
- ARMITAGE, D. W. & OBER, H. K. (2010). A comparison of supervised learning techniques in the classification of bat echolocation calls. *Ecological Informatics* **5**, 465–473.
- ARNOLD, C. & NUNN, C. L. (2010). Phylogenetic targeting of research effort in evolutionary biology. *American Naturalist* **176**, 601–612.
- BAKER, J., MEADE, A., PAGEL, M. D. & VENDITTI, C. (2016). Positive phenotypic selection inferred from phylogenies. *Biological Journal of the Linnean Society* **118**, 95–115.
- BEECHER, M. D. (1988). Spectrographic analysis of animal vocalizations: implications of the “uncertainty principle”. *Bioacoustics* **1**, 187–208.
- BERLIN, S., TOMARAS, D. & CHARLESWORTH, B. (2007). Low mitochondrial variability in birds may indicate Hill-Robertson effects on the W chromosome. *Heredity* **99**, 389–396.
- BILLINGS, A. C. (2018). The low-frequency acoustic structure of mobbing calls differs across habitat types in three passerine families. *Animal Behaviour* **138**, e74.
- BjÖRKLUND, M. (2019). Be careful with your principal components. *Evolution* **73**, 2151–2158.
- BLOMBERG, S. P., GARLAND, T. & IVES, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**, 717–745.
- BLUMSTEIN, D. T. (2010). Golden-marmot alarm calls. II. Asymmetrical production and perception of situationally specific vocalizations? *Ethology* **101**, 25–32.
- BLUMSTEIN, D. T. & ARMITAGE, K. B. (1997). Does sociality drive the evolution of communicative complexity? A comparative test with ground-dwelling sciurid alarm calls. *American Naturalist* **150**, 179–200.
- BLUMSTEIN, D. T. & CHI, Y. Y. (2012). Scared and less noisy: glucocorticoids are associated with alarm call entropy. *Biology Letters* **8**, 189–192.
- BOETTIGER, C., COOP, G. & RALPH, P. (2012). Is your phylogeny informative? Measuring the power of comparative methods. *Evolution* **66**, 2240–2251.
- BONNEVIE, B. T. & CRAIG, A. J. F. K. (2018). A song for the South also: defining birdsong in global terms. *Ibis* **160**, 926–928.
- BOTERO, C. A., MUDGE, A. E., KOLTZ, A. M., HOCHACHKA, W. M. & VEHCAMP, S. L. (2008). How reliable are the methods for estimating repertoire size? *Ethology* **114**, 1227–1238.
- BRADBURY, J. W. & VEHCAMP, S. L. (2011). *Principles of Animal Communication*, 2nd Edition. Oxford University Press, Oxford.
- BROUGHTON, W. B. (1963). *Method in Bioacoustic Terminology*. Elsevier, Amsterdam.
- BROWN, C. & RIEDE, T. (eds) (2017). *Comparative Bioacoustics: An Overview*. Bentham Science Publishers, Sharjah.
- BRUMM, H. (2009). Song amplitude and body size in birds. *Behavioral Ecology and Sociobiology* **63**, 1157–1165.
- BRUMM, H., ZOLLINGER, S. A., NIEMELÄ, P. T. & SPRAU, P. (2017). Measurement artefacts lead to false positives in the study of birdsong in noise. *Methods in Ecology and Evolution* **8**, 1617–1625.
- BUDAEV, S. V. (2010). Using principal components and factor analysis in animal behaviour research: caveats and guidelines. *Ethology* **116**, 472–480.
- BUSNEL, R. G. (1963). *Acoustic Behaviour of Animals*. Elsevier, Amsterdam.
- BUTLER, M. A. & KING, A. A. (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *American Naturalist* **164**, 683–695.
- CAETANO, D. S. & AISENBERG, A. (2014). Forgotten treasures: the fate of data in animal behaviour studies. *Animal Behaviour* **98**, 1–5.
- CAETANO, D. S. & BEAULIEU, J. M. (2020). Comparative analyses of phenotypic sequences using phylogenetic trees. *American Naturalist* **195**, E38–E50.
- CALDER, W. A. (1990). The scaling of sound output and territory size: are they matched? *Ecology* **71**, 1810–1816.
- CARDOSO, G. C. (2010). Loudness of birdsong is related to the body size, syntax and phonology of passerine species. *Journal of Evolutionary Biology* **23**, 212–219.
- CARDOSO, G. C. (2012). Paradoxical calls: the opposite signaling role of sound frequency across bird species. *Behavioral Ecology* **23**, 237–241.
- CARDOSO, G. C. (2013). Using frequency ratios to study vocal communication. *Animal Behaviour* **85**, 1529–1532.
- CARDOSO, G. C. (2019). Double quantile regression accurately assesses distance to boundary trade-offs. *Methods in Ecology and Evolution* **10**, 1322–1331.
- CARDOSO, G. C. & ATWELL, J. W. (2011). On the relation between loudness and the increased song frequency of urban birds. *Animal Behaviour* **82**, 831–836.
- CARDOSO, G. C. & HU, Y. (2011). Birdsong performance and the evolution of simple (rather than elaborate) sexual signals. *American Naturalist* **178**, 679–686.

- CARDOSO, G. C. & MOTA, P. G. (2007). Song diversification and complexity in canaries and seedeaters (*Serinus* spp.). *Biological Journal of the Linnean Society* **92**, 183–194.
- CATALANO, S. A., GOLOBOFF, P. A. & GIANNINI, N. P. (2010). Phylogenetic morphometrics (I): the use of landmark data in a phylogenetic framework. *Cladistics* **26**, 539–549.
- CATCHPOLE, C. & SLATER, P. J. B. (2008). *Bird Song: Biological Themes and Variations*, 2nd Edition. Cambridge University Press, Cambridge.
- CHARIF, R. A., WAACK, A. M. & STRICKMAN, L. M. (2010). *Raven Pro 1.4 User's Manual*. The Cornell Lab of Ornithology, Ithaca.
- CHARLTON, B. D. (2015). The acoustic structure and information content of female koala vocal signals. *PLoS One* **10**, e0138670.
- CHARLTON, B. D., OWEN, M. A. & SWAISGOOD, R. R. (2019). Coevolution of vocal signal characteristics and hearing sensitivity in forest mammals. *Nature Communications* **10**, 2778.
- CHARLTON, B. D., WATCHORN, D. J. & WHISSON, D. A. (2017). Subharmonics increase the auditory impact of female koala rejection calls. *Ethology* **123**, 571–579.
- CHOLEWIAK, D. M., SOUSA-LIMA, R. S. & CERCHIO, S. (2013). Humpback whale song hierarchical structure: historical context and discussion of current classification issues. *Marine Mammal Science* **29**, E312–E332.
- CLEMIN, P. J. & JOHNSON, M. T. (2006). Generalized perceptual linear prediction features for animal vocalization analysis. *Journal of the Acoustical Society of America* **120**, 527–534.
- COWLING, M. & SITTE, R. (2003). Comparison of techniques for environmental sound recognition. *Pattern Recognition Letters* **24**, 2895–2907.
- CUNNINGHAM, C. W. (1999). Some limitations of ancestral character-state reconstruction when testing evolutionary hypotheses. *Systematic Biology* **48**, 665–674.
- CUNNINGHAM, C. W., OMLAND, K. E. & OAKLEY, T. H. (1998). Reconstructing ancestral character states: a critical reappraisal. *Trends in Ecology & Evolution* **13**, 361–366.
- CURRIE, T. E., GREENHILL, S. J. & MACE, R. (2010). Is horizontal transmission really a problem for phylogenetic comparative methods? A simulation study using continuous cultural traits. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 3903–3912.
- DALE, J., DEY, C. J., DELHEY, K., KEMPENAERS, B. & VALCU, M. (2015). The effects of life history and sexual selection on male and female plumage colouration. *Nature* **527**, 367–370.
- DALZIELL, A. H. (2012). *The Ecology of Vocal Mimicry in the Superb Lyrebird*, Menura novaehollandiae. PhD Dissertation. Australian National University.
- DALZIELL, A. H. & COCKBURN, A. (2008). Dawn song in superb fairy-wrens: a bird that seeks extrapair copulations during the dawn chorus. *Animal Behaviour* **75**, 489–500.
- DALZIELL, A. H., PETERS, R. A., COCKBURN, A., DORLAND, A. D., MAISEY, A. C. & MAGRATH, R. D. (2013). Dance choreography is coordinated with song repertoire in a complex avian display. *Current Biology* **23**, 1132–1135.
- DALZIELL, A. H., MAISEY, A. C., MAGRATH, R. D. & WELBERGEN, J. A. (2021). Male lyrebirds create an acoustic illusion of a mobbing flock during courtship and copulation. *Current Biology*. <https://doi.org/10.1016/j.cub.2021.02.003>
- DANCHIN, É., GIRALDEAU, L. A., VALONE, T. J. & WAGNER, R. H. (2004). Public information: from nosy neighbors to cultural evolution. *Science* **305**, 487–491.
- DANCHIN, É. & WAGNER, R. H. (2010). Inclusive heritability: combining genetic and non-genetic information to study animal behavior and culture. *Oikos* **119**, 210–218.
- DANTZKER, M. S., DEANE, G. B. & BRADBURY, J. W. (1999). Acoustics of the display of the sage grouse. *Journal of Experimental Biology* **202**, 2893–2909.
- DARCH, J., MILNER, B. & SHAO, X. (2004). *Formant Prediction from MFCC Vectors*. University of East Anglia, Norwich.
- DARLING, J. D. & SOUSA-LIMA, R. S. (2005). Songs indicate interaction between humpback whale (*Megaptera novaeangliae*) populations in the Western and Eastern South Atlantic Ocean. *Marine Mammal Science* **21**, 557–566.
- DAVIES, T. J., KRAFT, N. J. B., SALAMIN, N. & WOLKOVICH, E. M. (2012). Incompletely resolved phylogenetic trees inflate estimates of phylogenetic conservatism. *Ecology* **93**, 242–247.
- DEECKE, V. B., FORD, J. K. B. & SPONG, P. (1999). Quantifying complex patterns of bioacoustic variation: use of a neural network to compare killer whale (*Orcinus orca*) dialects. *Journal of the Acoustical Society of America* **105**, 2499–2507.
- DEECKE, V. B. & JANIK, V. M. (2005). Automated categorization of bioacoustic signals: avoiding perceptual pitfalls. *Journal of the Acoustical Society of America* **117**, 645–653.
- DELMORE, K. E., KENYON, H. L., GERMAIN, R. R. & IRWIN, D. E. (2015). Phenotypic divergence during speciation is inversely associated with differences in seasonal migration. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20151921.
- DELSUC, F., BRINKMANN, H. & PHILIPPE, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics* **6**, 361–375.
- DERRYBERRY, E. P., SEDDON, N., CLARAMUNT, S., TOBIAS, J. A., BAKER, A., ALEIXO, A. & BRUMFIELD, R. T. (2012). Correlated evolution of beak morphology and song in the neotropical woodcreeper radiation. *Evolution* **66**, 2784–2797.
- DEVOOGD, T. J., KREBS, J. R., HEALY, S. D. & PURVIS, A. (1993). Relations between song repertoire size and the volume of brain nuclei related to song: comparative evolutionary analyses amongst oscine birds. *Proceedings of the Royal Society B: Biological Sciences* **254**, 75–82.
- DOOLING, R. J. & PRIOR, N. H. (2017). Do we hear what birds hear in birdsong? *Animal Behaviour* **124**, 283–289.
- DOS REIS, M. & YANG, Z. (2013). The unbearable uncertainty of Bayesian divergence time estimation. *Journal of Systematics and Evolution* **51**, 30–43.
- DUNN, M., GREENHILL, S. J., LEVINSON, S. C. & GRAY, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* **473**, 79–82.
- ENDLER, J. A. (1992). Signals, signal conditions, and the direction of evolution. *American Naturalist* **139**, S125–S153.
- FARRIS, H. E. & RYAN, M. J. (2011). Relative comparisons of call parameters enable auditory grouping in frogs. *Nature Communications* **2**, 410.
- FELSENSTEIN, J. (1985). Phylogenies and the comparative method. *American Naturalist* **125**, 1–15.
- FILATOVA, O. A., FEDUTIN, I. D., BURDIN, A. M. & HOYT, E. (2007). The structure of the discrete call repertoire of killer whales *Orcinus orca* from southeast Kamchatka. *Bioacoustics* **16**, 261–280.
- FILATOVA, O. A., SAMARRA, F. I. P., DEECKE, V. B., FORD, J. K. B., MILLER, P. J. O. & YURK, H. (2015). Cultural evolution of killer whale calls: background, mechanisms and consequences. *Behaviour* **152**, 2001–2038.
- FITZJOHN, R. G. (2010). Quantitative traits and diversification. *Systematic Biology* **59**, 619–633.
- FITZJOHN, R. G. (2012). Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution* **3**, 1084–1092.
- FORTI, L. R., ZORNOSA-TORRES, C., MÁRQUEZ, R. & TOLEDO, L. F. (2018). Ancestral state, phylogenetic signal and convergence among anuran distress calls. *Zoologischer Anzeiger* **274**, 1–5.
- FRAILE, R. & GODINO-LLORENTE, J. I. (2014). Cepstral peak prominence: a comprehensive analysis. *Biomedical Signal Processing and Control* **14**, 42–54.
- FRECKLETON, R. P., COOPER, N. & JETZ, W. (2011). Comparative methods as a statistical fix: the dangers of ignoring an evolutionary model. *American Naturalist* **178**, E10–E17.
- FUENTES-G, J. A., HOUSWORTH, E. A., WEBER, A. & MARTINS, E. P. (2016). Phylogenetic ANCOVA: estimating changes in evolutionary rates as well as relationships between traits. *American Naturalist* **188**, 615–627.
- GARAMSZEGI, L. Z. (2014). *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*. Springer-Verlag, Berlin.
- GARAMSZEGI, L. Z. & GONZALEZ-VOYER, A. (2014). Working with the tree of life in comparative studies: how to build and tailor phylogenies to interspecific datasets. In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*, pp. 19–48. Springer-Verlag, Berlin.
- GARCÍA, N. C. & TUBARO, P. L. (2018). Dissecting the roles of body size and beak morphology in song evolution in the “blue” cardinalids (Passeriformes: Cardinalidae). *Auk* **135**, 262–275.
- GARLAND, T. JR. & IVES, A. R. (2000). Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *American Naturalist* **155**, 346–364.
- GARLAND, T. JR., BENNETT, A. F. & REZENDE, E. L. (2005). Phylogenetic approaches in comparative physiology. *Journal of Experimental Biology* **208**, 3015–3035.
- GEBERZAHN, N. & AUBIN, T. (2014). Assessing vocal performance in complex birdsong: a novel approach. *BMC Biology* **12**, 1–9.
- GERHARDT, H. C. (1998). Acoustic signals of animals: recording, field measurements, analysis and description. In *Animal Acoustic Communication* (eds S. L. HOPP, M. J. OWREN and C. S. EVANS), pp. 1–25. Springer, Berlin.
- GERHARDT, H. C. & HUBER, F. (2002). *Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions*. The University of Chicago Press, Chicago.
- GINGERICH, P. D. (2009). Rates of evolution. *Annual Review of Ecology, Evolution, and Systematics* **40**, 657–675.
- GINGRAS, B., MOHANDASAN, E., BOKO, D. & FITCH, W. T. (2013). Phylogenetic signal in the acoustic parameters of the advertisement calls of four clades of anurans. *BMC Evolutionary Biology* **13**, 134.
- GOICOECHEA, N., DE LA RIVA, I. & PADIAL, J. M. (2010). Recovering phylogenetic signal from frog mating calls. *Zoologica Scripta* **39**, 141–154.
- GONZALEZ-VOYER, A. & VON HARDENBERG, A. (2014). An introduction to phylogenetic path analysis. In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*, pp. 105–130. Springer-Verlag, Berlin.
- GONZALEZ-VOYER, A., DEN TEX, R.-J., CASTELLÓ, A. & LEONARD, J. A. (2013). Evolution of acoustic and visual signals in Asian barbets. *Journal of Evolutionary Biology* **26**, 647–659.
- GOOLSBY, E. W. (2015). Phylogenetic comparative methods for evaluating the evolutionary history of function-valued traits. *Systematic Biology* **64**, 568–578.
- GOUTTE, S., DUBOIS, A., HOWARD, S. D., MARQUEZ, R., ROWLEY, J. J. L., DEHLING, J. M., GRANDCOLAS, P., RONGCHUAN, X. & LEGENDRE, F. (2016). Environmental constraints and call evolution in torrent-dwelling frogs. *Evolution* **70**, 811–826.

- GRAFEN, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society B: Biological Sciences* **326**, 119–157.
- GRAY, R. D. & WATTS, J. (2017). Cultural macroevolution matters. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 7846–7852.
- GREG, E. I., PRICE, J. J. & PRUETT-JONES, S. (2013). Song evolution in Maluridae: influences of natural and sexual selection on acoustic structure. *Emu* **113**, 270–281.
- GREG, E. I. & WEBSTER, M. S. (2014). How do novel signals originate? The evolution of fairy-wren songs from predator to display contexts. *Animal Behaviour* **88**, 57–65.
- GRIESSER, M., DROBNIK, S. M., NAKAGAWA, S. & BOTERO, C. A. (2017). Family living sets the stage for cooperative breeding and ecological resilience in birds. *PLoS Biology* **15**, e2000483.
- HADFIELD, J. D. (2010a). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software* **33**, 1–22.
- HADFIELD, J. D. (2010b). MCMCglmm: Markov chain Monte Carlo methods for generalised linear mixed models. *Tutorial for MCMCglmm package in R*, 1–25.
- HADFIELD, J. D. & NAKAGAWA, S. (2010). General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology* **23**, 494–508.
- HANRAHAN, N. (2020) *The Acoustic Ecology of the Ghost Bat (Macroderma gigas): Form, Function and Applied Uses of Vocalisations*. PhD Dissertation, Western Sydney University.
- HANRAHAN, N., TURBILL, C., ARMSTRONG, K., DALZIELL, A. H. & WELBERGEN, J. A. (2021). Ghost bats exhibit informative daily and seasonal temporal patterns in the production of social vocalisations. *Australian Journal of Zoology*. Advance online publication. <https://doi.org/10.1071/ZO20055>
- HARMON, L. J. & GLOR, R. E. (2010). Poor statistical performance of the mantel test in phylogenetic comparative analyses. *Evolution* **64**, 2173–2178.
- HARMON, L. J., LOSOS, J. B., JONATHAN DAVIES, T., GILLESPIE, R. G., GITTLEMAN, J. L., BRYAN JENNINGS, W., KOZAK, K. H., MCPEEK, M. A., MORENO-ROARK, F., NEAR, T. J. & PURVIS, A. (2010). Early bursts of body size and shape evolution are rare in comparative data. *Evolution* **64**, 2385–2396.
- HARRIS, A. J., WILSON, D. R., GRAHAM, B. A. & MENNILL, D. J. (2016). Estimating repertoire size in a songbird: a comparison of three techniques. *Bioacoustics* **25**, 211–224.
- HARVEY, P. H. & PAGEL, M. D. (1991). *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford.
- HEATH, T. A., HEDTKE, S. M. & HILLIS, D. M. (2008). Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution* **46**, 239–257.
- HEDLEY, R. W., LOGUE, D. M., BENEDICT, L. & MENNILL, D. J. (2018). Assessing the similarity of song-type transitions among birds: evidence for interspecies variation. *Animal Behaviour* **140**, 161–170.
- HEDWIG, D., ROBBINS, M. M., MUNDRY, R., HAMMERSCHMIDT, K. & BOESCH, C. (2014). Acoustic structure and variation in mountain and western gorilla close calls: a syntactic approach. *Behaviour* **151**, 1091–1120.
- HEDWIG, D., VERAHRAMI, A. K. & WREGE, P. H. (2019). Acoustic structure of forest elephant rumbles: a test of the ambiguity reduction hypothesis. *Animal Cognition* **22**, 1115–1128.
- HERNÁNDEZ, C. E., RODRÍGUEZ-SERRANO, E., AVARIA-LLAUTUREO, J., INOSTROZA-MICHAEL, O., MORALES-PALLERO, B., BORIC-BARGETTO, D., CANALES-AGUIRRE, C. B., MARQUET, P. A. & MEADE, A. (2013). Using phylogenetic information and the comparative method to evaluate hypotheses in macroecology. *Methods in Ecology and Evolution* **4**, 401–415.
- HOPP, S. L., OWREN, M. J. & EVANS, C. S. (1998). *Animal Acoustic Communication: Sound Analysis and Research Methods*. Springer-Verlag, Berlin.
- HOPPITT, W. & LALAND, K. N. (2013). *Social Learning: An Introduction to Mechanisms, Methods, and Models*. Princeton University Press, Princeton.
- HOUSWORTH, E. A., MARTINS, E. P. & LYNCH, M. (2004). The phylogenetic mixed model. *American Naturalist* **163**, 84–96.
- HUELSENBECK, J. P. & BOLLBACK, J. P. (2001). Empirical and hierarchical bayesian estimation of ancestral states. *Systematic Biology* **50**, 351–366.
- IZENMAN, A. J. (2008). *Modern Multivariate Statistical Techniques*. Springer, New York.
- JANIK, V. M. & SLATER, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour* **60**, 1–11.
- JETZ, W., THOMAS, G. H., JOY, J. B., HARTMANN, K. & MOOERS, A. O. (2012). The global diversity of birds in space and time. *Nature* **491**, 444–448.
- JOHNSON, R. A. & WICHERN, D. W. (1982). *Applied Multivariate Statistical Analysis*. Prentice-Hall, Englewood Cliffs, NJ.
- JOY, J. B., LIANG, R. H., MCCLOSKEY, R. M., NGUYEN, T. & POON, A. F. Y. (2016). Ancestral reconstruction. *PLoS Computational Biology* **12**, e1004763.
- KAMILAR, J. M. & COOPER, N. (2013). Phylogenetic signal in primate behaviour, ecology and life history. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**, 20120341.
- KATAHIRA, K., SUZUKI, K., KAGAWA, H. & OKANOYA, K. (2013). A simple explanation for the evolution of complex song syntax in Bengalese finches. *Biology Letters* **9**, 20130842.
- KAUFMAN, A. B. & ROSENTHAL, R. (2009). Can you believe my eyes? The importance of interobserver reliability statistics in observations of animal behaviour. *Animal Behaviour* **78**, 1487–1491.
- KEEN, S., ROSS, J. C., GRIFFITHS, E. T., LANZONE, M. & FARNSWORTH, A. (2014). A comparison of similarity-based approaches in the classification of flight calls of four species of North American wood-warblers (Parulidae). *Ecological Informatics* **21**, 25–33.
- KEEN, S., SHIU, Y., WREGE, P. H. & ROWLAND, E. D. (2017). Automated detection of low-frequency rumbles of forest elephants: a critical tool for their conservation. *Journal of the Acoustical Society of America* **141**, 2715–2726.
- KERSHENBAUM, A. (2014). Entropy rate as a measure of animal vocal complexity. *Bioacoustics* **23**, 195–208.
- KERSHENBAUM, A., BLUMSTEIN, D. T., ROCH, M. A., AKÇAY, Ç., BACKUS, G., BEE, M. A., BOHN, K., CAO, Y., CARTER, G., CĂŞAR, C., COEN, M., DERUITER, S. L., DOYLE, L., EDELMAN, S., FERRER-I-CANCHO, R., et al. (2016). Acoustic sequences in non-human animals: a tutorial review and prospectus. *Biological Reviews* **91**, 13–52.
- KERSHENBAUM, A., FREEBERG, T. M. & GAMMON, D. E. (2015). Estimating vocal repertoire size is like collecting coupons: a theoretical framework with heterogeneity in signal abundance. *Journal of Theoretical Biology* **373**, 1–11.
- KOGAN, J. A. & MARGOLIASH, D. (1998). Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: a comparative study. *Journal of the Acoustical Society of America* **103**, 2185–2196.
- KÖHLER, J., JANSEN, M., RODRÍGUEZ, A., KOK, P. J. R., TOLEDO, L. F., EMMRICH, M., GLAW, F., HADDAD, C. F. B., RÖDEL, M. O. & VENCES, M. (2017). The use of bioacoustics in anuran taxonomy: theory, terminology, methods and recommendations for best practice. *Zootaxa* **4251**, 1–124.
- KUHL, P. K. (1981). Discrimination of speech by nonhuman animals: basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America* **70**, 340–349.
- LACHLAN, R. F. (2007). Luscinia: a bioacoustics analysis computer program. <https://sourceforge.net/projects/luscinia/>
- LACHLAN, R. F., RATMANN, O. & NOWICKI, S. (2018). Cultural conformity generates extremely stable traditions in bird song. *Nature Communications* **9**, 2417.
- LACHLAN, R. F., VERZIJDEN, M. N., BERNARD, C. S., JONKER, P.-P., KOESE, B., JAARMA, S., SPOOR, W., SLATER, P. J. B. & TEN CATE, C. (2013). The progressive loss of syntactical structure in bird song along an island colonization chain. *Current Biology* **23**, 1896–1901.
- LAUDER, G. V. (1986). *Homology, Analogy, and the Evolution of Behavior*. Oxford University Press, New York.
- LEGENDRE, P. & LEGENDRE, L. (2012). *Numerical Ecology*. Elsevier, Amsterdam.
- LEIGHTON, G. M. (2017). Cooperative breeding influences the number and type of vocalizations in avian lineages. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20171508.
- LEIPPERT, D. (1994). Social behaviour on the wing in the False Vampire, *Megaderma byra*. *Ethology* **98**, 111–127.
- LI, G., STEEL, M. & ZHANG, L. (2008). More taxa are not necessarily better for the reconstruction of ancestral character states. *Systematic Biology* **57**, 647–653.
- LIGGES, U., KREY, S., MERSMANN, O. & SCHNACKENBERG, S. (2018). *tuneR: Analysis of Music and Speech*. <https://CRAN.R-project.org/package=tuneR>
- LIGON, R. A., DIAZ, C. D., MORANO, J. L., TROSCIANKO, J., STEVENS, M., MOSKELAND, A., LAMAN, T. G. & SCHOLES, E. (2018). Evolution of correlated complexity in the radically different courtship signals of birds-of-paradise. *PLoS Biology* **16**, e2006962.
- LOSOS, J. B. (1999). Uncertainty in the reconstruction of ancestral character states and limitations on the use of phylogenetic comparative methods. *Animal Behaviour* **58**, 1319–1324.
- LOSOS, J. B. (2011). Seeing the forest for the trees: the limitations of phylogenies in comparative biology. *American Naturalist* **177**, 709–727.
- LOUGA, S. & PENNELL, M. W. (2020). Extant timetrees are consistent with a myriad of diversification histories. *Nature* **580**, 502–505.
- LUTTRELL, S. A. M., GALLAGHER, M. E. & LOHR, B. (2016). Multiple estimation methods suggest similar repertoire sizes for Gulf Coast and eastern marsh wrens with no correlation between repertoire size and migratory distance. *Behaviour* **153**, 287–311.
- LYON, R. H. & ORDUBADI, A. (1982). Use of cepstra in acoustical signal analysis. *Journal of Mechanical Design* **104**, 303–306.
- MACE, R. & HOLDEN, C. J. (2005). A phylogenetic approach to cultural evolution. *Trends in Ecology & Evolution* **20**, 116–121.
- MADDISON, W. & COLU, B. (2015). Parsimony, character-state reconstructions, and evolutionary inferences. In *Systematics, Historical Ecology, and North American Freshwater Fishes*.
- MADDISON, W. P., MIDFORD, P. E. & OTTO, S. P. (2007). Estimating a binary character's effect on speciation and extinction. *Systematic Biology* **56**, 701–710.
- MADSEN, P. T. & WAHLBERG, M. (2007). Recording and quantification of ultrasonic echolocation clicks from free-ranging toothed whales. *Deep Sea Research Part I: Oceanographic Research Papers* **54**, 1421–1444.

- MARLER, P. (1961). The logical analysis of animal communication. *Journal of Theoretical Biology* **1**, 295–317.
- MARLER, P. (1967). Animal communication signals. *Science* **157**, 769–774.
- MARLER, P. & SLABBEKOORN, H. (2004). *Nature's Music: The Science of Birdsong*. Elsevier Academic Press, San Diego.
- MARSHALL, C. R. (2017). Five palaeobiological laws needed to understand the evolution of the living biota. *Nature Ecology & Evolution* **1**, 0165.
- MARTINS, E. P. & GARLAND, T. (1991). Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. *Evolution* **45**, 534.
- MARTINS, E. P. & HANSEN, T. F. (1997). Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist* **149**, 646–667.
- MASON, N. A. & BURNS, K. J. (2015). The effect of habitat and body size on the evolution of vocal displays in Thraupidae (tanagers), the largest family of songbirds. *Biological Journal of the Linnean Society* **114**, 538–551.
- MASON, N. A., BURNS, K. J., TOBIAS, J. A., CLARAMUNT, S., SEDDON, N. & DERRYBERRY, E. P. (2017a). Song evolution, speciation, and vocal learning in passerine birds. *Evolution* **1788**, 786–796.
- MASON, N. A., PASCH, B., BURNS, K. J. & DERRYBERRY, E. P. (2017b). Integrating museum and media collections to study vocal ecology and evolution. In *The Extended Specimen: Emerging Frontiers in Collections-Based Ornithological Research*. CRC Press, Boca Raton.
- MATTHEWS, J. N., RENDELL, L. E., GORDON, J. C. D. & MACDONALD, D. W. (2012). A review of frequency and time parameters of cetacean tonal calls. *Bioacoustics* **10**, 47–71.
- MAY-COLLADO, L. J., AGNARSSON, I. & WARTZOK, D. (2007). Phylogenetic review of tonal sound production in whales in relation to sociality. *BMC Evolutionary Biology* **7**, 136.
- MCCOWAN, B., HANSER, S. F. & DOYLE, L. R. (2002). Using information theory to assess the diversity, complexity, and development of communicative repertoires. *Journal of Comparative Psychology* **116**, 166.
- MEDINA-GARCÍA, A., ARAYA-SALAS, M. & WRIGHT, T. F. (2015). Does vocal learning accelerate acoustic diversification? Evolution of contact calls in neotropical parrots. *Journal of Evolutionary Biology* **28**, 1782–1792.
- MELIZA, C. D., KEEN, S. C. & RUBENSTEIN, D. R. (2013). Pitch- and spectral-based dynamic time warping methods for comparing field recordings of harmonic avian vocalizations. *Journal of the Acoustical Society of America* **134**, 1407–1415.
- MELLINGER, D. K. & BRADBURY, J. W. (2007). Acoustic measurement of marine mammal sounds in noisy environments. *Proceedings of the Second International Conference on Underwater Acoustic Measurements: Technologies and Results, Heraklion, Greece, 25–29 June 2007*, June, 8–8.
- MESOUDI, A. (2016). Cultural evolution: a review of theory, findings and controversies. *Evolutionary Biology* **43**, 481–497.
- MESOUDI, A. (2017). Pursuing Darwin's curious parallel: prospects for a science of cultural evolution. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 7853–7860.
- MIKULA, P., PETRUSKOVÁ, T. & ALBRECHT, T. (2018). Song complexity—no correlation between standard deviation of frequency and traditionally used song complexity metrics in passerines: a comment on Pearse et al. (2018). *Evolution* **72**, 2832–2835.
- MILLER, E. T., LEIGHTON, G. M., FREEMAN, B. G., LEES, A. C. & LIGON, R. A. (2019). Ecological and geographical overlap drive plumage evolution and mimicry in woodpeckers. *Nature Communications* **10**, 1602.
- MOOERS, A. Ø. & SCHLUTER, D. (1999). Reconstructing ancestor states with maximum likelihood: support for one- and two-rate models. *Systematic Biology* **48**, 623–633.
- MORFEY, C. L. (2001). *Dictionary of Acoustics*. Academic Press, San Diego.
- MOTA, P. G. & CARDOSO, G. C. (2001). Song organisation and patterns of variation in the serin (*Serinus serinus*). *Acta Ethologica* **3**, 141–150.
- MÜNKEMÜLLER, T., LAVERGNE, S., BZEZNIK, B., DRAY, S., JOMBART, T., SCHIFFERS, K. & THULLER, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution* **3**, 743–756.
- NARINS, P. & CAPRANICA, R. (1976). Sexual differences in the auditory system of the tree frog *Eleutherodactylus coqui*. *Science* **192**, 378–380.
- NELSON, D. A. & MARLER, P. (1990). *The Perception of Birdsong and An Ecological Concept of Signal Space*. John Wiley & Sons, Oxford.
- NUNN, C. L. (2011). *The Comparative Approach in Evolutionary Anthropology and Biology*. University of Chicago Press, Chicago.
- O'MEARA, B. C., ANÉ, C., SANDERSON, M. J. & WAINWRIGHT, P. C. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution* **60**, 922.
- OAKLEY, T. H. & CUNNINGHAM, C. W. (2000). Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny. *Evolution* **54**, 397–405.
- ODOM, K. J., HALL, M. L., RIEBEL, K., OMLAND, K. E. & LANGMORE, N. E. (2014). Female song is widespread and ancestral in songbirds. *Nature Communications* **5**, 3379.
- ODOM, K. J., OMLAND, K. E. & PRICE, J. J. (2015). Differentiating the evolution of female song and male-female duets in the New World blackbirds: can tropical natural history traits explain duet evolution? *Evolution* **69**, 839–847.
- OMLAND, K. E. (1999). The assumptions and challenges of ancestral state reconstructions. *Systematic Biology* **48**, 604–611.
- OPPENHEIM, A. V. & SCHAFFER, R. (2009). *Discrete-Time Signal Processing*, 3rd Edition. Pearson, London.
- OWREN, M. J., RENDALL, D. & RYAN, M. J. (2010). Redefining animal signaling: influence versus information in communication. *Biology & Philosophy* **25**, 755–780.
- PÄCKERT, M., MARTENS, J., KOSUCH, J., NAZARENKO, A. A. & VEITH, M. (2003). Phylogenetic signal in the song of crests and kinglets (Aves: *Regulus*). *Evolution* **57**, 626–629.
- PAGEL, M. D. (1994). Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society B: Biological Sciences* **255**, 37–45.
- PAGEL, M. D. (1999a). Inferring the historical patterns of biological evolution. *Nature* **401**, 877.
- PAGEL, M. D. (1999b). The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology* **48**, 612–622.
- PAGEL, M. D. (2009). Human language as a culturally transmitted replicator. *Nature Reviews Genetics* **10**, 405–415.
- PAGEL, M. D., MEADE, A. & BARKER, D. (2004). Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology* **53**, 673–684.
- PARADIS, E. (2011). Time-dependent speciation and extinction from phylogenies: a least squares approach. *Evolution* **65**, 661–672.
- PARADIS, E. (2013). Molecular dating of phylogenies by likelihood methods: a comparison of models and a new information criterion. *Molecular Phylogenetics and Evolution* **67**, 436–444.
- PARADIS, E. (2014). An introduction to the phylogenetic comparative method. In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*, pp. 3–18. Springer-Verlag, Berlin.
- PARRA-HERNÁNDEZ, R. M., POSADA-QUINTERO, J. I., ACEVEDO-CHARRY, O. & POSADA-QUINTERO, H. F. (2020). Uniform manifold approximation and projection for clustering taxa through vocalizations in a neotropical passerine (rough-legged tyrannulet, *Phyllomyias burmeisteri*). *Animals* **10**, 1406–1419.
- PATRICELLI, G. L., DANTZKER, M. S. & BRADBURY, J. W. (2007). Differences in acoustic directionality among vocalizations of the male red-winged blackbird (*Agelaius phoeniceus*) are related to function in communication. *Behavioral Ecology and Sociobiology* **61**, 1099–1110.
- PEARSE, W. D., MORALES-CASTILLA, I., JAMES, L. S., FARRELL, M., BOIVIN, F. & DAVIES, T. J. (2018). Global macroevolution and macroecology of passerine song. *Evolution* **72**, 944–960.
- PESHEK, K. R. & BLUMSTEIN, D. T. (2011). Can rarefaction be used to estimate song repertoire size in birds? *Current Zoology* **57**, 300–306.
- PICONE, J. W. (1993). Signal modeling techniques in speech recognition. *Proceedings of the IEEE* **81**, 1215–1247.
- PODOS, J. (1996). Motor constraints on vocal development in a songbird. *Animal Behaviour* **51**, 1061–1070.
- PODOS, J. (1997). A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). *Evolution* **51**, 537–551.
- PODOS, J. (2001). Correlated evolution of morphology and vocal structure in Darwin's finches. *Nature* **409**, 185–188.
- PODOS, J., HUBER, S. K. & TAFT, B. (2004). Bird song: the interface of evolution and mechanism. *Annual Review of Ecology, Evolution, and Systematics* **35**, 55–87.
- PODOS, J., MOSELEY, D. L., GOODWIN, S. E., MCCLURE, J., TAFT, B. N., STRAUSS, A. V. H., REGA-BRODSKY, C. & LAHTI, D. C. (2016). A fine-scale, broadly applicable index of vocal performance: frequency expansion. *Animal Behaviour* **116**, 203–212.
- PRICE, J. J., EARNSHAW, S. M. & WEBSTER, M. S. (2006). Montezuma oropendolas modify a component of song constrained by body size during vocal contests. *Animal Behaviour* **71**, 799–807.
- PRICE, J. J., FRIEDMAN, N. R. & OMLAND, K. E. (2007). Song and plumage evolution in the New World orioles (*Icterus*) show similar lability and convergence in patterns. *Evolution* **61**, 850–863.
- PRICE, J. J. & LANYON, S. M. (2002). Reconstructing the evolution of complex song in the oropendolas. *Evolution* **56**, 1514–1529.
- QUENTAL, T. B. & MARSHALL, C. R. (2010). Diversity dynamics: molecular phylogenies need the fossil record. *Trends in Ecology & Evolution* **25**, 434–441.
- RABOSKY, D. L., CHANG, J., TITLE, P. O., COWMAN, P. F., SALLAN, L., FRIEDMAN, M., KASCHNER, K., GARILAO, C., NEAR, T. J., COLL, M. & ALFARO, M. E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature* **559**, 392–395.
- RABOSKY, D. L., GRUNDLER, M., ANDERSON, C., TITLE, P., SHI, J. J., BROWN, J. W., HUANG, H. & LARSON, J. G. (2014). BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution* **5**, 701–707.
- RABOSKY, D. L., SANTINI, F., EASTMAN, J., SMITH, S. A., SIDLAUSKAS, B., CHANG, J. & ALFARO, M. E. (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications* **4**, 1958.
- RAMASUBRAMANIAN, K. & SINGH, A. (2016). *Machine Learning Using R: With Time Series and Industry-Based Use Cases in R*, 2nd ed. Edition. Apress, Berkeley.

- RAND, A. S. & RYAN, M. J. (1987). Female responses to ancestral advertisement calls in tungara frogs. *Science* **269**, 390–392.
- READ, A. F. & WEARY, D. M. (1992). The evolution of bird song: comparative analyses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **338**, 165–187.
- REN, Y., JOHNSON, M. T., CLEMIN, P. J., DARRE, M., GLAESER, S. S., OSIEJUK, T. S. & OUT-NYARKO, E. (2009). A framework for bioacoustic vocalization analysis using hidden Markov models. *Algorithms* **2**, 1410–1428.
- RENDELL, L., MATTHEWS, J., GILL, A., GORDON, J. & MACDONALD, D. (1999). Quantitative analysis of tonal calls from five odontocete species, examining interspecific and intraspecific variation. *Journal of Zoology* **249**, 403–410.
- REVELL, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217–223.
- REVELL, L. J. (2013). Two new graphical methods for mapping trait evolution on phylogenies. *Methods in Ecology and Evolution* **4**, 754–759.
- REVELL, L. J. (2014). Ancestral character estimation under the threshold model from quantitative genetics. *Evolution* **68**, 743–759.
- REVELL, L. J., HARMON, L. J. & COLLAR, D. C. (2008). Phylogenetic Signal, evolutionary process, and rate. *Systematic Biology* **57**, 591–601.
- RICE, A. N. & BASS, A. H. (2009). Novel vocal repertoire and paired swimbladders of the three-spined toadfish, *Batrachomoeus trispinosus*: insights into the diversity of the Batrachoididae. *Journal of Experimental Biology* **212**, 1377–1391.
- RICHARDS, D. G. & WILEY, R. H. (1980). Reverberations and amplitude fluctuations in the propagation of sound in a forest: implications for animal communication. *American Naturalist* **115**, 381–399.
- RIEDE, T., ELIASON, C. M., MILLER, E. H., GOLLER, F. & CLARKE, J. A. (2016). Coos, booms, and hoots: the evolution of closed-mouth vocal behavior in birds. *Evolution* **70**, 1734–1746.
- RIEDE, T. & GOLLER, F. (2014). Morphological basis for the evolution of acoustic diversity in oscine songbirds. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20132306.
- RÍOS-CHELÉN, A. A., McDONALD, A. N., BERGER, A., PERRY, A. C., KRAKAUER, A. H. & PATRICELLI, G. L. (2017). Do birds vocalize at higher pitch in noise, or is it a matter of measurement? *Behavioral Ecology and Sociobiology* **71**, 29.
- ROSENBERG, M. S. & KUMAR, S. (2001). Incomplete taxon sampling is not a problem for phylogenetic inference. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 10751–10756.
- ROYER-CARENZI, M., PONTAROTTI, P. & DIDIER, G. (2013). Choosing the best ancestral character state reconstruction method. *Mathematical Biosciences* **242**, 95–109.
- RUBOLINI, D., LIKER, A., GARAMSZEGI, L. Z., MÖLLER, A. P. & SAINO, N. (2015). Using the BirdTree.org website to obtain robust phylogenies for avian comparative studies: a primer. *Current Zoology* **61**, 959–965.
- RUSO, D., ANCILLOTTI, L. & JONES, G. (2018). Bats are still not birds in the digital era: echolocation call variation and why it matters for bat species identification. *Canadian Journal of Zoology* **96**, 63–78.
- RUSO, D. & VOIGT, C. C. (2016). The use of automated identification of bat echolocation calls in acoustic monitoring: a cautionary note for a sound analysis. *Ecological Indicators* **66**, 598–602.
- SAKATA, J. T. & VEHRENCAMP, S. L. (2012). Integrating perspectives on vocal performance and consistency. *Journal of Experimental Biology* **215**, 201–209.
- SALISBURY, B. A. & KIM, J. (2001). Ancestral state estimation and taxon sampling density. *Systematic Biology* **50**, 557–564.
- SANDSTEN, M., GROSSE RUSE, M. & JÖNSSON, M. (2016). Robust feature representation for classification of bird song syllables. *EURASIP Journal on Advances in Signal Processing* **2016**, 68.
- SAUQUET, H., VON BALTHAZAR, M., MAGALLÓN, S., DOYLE, J. A., ENDRESS, P. K., BAILES, E. J., BARROSO DE MORAIS, E., BULL-HEREÑU, K., CARRIVE, L., CHARTIER, M., CHOMICIKI, G., COIRO, M., CORNETTE, R., EL OTTRA, J. H. L., EPICOCO, C., et al. (2017). The ancestral flower of angiosperms and its early diversification. *Nature Communications* **8**, 16047.
- SCHULTZ, T. R. & CHURCHILL, G. A. (1999). The role of subjectivity in reconstructing ancestral character states: a Bayesian approach to unknown rates, states, and transformation asymmetries. *Systematic Biology* **48**, 651–664.
- SHELLEY, E. L. & BLUMSTEIN, D. T. (2005). The evolution of vocal alarm communication in rodents. *Behavioral Ecology* **16**, 169–177.
- SINGH, P. & PRICE, T. D. (2015). Causes of the latitudinal gradient in birdsong complexity assessed from geographical variation within two Himalayan warbler species. *Ibis* **157**, 511–527.
- SMOTHERMAN, M., KNÖRNSCHILD, M., SMARSH, G. & BOHN, K. (2016). The origins and diversity of bat songs. *Journal of Comparative Physiology A* **202**, 535–554.
- SNYDER, K. T. & CREANZA, N. (2019). Polygyny is linked to accelerated birdsong evolution but not to larger song repertoires. *Nature Communications* **10**, 3884.
- SPECTOR, D. A. (1994). Definition in biology: the case of 'bird song'. *Journal of Theoretical Biology* **168**, 373–381.
- STODDARD, M. C. & PRUM, R. O. (2008). Evolution of avian plumage color in a tetrahedral color space: a phylogenetic analysis of new world buntings. *American Naturalist* **171**, 755–776.
- SUEUR, J. (2018). *Sound Analysis and Synthesis with R*. Springer International Publishing, Cham.
- SULLIVAN, J. & JOYCE, P. (2005). Model selection in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* **36**, 445–466.
- SUTHERS, R. A., FITCH, W. T., FAY, R. R. & POPPER, A. N. (2016). *Vertebrate Sound Production and Acoustic Communication*. Springer Handbook of Auditory Research, Heidelberg.
- SYMONDS, M. R. E. & BLOMBERG, S. P. (2014). A primer on phylogenetic generalised least squares. In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*, pp. 105–130. Springer-Verlag, Berlin.
- TARVER, J. E. & DONOGHUE, P. C. J. (2011). The trouble with topology: phylogenies without fossils provide a revisionist perspective of evolutionary history in topological analyses of diversity. *Systematic Biology* **60**, 700–712.
- TAYLOR, A. M., CHARLTON, B. D. & REBY, D. (2016). Vocal production by terrestrial mammals: source, filter, and function. In *Vertebrate Sound Production and Acoustic Communication*, pp. 229–259. Springer, Cham.
- TCHERNICHOVSKI, O., NOTTEBOHM, F., HO, C. E., PESARAN, B. & MITRA, P. P. (2000). A procedure for an automated measurement of song similarity. *Animal Behaviour* **59**, 1167–1176.
- TEN CATE, C., LACHLAN, R. F. & ZUIDEMA, W. (2013). Analyzing the structure of bird vocalizations and language: finding common ground. In *Birdsong, Speech, and Language: Exploring the Evolution of Mind and Brain*. The MIT Press, Cambridge.
- THARWAT, A., GABER, T., IBRAHIM, A. & HASSANIEN, A. E. (2017). Linear discriminant analysis: a detailed tutorial. *AI Communications* **30**, 169–190.
- THOMPSON, N. S., LEDOUX, K. & MOODY, K. (1994). A system for describing bird song units. *Bioacoustics* **5**, 267–279.
- TITLE, P. O. & RABOSKY, D. L. (2019). Tip rates, phylogenies and diversification: what are we estimating, and how good are the estimates? *Methods in Ecology and Evolution* **10**, 821–834.
- TOBIAS, J. A., PLANQUE, R., CRAM, D. L. & SEDDON, N. (2014). Species interactions and the structure of complex communication networks. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 1020–1025.
- TOBIAS, J. A., SHEARD, C., SEDDON, N., MEADE, A., COTTON, A. J. & NAKAGAWA, S. (2016). Territoriality, social bonds, and the evolution of communal signaling in birds. *Frontiers in Ecology and Evolution* **4**, 74.
- TODT, D. & HULTSCH, H. (1998). How songbirds deal with large amounts of serial information: retrieval rules suggest a hierarchical song memory. *Biological Cybernetics* **79**, 487–500.
- TOKUDA, I., RIEDE, T., NEUBAUER, J., OWREN, M. J. & HERZEL, H. (2002). Nonlinear analysis of irregular animal vocalizations. *Journal of the Acoustical Society of America* **111**, 2908–2919.
- TYSON, R. B., NOWACEK, D. P. & MILLER, P. J. O. (2007). Nonlinear phenomena in the vocalizations of North Atlantic right whales (*Eubalaena glacialis*) and killer whales (*Orcinus orca*). *Journal of the Acoustical Society of America* **122**, 1365–1373.
- UYEDA, J. C., CAETANO, D. S. & PENNELL, M. W. (2015). Comparative analysis of principal components can be misleading. *Systematic Biology* **64**, 677–689.
- WADEWITZ, P., HAMMERSCHMIDT, K., BATTAGLIA, D., WITT, A., WOLF, F. & FISCHER, J. (2015). Characterizing vocal repertoires—hard vs. soft classification approaches. *PLoS One* **10**, e0125785.
- WANG, X., MUEEN, A., DING, H., TRAJCEVSKI, G., SCHEUERMANN, P. & KEOGH, E. (2013). Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery* **26**, 275–309.
- WEIR, J. T. & LAWSON, A. (2015). Evolutionary rates across gradients. *Methods in Ecology and Evolution* **6**, 1278–1286.
- WEIR, J. T. & WHEATCROFT, D. (2011). A latitudinal gradient in rates of evolution of avian syllable diversity and song length. *Proceedings of the Royal Society B: Biological Sciences* **278**, 1713–1720.
- WEIR, J. T., WHEATCROFT, D. J. & PRICE, T. D. (2012). The role of ecological constraint in driving the evolution of avian song frequency across a latitudinal gradient. *Evolution* **66**, 2773–2783.
- WILDEN, I., TEMBROCK, G., HERZEL, H. & PETERS, G. (1998). Subharmonics, biphonation, and deterministic chaos in mammal vocalization. *Bioacoustics* **9**, 171–196.
- WILEY, R. H. & RICHARDS, D. G. (1982). Adaptations for acoustic communication in birds: Sound transmission and signal detection. *Acoustic Communication in Birds*. **vol. 1**, New York: Academic Press, 131–181.
- WILMAN, H., BELMAKER, J., SIMPSON, J., DE LA ROSA, C., RIVADENEIRA, M. M. & JETZ, W. (2014). EltonTraits 1.0: species-level foraging attributes of the world's birds and mammals. *Ecology* **95**, 2027–2027.
- WILSON, D. R., BITTON, P. P., PODO, J. & MENNILL, D. J. (2014). Uneven sampling and the analysis of vocal performance constraints. *American Naturalist* **183**, 214–228.
- ZOLLINGER, S. A., PODO, J., NEMETH, E., GOLLER, F. & BRUMM, H. (2012). On the relationship between, and measurement of, amplitude and frequency in birdsong. *Animal Behaviour* **84**, e1–e9.
- ZOLLINGER, S. A., RIEDE, T. & SUTHERS, R. A. (2003). Nonlinear phenomena in northern mockingbird (*Mimus polyglottos*) vocalizations: acoustics and physiology. *Journal of the Acoustical Society of America* **113**, 2305–2305.

VII. Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Examples of useful acoustic metrics and measurement approaches for comparative studies of vertebrate vocal behaviour.

Appendix S1. Some principles, practices, and pitfalls of bioacoustic recording and measurement.

(Received 17 December 2019; revised 3 February 2021; accepted 5 February 2021)