

# Understanding sources of variance and correlation among features of Bornean gibbon (*Hylobates muelleri*) female calls

Dena J. Clink,<sup>1,a)</sup> Mark N. Grote,<sup>1</sup> Margaret C. Crofoot,<sup>1,b)</sup> and Andrew J. Marshall<sup>2</sup>

<sup>1</sup>Department of Anthropology, University of California, Davis, One Shields Avenue, Davis, California 95616, USA

<sup>2</sup>Department of Anthropology, Program in the Environment, Department of Ecology and Evolutionary Biology, and School for Environment and Sustainability, University of Michigan, 101 West Hall, 1085 South University Avenue, Ann Arbor, Michigan 48109, USA

Acoustic signals serve important functions in mate choice, resource defense, and species recognition. Quantifying patterns and sources of variation in acoustic signals can advance understanding of the evolutionary processes that shape behavioral diversity more broadly. Animal vocalization datasets are inherently multivariate and hierarchical, wherein multiple features are estimated from calls of many individuals across different recording locations. Patterns of variation within different hierarchical levels—notwithstanding the challenges they present for modeling and inference—can provide insight into processes shaping vocal variation. The current work presents a multivariate, variance components model to investigate three levels of variance (within-female, between-female, and between-site) in Bornean gibbon calls. For six of the eight features estimated from call spectrograms, between-female variance was the most important contributor to total variance. For one feature, trill rate, there were site-level differences, which may be related to geographic isolation of certain gibbon populations. There was also a negative relationship between trill rate and duration of the introduction, suggesting trade-offs in the production of gibbon calls. Given substantial inter-individual variation in gibbon calls, it seems likely that there has been selection to confer information regarding caller identity, but mechanisms leading to site-level variation in trill rate remain to be determined.

## I. INTRODUCTION

Quantifying the sources of variation in animal vocalizations is of broad interest to ecologists and evolutionary biologists, due to the important role vocalizations play in a diverse set of biological and behavioral processes such as mate choice, resource defense, and species recognition (Wilkins *et al.*, 2012). Animal vocal communication systems are ideal for testing hypotheses related to the evolution and maintenance of behavioral diversity, as acoustic data are often easier to collect than other types of behavioral data. Like other types of behavioral data, acoustic signals may vary across sites or populations, individuals may exhibit distinct call features, and signals may vary across different temporal scales or in response to different social conditions. Understanding how call features vary across sites, individuals and contexts is a crucial first step for testing hypotheses related to the evolution of call diversity.

Geographic variation in vocalizations has been documented across a wide range of taxa including birds (greenish warblers, *Phylloscopus trochiloides*, Irwin *et al.*, 2008; Amazon parrots, *Amazona auropalliata*, Wright *et al.*, 2001), bats (greater horseshoe bats, *Rhinolophus ferrumequinum*,

Sun *et al.*, 2013), and mice (e.g., singing mice, *Scotinomys teguina*, Campbell *et al.*, 2010). In non-human primates, evidence for geographic variation in vocalizations has been found in tarsiers (*Tarsius tarsier*; Burton and Nietsch, 2010), pygmy marmosets (*Cebuella pygmaea*; de la Torre and Snowdon, 2009), Thomas langurs (*Presbytis thomasi*; Wich *et al.*, 2008), gibbons (*Hylobates spp.*; Dallmann and Geissmann, 2009; Heller *et al.*, 2010), orangutans (*Pongo spp.*; Delgado, 2007), and chimpanzees (*Pan troglodytes*; Mitani *et al.*, 1999). Differences in ecology may lead to substantial variation in acoustic parameters between geographically distinct populations (the acoustic adaptation hypothesis; Morton, 1975), but such patterns can also arise through the accumulation of random changes in call structure over time (i.e., behavioral drift, Yurk *et al.*, 2002). Directional and random evolutionary processes need not be mutually exclusive. It is reasonable to predict that certain aspects of calls, such as the spectral properties or note repetition rate, will be influenced by the environment (Boncoraglio and Saino, 2007). For example, forest habitats have greater reverberation than open habitats, and higher frequencies attenuate more quickly in forested environments (Marten and Marler, 1977), which may select for calls with fewer repeated elements and lower frequencies. Other aspects of the calls, such as note duration or sequential order of call types, may be more influenced by drift (e.g., Slabbekoorn and Smith, 2002).

In addition to population-level variation in call structure, substantial between-individual variation is expected

<sup>a)</sup>Also at: Bioacoustics Research Program, Cornell Laboratory of Ornithology, Cornell University, Ithaca, NY 14850, USA. Electronic mail: [djc426@cornell.edu](mailto:djc426@cornell.edu)

<sup>b)</sup>Also at: Smithsonian Tropical Research Institute, Balboa, Ancon, Panama.

when calls transmit information regarding individual identity (Tibbetts and Dale, 2007). Vocal individuality has been documented in a variety of nonhuman primates, including mouse lemurs (*Microcebus murinus*; Leliveld *et al.*, 2011), ring-tailed lemurs (*Lemur catta*; Macedonia, 1986), tarsiers (*Tarsius syrichta*; Řeháková-Petrů *et al.*, 2012), squirrel monkeys (*Saimiri sciureus*; Symmes *et al.*, 1979), baboons (*Papio cynocephalus*; Fischer *et al.*, 2002; Semple, 2001), Japanese macaques (*Macaca fuscata*; Ceugniet and Izumi, 2004), gibbons (*Hylobates spp.*; Haimoff and Gittins, 1985; Haimoff and Tilson, 1985; Dallmann and Geissmann, 2009; Terleph *et al.*, 2015; Clink *et al.*, 2017a), orangutans (*Pongo pygmaeus*; Lameira and Wich, 2008), gorillas (*Gorilla gorilla*; Salmi *et al.*, 2014), and chimpanzees (*Pan troglodytes*; Mitani *et al.*, 1996). Indeed, most vocal animals that have been examined exhibit vocal individuality, and there is evidence via playback experiments that primates can distinguish between individuals solely on the basis of their vocalizations (Cheney and Seyfarth, 1980, 1999; Rendall *et al.*, 1996). Importantly, strong patterns of variation at the individual-level do not preclude variation at other levels, such as between-population or between-site variation (Tibbetts and Dale, 2007).

Studies of animal vocalizations generally result in data with a hierarchical structure, wherein there are replicate calls from multiple individuals, often from multiple populations or sites (Mundry and Sommer, 2007). This produces nested clusters of calls that, despite being statistically non-independent, can reveal important sources of variation (Merlo *et al.*, 2005). Pairs of features (such as note duration and frequency) within a call may also exhibit non-independence, and patterns of variation and co-variation within and among clusters may hint at underlying ecological and evolutionary processes. While vocalizations are generally considered to be more plastic than many other phenotypic traits, they are still constrained by genetic variation and mechanical limits of sound production (Podos *et al.*, 2004). For example, there has been substantial interest in the trade-off between trill rate and bandwidth of trill notes in bird song (Ballentine *et al.*, 2004; Illes *et al.*, 2006; Podos, 1997; Wilson *et al.*, 2014). This tradeoff is also present in Bornean gibbon female great calls (Clink *et al.*, 2018), and in male lar gibbons (*Hylobates lar*) where the trill phrases of the male coda are restricted in their frequency range relative to other parts of the coda (Terleph *et al.*, 2018). It has been proposed that calls having both high trill rate and bandwidth are costly to produce, and that individuals better able to produce these costly calls may be higher quality mates (Illes *et al.*, 2006; Podos, 2016). Therefore, investigating the coevolution of—and tradeoffs among—call features has important implications for understanding the mechanisms and limitations of call production, as well as permitting assessment of the information conferred to conspecifics about caller quality.

Here, we use the Bornean gibbon (*Hylobates muelleri*) as a model to investigate three levels of variance (within-female, between-female and between-site) in features estimated from spectrograms of gibbon calls recorded in Sabah, Malaysia. Gibbons calls are largely inherited (Brockelman

and Schilling, 1984; Geissmann, 1984) and do not require a substantial learning component—but see Koda *et al.* (2013)—which means that they may be more appropriate models for understanding the evolutionary sources of variation in vocal behavior, relative to vocal learners such as songbirds. Similar to most studies of variation in animal vocalizations, our data exhibited a multi-level, multivariate structure because we estimated multiple features from each call and analyzed replicate calls from multiple females at multiple sites. To understand how variance in call features was partitioned at each level, as well as to investigate the covariance among features, we developed a Bayesian multivariate, variance components model. We estimated intraclass correlation coefficients for eight features based on posterior samples generated by a Hamiltonian Markov-chain Monte Carlo method, allowing us to make inferences regarding variance partitioning at each level (Merlo *et al.*, 2005). We also estimated correlation matrices for features at each of our three levels of analysis, which provide insight into possible tradeoffs inherent to gibbon call production. The benefits of our approach include the ability to estimate variance, along with uncertainty, of each call feature for each level of analysis, while simultaneously investigating correlations among call features across each level.

## II. MATERIALS AND METHODS

### A. Study subjects and data collection

D.J.C. collected recordings from unhabituated gibbons at seven sites across Sabah, Malaysia during multiple field seasons from January 2013 to September 2016 (map shown in Fig. 1; site information listed in Table I). We recorded vocalizations at a sampling rate of 44.1 kHz and 16-bit size using a Marantz PMD 660 flash recorder (Marantz, Kawasaki, Kanagawa Prefecture, Japan) equipped with a Røde NTG-2 directional condenser microphone (Røde Microphones, Sydney, Australia). To augment data collection, we used a Roland CUBE Street EX 4-Channel 50-W Battery Powered Amplifier (Roland Corporation, Osaka,

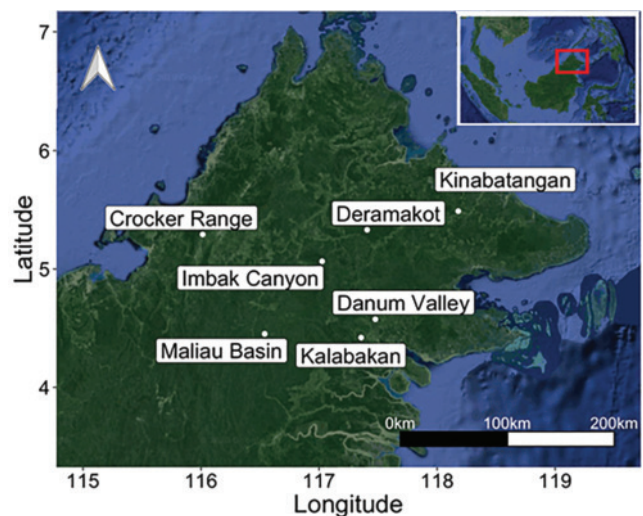


FIG. 1. (Color online) Map of seven recording locations in Sabah, Malaysia.

TABLE I. Site name, forest type, location, number of females, and number of calls from each site used for analysis.

Site	Forest type	Latitude	Longitude	Number of females	Number of calls
Crocker Range National Park	Montane, pristine	5.2934	116.01360	1	7
Deramakot Forest Reserve	Lowland, reduced impact logging	5.3322	117.40666	9	119
Danum Valley Conservation Area	Lowland, pristine	4.5752	117.47651	14	164
Imbak Canyon Conservation Area	Lowland, pristine	5.0662	117.02557	11	159
Lower Kinabatangan Wildlife Sanctuary	Riverine, logged	5.4978	118.17861	2	5
Maliau Basin Conservation Area	Lowland, pristine	4.4528	116.53899	5	62
Kalabakan Forest Reserve	Lowland, logged	4.4224	117.35560	33	368
			Total	75	884

Osaka Prefecture, Japan) to broadcast a recorded duet in assumed territories of gibbon groups. We considered recordings made more than 500 m apart to be from separate individuals, following territorial width estimates made by Brockelman and Srikosamatara (1993). In addition, we used group composition, along with unique calling behaviors that were detectable by a human observer (such as long calling bouts or distinct male contributions), to further differentiate between groups recorded in close proximity. In some cases, we had repeat recordings at the same recording location but on different days. We did not include these calls in our analysis since we could not be certain that we re-recorded the same group that we had recorded during the previous recording session.

Previously, we found that there were no quantitative differences in our results when we analyzed only calls collected during playbacks, compared to when we pooled calls collected during playbacks with those collected during spontaneous duetting (Clink *et al.*, 2017a). Therefore, we pooled calls in the same way for the present analysis. Moreover, we have no *a priori* reasons to predict that calls would vary by season because, although the forests of northern Borneo exhibit some temporal fluctuations in plant phenology and weather (Walsh and Newbery, 1999), they are aseasonal. Clink *et al.* (2017a) found that their results did not change when female great calls were combined across seasons, versus when great calls were restricted to only one season. This suggests that combining calls over all field recording seasons is appropriate for the present analysis.

## B. Acoustic analysis

The choice of features to use in analyses of animal vocalizations is highly subjective, and can range from many features (58 features; Oyakawa *et al.*, 2007) to relatively few (12 features; Haimoff and Gittins, 1985). Generally, the inclusion of a higher number of features does not provide more meaningful information regarding variation in gibbon calls (Heller *et al.*, 2010). To estimate call features in the present study we created spectrograms using the RAVEN PRO 1.5 sound analysis software (Cornell Lab of Ornithology Bioacoustics Research Program, Ithaca, New York) with a 512-point (11.6 ms) Hann window (3 dB bandwidth = 124 Hz), with 75% overlap, and a 1024-point discrete Fourier transform, yielding time and frequency measurement precision of 2.9 ms and 43.1 Hz. We then used the band-limited energy detector in RAVEN PRO to identify individual notes from each great call and used the

selection tables feature to estimate minimum frequency, maximum frequency, bandwidth, and duration for each note in the great call. Since most discriminative and modeling tasks require call feature vectors to have the same number of features, we further reduced our original dataset to 30 features estimated from the spectrogram—including minimum frequency, maximum frequency, bandwidth and duration of the first six introductory notes (duration greater than 0.135 s), along with total duration of each call and its various components. To estimate frequency, we used the robust measurements in RAVEN PRO, which are calculated based on the energy of the selected call segment, and are less sensitive to observer variability (Rice *et al.*, 2014); values reported here are 95% frequency (Charif *et al.*, 2008), therefore the actual maximum frequencies of gibbon calls are slightly higher. To estimate note duration, we used the delta time feature in RAVEN PRO. For our “rest duration” feature we summed the durations of pauses between the longer notes (greater than 0.135 s). We defined the trill as the portion of the call containing notes of less than 0.135 s (Clink *et al.*, 2017a), and calculated both the duration of the portion of the great call with notes longer than 0.135 s (the introductory duration) and the duration of the trill portion of the call.

We chose eight features from the original set of 30 features for subsequent modeling, based on the following reasoning. A trace plot of the singular values of the 30-feature dataset (Fig. 2) suggests that its actual dimension is considerably less than 30, and that an analysis based on 5 to 10 features would be appropriate. Furthermore, reducing the number of features is practical from a computational view: in contrast to discriminant analysis, the multivariate modeling contemplated here is computationally intensive. We first eliminated any feature from the initial set of 30 that was derived by calculation from one or more other features. For example, we did not include a note’s minimum frequency along with its maximum frequency and bandwidth, since minimum frequency is the difference between the latter features. Second, we eliminated features that were highly correlated with other features, based on visual inspection of a pairwise scatterplot matrix. The removal of features associated with later introductory notes (beyond introductory note 2) was a practical way to reduce dimension: this permitted the study of correlations between the successive notes 1 and 2, yet eliminated features of later introductory notes that were relatively redundant with those of the first two introductory notes, based on pairwise scatter plots. The final set

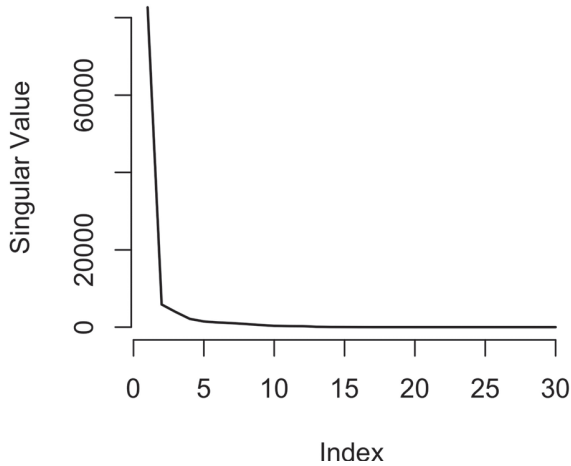


FIG. 2. Trace-plot of singular values, obtained from a singular value decomposition of the 30-feature dataset. The singular values are ordered from largest to smallest, and correspond to latent dimensions of the dataset in order of each dimension’s support. A well-supported dimension of variation is indicated by the first large singular value. Diminishing returns of support are apparent between the fifth and tenth singular values.

of eight features was selected prior to any modeling. A representative spectrogram of a Bornean gibbon female great call is shown in Fig. 3, and features used in the present analysis are summarized in Table II.

### C. Statistical analysis

Our data included three sources of variance: within-female, between-female, and between-site. Covariances between features are associated with each of these sources, for example the covariance between the duration and frequency of note 1 within a call, and so on. We assumed that these three sources of variance were independent and additive. All features are by definition positive and can therefore be log-transformed to stabilize variance, symmetrize feature distributions and scale different features compatibly. After log-transformation, we mean-centered the observations. Because variance components reflect deviations from the mean, mean-centering causes no loss of information, yet reduces the computational burden (Hofmann and Gavin,

TABLE II. Description of features extracted from spectrograms of Bornean gibbon female great calls.

Feature	Description
Note 1 duration (s)	Duration (s) of the first note
Note 1 maximum frequency (Hz)	95% of the maximum frequency of the first note
Note 2 duration (s)	Duration (s) of the second note
Note 2 maximum frequency (Hz)	95% of the maximum frequency of the second note
Rest duration (s)	Duration (s) of the resting period between the first five introductory notes
Intro duration (s)	Duration (s) of the portion of the call with notes longer than 0.135 s
Trill duration (s)	Duration (s) of the portion of the call with notes shorter than 0.135 s
Trill rate (notes per unit time)	Ratio of the number of notes in the trill over the total duration of the trill.

1998). The statistical challenges posed by a multivariate, hierarchical dataset led us to fit models using the computational Bayesian program Stan, implemented in the “RSTAN” package of the R programming environment (Guo *et al.*, 2016). Prior distributions for parameters—described below as part of model specification—play a stabilizing role in parameter estimation, and are a primary reason that computational Bayesian approaches are attractive for high-dimensional, structured data (Annis *et al.*, 2017; Gelman *et al.*, 2014; Gelman and Shirley, 2011).

### D. Model design

We defined our model for call  $c$ , female  $f$ , site  $s$  as

$$y_{s,f,c} = a_s + b_f + e_{s,f,c}, \quad (1)$$

where  $y$  is the log-transformed feature vector,  $a$  is a site-specific random intercept,  $b$  is a female-specific random intercept, and  $e$  is a call-specific error term. The additive

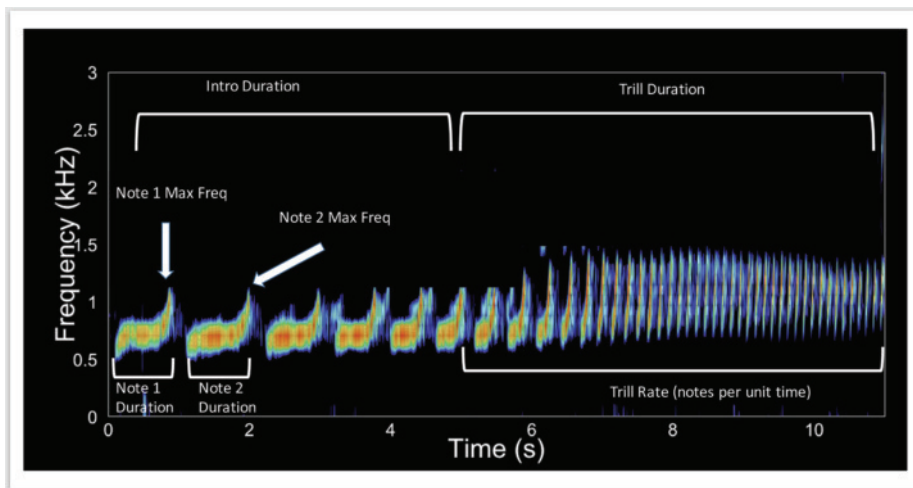


FIG. 3. (Color online) Representative spectrogram of a Bornean female great call along with features estimated from the spectrogram. The “rest duration” feature is not shown here, but is the sum of the duration of pauses in between the first five introductory notes. Spectrogram was created using the “seewave” package in the R programming environment using the default window and FFT size settings (R Development Core Team, 2017; Sueur *et al.*, 2008). Background noise and harmonics were also removed.

terms  $\mathbf{a}$ ,  $\mathbf{b}$ , and  $\mathbf{e}$ , correspond, respectively, to between-site, between-female, and within-female sources of variance. Variance/covariance matrices at each of the three levels measure the variability in each feature as well as the covariance among features. The variance/covariance matrices for  $\mathbf{a}$ ,  $\mathbf{b}$ , and  $\mathbf{e}$  are defined, respectively, as  $\Sigma_a$ ,  $\Sigma_b$ , and  $\Sigma_e$  and are assumed to be unstructured.

We carried out early computing experiments using multivariate Gaussian distributions for  $\mathbf{a}$ ,  $\mathbf{b}$ , and  $\mathbf{e}$ , but we found that Gaussian models fit our data poorly (see details below). The lack of fit came from over-dispersion in the observations so we chose a multivariate-t distribution, which allows for additional probability far from the mean of the observations, for each of the three additive terms (Roth, 2013). Variance/covariance matrices  $\Sigma_a$ ,  $\Sigma_b$ , and  $\Sigma_e$  derived from multivariate-t distributions have the formulas  $\Sigma_a = \nu_a \Phi_a / (\nu_a - 2)$ ,  $\Sigma_b = \nu_b \Phi_b / (\nu_b - 2)$ , and  $\Sigma_e = \nu_e \Phi_e / (\nu_e - 2)$ , where  $\nu_a$ ,  $\nu_b$ , and  $\nu_e$  are degrees-of-freedom parameters and  $\Phi_a$ ,  $\Phi_b$ , and  $\Phi_e$  are scale matrices. For computational efficiency, the scale matrices are further decomposed as  $\Phi_a = \mathbf{D}_a \Omega_a \mathbf{D}_a$  (for example), where  $\mathbf{D}_a$  is a diagonal matrix and  $\Omega_a$  is a correlation matrix (Stan Development Team, 2017).

We used a half-Cauchy prior for the elements of  $\mathbf{D}_a$ , and tested scale parameters 1, 5, and 25. We found that mixing was poor (i.e., the chains did not converge) with scale = 1, and ultimately settled on scale = 5. We used an LKJ prior for  $\Omega_a$ , tested parameters 0, 1, and 1.5 and found little sensitivity to this choice, ultimately settling on 1.5 (Stan Development Team, 2017).  $\Phi_b$ , and  $\Phi_e$  are similarly parameterized. We used a Gamma prior with shape = 2 and rate = 0.1 for  $\nu_a$ ,  $\nu_b$ , and  $\nu_e$ , truncated on the left at the value 2 to prevent singularities in expressions like  $\Sigma_a = \nu_a \Phi_a / (\nu_a - 2)$ .

We simulated 500 parameter samples from each of two Markov chains, after a warmup of 2500 samples, for a total of 1000 samples for posterior inference. Trace-plots of model parameters and convergence diagnostics implemented in RSTAN suggested that mixing was adequate for inference. Computing time was approximately 3.5 h using a MacBook Pro with 2.7 GHz Intel Core.

### E. Intraclass correlation coefficients

Intraclass correlation coefficients measure the relative contribution of each level (within-female, between-female and between-site) to the total variance. We calculated the intraclass correlation coefficient (ICC) at level  $l$  for each feature from posterior samples of  $\Sigma_a$ ,  $\Sigma_b$ , and  $\Sigma_e$  as

$$\text{ICC}_l = \frac{\text{Variance of feature at level } l}{\text{Total variance of feature}}.$$

As the variances along the main diagonals of  $\Sigma_a$ ,  $\Sigma_b$ , and  $\Sigma_e$  can only be positive, ICC values necessarily range between 0 and 1. An ICC near 1 for a given level indicates that the level is an important source of variation (Merlo *et al.*, 2005).

### F. Goodness of fit

We calculated a scaled, squared Mahalanobis distance between each observation and its predicted value, to

determine whether the model fit was adequate. For a given observation  $\mathbf{y}$ , the predicted value based on posterior sample  $t$  is  $\hat{\mathbf{y}}_t$ , and the distance is

$$d_t^2(\mathbf{y}) = \frac{1}{p} (\mathbf{y} - \hat{\mathbf{y}}_t)' \Phi_{e,t}^{-1} (\mathbf{y} - \hat{\mathbf{y}}_t),$$

where  $p$  is the number of features and  $\Phi_{e,t}$  is the scale matrix, at the within-female level, for posterior sample  $t$ . The values  $d_t^2(\mathbf{y})$  were averaged over posterior samples to obtain a posterior mean distance,  $\bar{d}^2(\mathbf{y})$ , for each observation. These were compared, by use of a quantile-quantile plot, to the F distribution having  $p$  and  $\bar{\nu}_e$  degrees of freedom, where  $\bar{\nu}_e$  is the posterior mean of  $\nu_e$  (Roth, 2013).

### G. Data availability

All data and R code necessary to reproduce our models and figures are included as supplementary material<sup>1</sup> and posted on GitHub (Clink *et al.*, 2017b).

### H. Ethical note

All research complied with local and international laws and regulations and the ethical principles of the Acoustical Society of America were followed in regard to the use of vertebrate animals in research.

## III. RESULTS

### A. Intraclass correlation coefficients

We report the results of the analysis of 884 great calls from 75 females (median number of calls per individual: 11; range: 1–43) recorded from seven sites across Sabah, Malaysia. We show the posterior density estimates of ICCs for three levels (within-female, between-female, and between-site) for each feature in Fig. 4. Based on ICC values estimated from posterior densities we found that for one feature, rest duration, variation at the level of the call dominated other sources of variation [ICC posterior mean = 0.58; 95% posterior credibility interval = (0.44, 0.66)]. Call-level variance was also important for the note 1 maximum frequency [mean = 0.36; CI = (0.27, 0.45)] and note 2 maximum frequency [mean = 0.36; CI = (0.27, 0.45)], although female-level variation was even more important for these features (Fig. 4).

For all of the features except for rest duration and trill rate, female-level variance dominated other sources of variation (Fig. 4). For one feature, trill rate, site-level variance was the most important contributor to total variance [mean = 0.52; CI = (0.18, 0.89)], but female-to-female variance was also important [mean = 0.44; CI = (0.10, 0.75)]. The site-specific intercept for trill rate is an element of  $\mathbf{a}_s$ , one of the additive sources of variance for  $\mathbf{y}_{s,f,c}$ . The intercepts measure deviations of trill rates at each site from an overall baseline trill rate. Posterior densities of the site-specific intercepts for trill rate are shown in Fig. 4. The intercepts were well separated, particularly for females recorded in the Deramakot Forest Reserve and Lower Kinabatangan Wildlife Sanctuary (Fig. 5), indicating that these females have lower trill rates than females from the other sites.

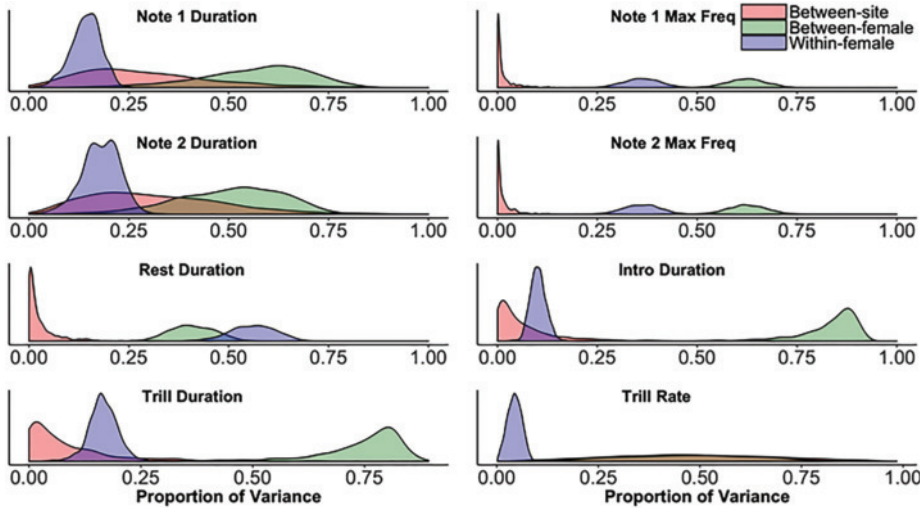


FIG. 4. (Color online) Posterior densities for the intraclass correlation coefficients for the three levels in our dataset (within-female, between-female, and between-site) for each feature of the Bornean gibbon great call. Call-level variance was the most important contributor to total variance for rest duration, and site-level variance was the most important for trill rate. For six of the eight features inter-individual (between-female) variation was the most important contributor to total variance. In each graph the y-axis refers to “density” and is not shown. Only the relative densities between classes matter.

## B. Correlation matrices

We found that features tended to be correlated at the within-female level and at the between-female level (Fig. 6), but that there was less correlation among features between sites. Notably, at the within-female level, the maximum frequency of note 1 and note 2 were positively correlated, along with note 1 and note 2 durations, and note 2 duration and rest duration. At the between-female level, maximum frequency of note 1 and note 2 were positively correlated, along with note 1 and note 2 durations, and intro duration and note 2 duration. In addition, there was a negative correlation between note 1 duration and note 1 maximum frequency, note 2 duration and note 2 maximum frequency, and intro duration and trill rate. At the between-site level there was little correlation among features (Fig. 6).

## C. Multivariate-t degrees of freedom

The posterior degrees of freedom were fairly small for call [ICC posterior mean = 4.8; 95% posterior credibility interval = (4.29, 5.58)], female [mean = 17.12; CI = (7.70, 37.95)] and site [mean = 19.39; CI = (2.70, 53.68)]. The

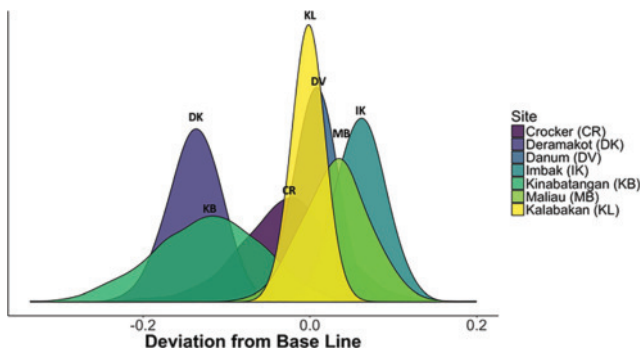


FIG. 5. (Color online) Posterior densities for site-specific intercepts for trill rate in the Bornean gibbon female great call. The site-specific intercepts for females recorded in the Deramakot Forest Reserve and Lower Kinabatangan Wildlife Sanctuary (denoted as DK and KB, respectively, in the figure) were separated from the rest, indicating that females at these sites had lower trill rate than females at other sites. The y-axis refers to “density” and is not shown.

small degrees of freedom parameters reaffirm the need for multivariate-t distributions for the three sources of variance, as opposed to multivariate Gaussian distributions, the latter of which would be suggested by large degrees of freedom parameters.

## D. Goodness of fit test

In order to see how well our model fit the data, we compared a Q-Q plot of posterior mean distances between observations and their predicted values, versus an appropriate F distribution (Fig. 7). The agreement between the empirical and theoretical quantiles is good for almost all observations, indicating that the model in Eq. (1), with additive terms  $a$ ,  $b$ , and  $e$  having multivariate-t distributions, fits the observations well. The observation at the extreme right of the Q-Q plot draws attention, but further confirms an adequate fit. This observation lies below the 1:1 line and is therefore closer than expected to its predicted value.

## IV. DISCUSSION

Gibbon calls are generally thought to be genetically determined, as seen in studies of both captive (Geissmann, 1984; Tenaza, 1985) and wild (Brockelman and Schilling, 1984) gibbon hybrids where offspring produce sex-specific calls of a parental species which in some cases they have never heard. We note, however, that high heritability in call parameters does not preclude population- and individual-level variation and refinement. Here, we investigated within-female, between-female and between-site variance of 884 great calls from 75 Bornean gibbon females. We showed that inter-individual variation was the most important source of variance for all features except the duration of rest in the introduction and trill rate. For rest duration, within-female variance was the most important source of variance, whereas between-site variance was the most important source of variance for trill rate. We also found interesting patterns of correlation among features, particularly at the individual-level, suggesting tradeoffs in the mechanical production of calls (as described below).

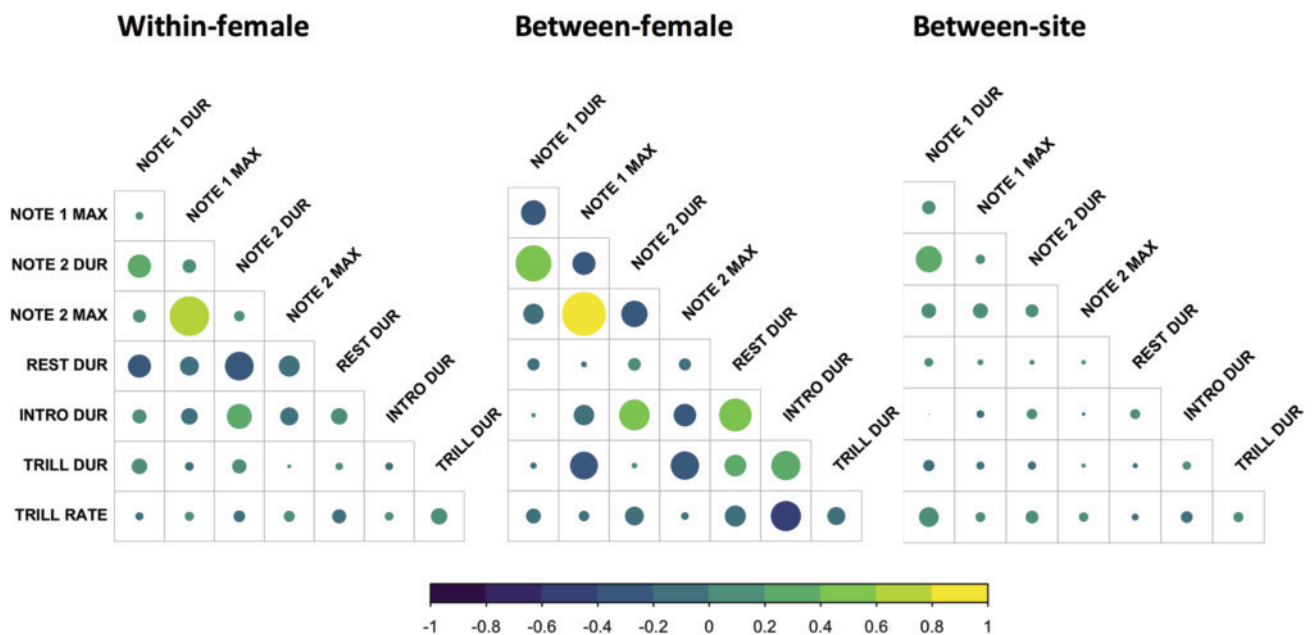


FIG. 6. (Color online) Correlation matrices for the eight features at each level of analysis: within-female, between-female, and between-site. Circle size indicates the magnitude of correlation, and different colors indicate whether the correlation is positive or negative.

### A. Benefits of our analytical approach

Like most animal vocalization studies, our dataset exhibited a hierarchical structure with multiple calls recorded from individuals at different sites. In addition, as we estimated more than one feature for each call, our dataset was also multivariate. Our multivariate, variance components model has broad-scale applicability for many bioacoustics applications related to understanding patterns of geographic variation or individual differences in animal vocalizations. A benefit of our approach over other approaches is that it allowed us to estimate variances for each feature at each level of analysis, and also to characterize the uncertainty of these estimates. Previous studies on geographic variation in primates have relied on discriminant function analysis to distinguish between populations (Delgado, 2007; Mitani *et al.*, 1999; Wich *et al.*, 2008), but there are important methodological concerns about using discriminant function analysis for multi-level data, and thus accounts of site- or population-level differences in vocalizations that are based on these methods may be over-estimated (Mundry and

Sommer, 2007). Mundry and Sommer (2007) provide an alternative approach—the permuted discriminant function analysis—which can be used to classify non-independent calls to various classes such as sites, sexes, or individuals. A major difference between the permuted discriminant function analysis and our approach is that the permuted discriminant function analysis is better suited for classification tasks, whereas our approach is better suited for answering questions about how call features vary across different levels of analysis. In addition, a major strength of our approach is that it allowed us to investigate correlations among features at different levels. Understanding how features are correlated can provide important insights into the evolution and co-evolution of call features (Fedurek *et al.*, 2017; Gustison *et al.*, 2016; Podos *et al.*, 2004). In addition, our approach can be adapted to include fewer or more levels of analysis and/or covariate predictors. We provide the model and R code as online supporting material<sup>1</sup> and encourage other researchers to adapt and expand our method for their research objectives.

### B. Justification of choice of features

A requirement for most discriminative and modelling approaches is that the same number of features must be included for each call, which requires researchers to make decisions about which features to use if calls are of variable length or have varying numbers of notes. The choice of features is somewhat subjective, and will be determined in part by the types of questions and the analysis (e.g., discrimination tasks or hypothesis testing) that is being done. If multivariate modelling is the goal, as it was here, practical computing constraints will be decisive—possibly requiring that a large number of available features be reduced to a smaller, computationally tractable set. In a recent paper

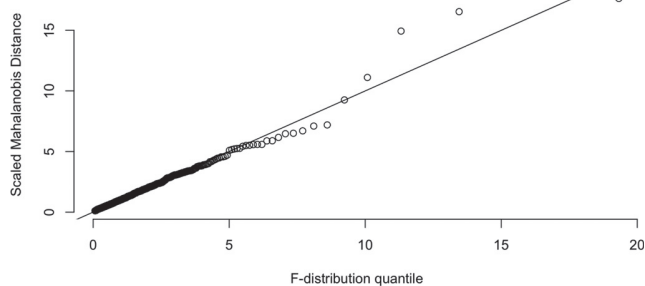


FIG. 7. Posterior mean Mahalanobis distances, squared and scaled by the number of features, versus F distribution quantiles. A Q-Q plot in which points stay close to the central line suggests a well-fitting model.

investigating variation in Bornean gibbon female calls from a single site, [Clink \*et al.\* \(2017a\)](#) showed that the duration of the introductory notes varied consistently among females. This partially guided our decision to focus on these notes for the present analysis. We focused on the first two introductory notes, as we found during exploratory data analysis that features of the first introductory notes were highly correlated with those of the later introductory notes. The inclusion of both the first and second introductory notes allows for the study and modelling of correlations across notes, whereas the further inclusion of the third, fourth and later introductory notes would seem to increase the computational burden while offering relatively little inferential returns.

We were also interested in understanding if there was variation in the duration of rest between notes, as it was shown in lar gibbons (*H. lar*) that this feature varied among females of different ages ([Terleph \*et al.\*, 2016](#)). Specifically, [Terleph \*et al.\* \(2016\)](#) measured the duration of the notes divided by the duration of the notes plus the inter-note intervals, providing a metric of note output relative to the breaks between notes. A calculation similar to [Terleph \*et al.\* \(2016\)](#) produced a variable that was highly collinear with other features of interest in our data, and therefore problematic for multivariate modelling. We opted instead to measure “rest duration” as the total duration of rest between the first five introductory notes. We did not include rest duration between trill notes as the notes often exhibited a high-degree of reverberation, along with very short inter-note intervals, and we did not feel that our method of feature extraction would allow us to effectively measure this feature. Last, we decided to include trill rate as it has been shown that there are patterns of geographic variation in this feature across taxa ([Campbell \*et al.\*, 2010](#); [Derryberry, 2009](#); [de la Torre and Snowdon, 2009](#)); furthermore, the acoustic adaptation hypothesis suggests that the rate of note repetition may vary across different habitat types ([Morton, 1975](#); [Wiley, 1991](#)).

### C. Individual-level variation

A previous study on agile gibbons from populations on Sumatra and Borneo showed considerable within-individual variability, particularly in the spectral parameters ([Heller \*et al.\*, 2010](#)), which is consistent with our finding of substantial intra-individual variation in note maximum frequency. We collected recordings from unhabituated gibbons at variable recording distances, and recording distance can have an effect on the frequency estimates ([Kroodsma, 2017](#); [Zollinger \*et al.\*, 2012](#)). In addition, gibbons tended to move during the course of a duet, and this change in recording distance over the course of a calling bout may have affected our estimates. It is also possible that animals alter the spectral properties of their calls in response to behavioral or motivational state ([Morton, 1977](#)). For example, chimpanzees alter the temporal and spectral features of their calls depending on whether they are in a neutral or aggressive social context ([Siebert and Parr, 2003](#)), and also depending on their social role (e.g., victim or aggressor) in a conflict ([Slocombe and Zuberbühler, 2005](#)), although it is unclear how or if the behavioral states of gibbons change over the course of a

calling bout. Alternately, it could be that the observed variation in call frequency is essentially noise that the nervous system is unable to control ([Tumer and Brainard, 2007](#)). We also found that there was substantial intra-individual variation in the duration of rest during the introductory portion of the great call. In lar gibbons, younger females were shown to have longer note output relative to the duration of rest in between notes ([Terleph \*et al.\*, 2016](#)), and we predicted that we would find substantial inter-individual variation in our rest duration feature. There is no evidence that the duration of rest increases over the course of a calling bout, and we do not know why rest duration is so variable within individuals.

Previously, we found that there was substantial between-female variation in the Bornean gibbon great calls recorded from a single site in Sabah, Malaysia ([Clink \*et al.\*, 2017a](#)), and that the duration of notes was most effective for discriminating between individuals. Here, we show that for both temporal and spectral features of the calls, between-female variation is the most important source of variance. The substantial inter-individual variation, particularly from a large dataset of 884 calls from 75 different females, suggests that there has been selection for the gibbon female great call to encode information regarding caller identity ([Tibbetts and Dale, 2007](#)). It is also possible that some of the between-female variation is related to differences in ages of individuals in our populations, as younger female lar gibbon calls have been shown to have higher fundamental frequency ([Terleph \*et al.\*, 2016](#)).

In addition to age-related changes, it is also possible that call features of an individual change over time in response to hormonal condition ([Barelli \*et al.\*, 2013](#)). As we collected data on unhabituated animals, our dataset did not allow for us to examine variation across time or calling bouts for a particular animal. Our model can easily be adapted to include another level of analysis, for example variance between calling bouts, and future analyses using datasets from known individuals recorded across time will be informative. In addition, since we were working with unhabituated groups with unknown territory sizes there was a possibility of re-recording the same female and considering her to be two separate females. This error of assignment would seemingly reduce estimates of between-female variance. For most features in our dataset, between-female variance was the largest source of variance. It could thus be argued that even this important variance component is a conservative lower bound for the true component.

We do not know whether gibbon individuals discriminate between other individuals on the basis of their calls. A previous study investigating gibbon response to playback of self, neighbor, and stranger calls did not show that gibbons respond differently to the different playback treatments ([Mitani, 1985](#)). [Mitani \(1985\)](#) noted that the lack of difference in behavioral response to the different treatments may be related to the experimental design, or that there may not be a selective advantage in responding differently to the various treatments. Further research examining whether gibbon female calls confer information regarding individual identity will be informative.

#### D. Site-level variation in trill rate

In the present study on Bornean gibbons, we showed that for one feature in our dataset, trill rate, site-level variance was the most important source of variance. Our results are consistent with other studies of birds and mammals documenting geographic variation in trill rate. In white-crowned sparrows (*Zonotrichia leucophrys*), trill rate was shown to vary among populations, and variation was best explained by bird bill size and vegetation density (Derryberry, 2009). In a meta-analysis of 120 North American territorial oscine birds, birds found in open habitats had songs with short repetition periods, whereas many (but not all) bird songs in forested habitats lacked this feature (Wiley, 1991). In neotropical singing mice (*Scotinomys spp.*), the rate of note repetition was found to vary within-populations and between species (Campbell *et al.*, 2010). The authors concluded that population-level differences in vocalizations were largely the result of neutral evolutionary processes, given the correlation between acoustic and genetic distance, but selection was important in shaping differences between species. In pygmy marmosets (*C. pygmaea*), there were also population-level differences in trills, and the authors proposed that these differences may be related to differences in the environment, specifically ambient noise levels, or the result of phenotypic plasticity and/or genetic drift (de la Torre and Snowdon, 2009).

For our study, we recorded gibbons in montane, riverine and lowland forests. Although the distinct forest types in Sabah have not been definitively studied, we surmised that their different environmental characteristics, such as ambient temperature and humidity, differences in acoustic competitors or forest structure, could influence gibbon calls. We found that two sites exhibited lower trill rates than the rest, Deramakot Forest Reserve and Lower Kinabatangan Wildlife Sanctuary, and these sites are isolated from the other sites by the Kinabatangan river. In orangutans, there is a high degree of genetic differentiation between populations on either side of the river, and the river appears to be a geographic barrier to gene flow (Goossens *et al.*, 2005). It seems likely that the river serves as a barrier to gene flow for gibbons as well, and that differences we found in trill rates between sites north and south of the Kinabatangan are due to the geographic isolation of these separate populations. However, it is possible that habitat differences north and south of the river may also contribute to trill rate differences. Further research that incorporates environmental variables as well as genetic data will be informative in teasing apart the relative importance of neutral and adaptive processes in shaping the observed patterns of variation in trill rate.

#### E. Tradeoffs

Primate acoustic signals exhibit a broad range of variation, presumably shaped by a combination of evolutionary forces, but are not infinitely plastic. There are both physiological and cognitive constraints on the limits of call production and innovation (Cheney and Seyfarth, 2005), and the morphology and neurophysiology of both the sender and receiver will limit the potential evolutionary outcomes of

signal evolution (Wilkins *et al.*, 2012). For example, body size and beak shape are morphological traits that can limit signal evolution in birds (Podos, 2001). There is also a correlation between song complexity and the volume of regions of the brain that male songbirds use to learn and produce songs (DeVoogd, 2004), and there are presumably physiological limits to the growth of these regions, which would in turn limit song complexity. Quantifying tradeoffs in the production of acoustic signals can potentially provide information about the evolutionary mechanisms that shaped these signals.

A major benefit of our approach over other types of analyses is that it allowed for us to investigate correlations among call features at different levels of analysis. We show that correlations among features are most detectable at the between-female level. This reinforces the notion that caller identity can be captured by a suite of features that vary together as an ensemble. Several negative correlations among features at this level further suggest tradeoffs in the production of gibbon great calls. For example, for a given female, we showed that there was a negative correlation between note duration and frequency, indicating that there may be a trade-off between the length of the introductory notes in her calls and the maximum frequency of the notes; a female may either have longer notes or higher maximum frequency but perhaps not both. We previously showed a trade-off between note bandwidth and rate of note repetition in Bornean gibbon female trills (Clink *et al.*, 2018), and this tradeoff may be the result of physical limitations on the animals' ability to rapidly modulate their vocal tracts while maintaining tonal frequencies (Wilson *et al.*, 2014). It seems that this may also be the case with the duration of a note and the maximum frequency of that particular note, as seen in the Bornean gibbon female introductory notes, but the morphological or physiological reasons for this are yet to be determined.

We also detected a negative correlation between the duration of the introduction and the rate of the trill portion of the call; a female may either have a long introduction or a faster trill rate, but perhaps not both. Although the mechanisms which could impose such a tradeoff are not known, it seems likely that rapid repetition of trill notes is physically challenging for the caller, and it may be that longer introductory notes are also physically challenging to produce. Physiological constraints of this kind could lead to the negative correlation we detected. We provide an investigation of the correlation among features in Bornean gibbon female calls as an example of what can be done with our approach, and invite future researchers to adapt this approach to investigate tradeoffs in acoustic signals in the context of their own research questions.

#### ACKNOWLEDGMENTS

Our research was approved by the Sabah Biodiversity Centre JKM/MBS 1000–2/2(90). Permission to conduct research in Crocker Range National Park was kindly provided by Sabah Parks. Permission to conduct research in Deramakot Forest Reserve was provided by the Sabah

Forestry Department. Permission to conduct research in the Lower Kinabatangan Wildlife Sanctuary was granted by the Kinabatangan Wildlife Department. Permission to conduct research in Danum Valley, Imbak Canyon, and Maliau Basin was kindly provided by their respective management committees and Yaysan Sabah. We gratefully acknowledge our local counterpart, Dr. Henry Bernard of Universiti Malaysia, Sabah, for his support with permits and logistics. Funding to conduct this research was kindly provided by Primate Conservation, Inc., Cleveland Metroparks Zoo, the American Primatological Association, and the Fulbright U.S. Student Program. We also acknowledge Ahbam Abulani, Johnny Bin Kissing, and the HUTAN-Kinabatangan Orangutan Conservation Project for their assistance in facilitating data collection. We also gratefully acknowledge Rob Ewers of Imperial College London for providing logistical support for a substantial portion of this project, along with all of the staff at the Stability of Altered Forest Ecosystems project and the South East Asia Rainforest Research Programme (SEARRP) management. We also thank the many research assistants who helped with logistics and data collection, including M. Ramadhan, S. Watson, R. Grey, M. Khoo, S. Rao, K. Kramarczuk, and C. Sullivan. Last, we extend a huge thanks to T. Thomas, M. Mercedes Vollmar, M. Borden, and A. Lau for their help processing countless hours of gibbon recordings. We also thank two anonymous reviewers for their helpful comments on previous versions of this manuscript.

Annis, J., Miller, B. J., and Palmeri, T. J. (2017). "Bayesian inference with Stan: A tutorial on adding custom distributions," *Behav. Res. Methods* **49**, 863–886.

Ballentine, B., Hyman, J., and Nowicki, S. (2004). "Vocal performance influences female response to male bird song: An experimental test," *Behav. Ecol.* **15**, 163–168.

Barelli, C., Mundry, R., Heistermann, M., and Hammerschmidt, K. (2013). "Cues to androgens and quality in male gibbon songs," *PLoS One* **8**, e82748.

Boncoraglio, G., and Saino, N. (2007). "Habitat structure and the evolution of bird song: A meta-analysis of the evidence for the acoustic adaptation hypothesis," *Funct. Ecol.* **21**, 134–142.

Brockelman, W. Y., and Schilling, D. (1984). "Inheritance of stereotyped gibbon calls," *Nature* **312**, 634–636.

Brockelman, W. Y., and Srikosamatara, S. (1993). "Estimation of density of gibbon groups by use of loud songs," *Am. J. Primatol.* **29**, 93–108.

Burton, J. A., and Nietsch, A. (2010). "Geographical variation in duet songs of Sulawesi tarsiers: Evidence for new cryptic species in south and south-east Sulawesi," *Int. J. Primatol.* **31**, 1123–1146.

Campbell, P., Pasch, B., Pino, J. L., Crino, O. L., Phillips, M., and Phelps, S. M. (2010). "Geographic variation in the songs of neotropical singing mice: Testing the relative importance of drift and local adaptation," *Evolution* **64**, 1955–1972.

Ceugniet, M., and Izumi, A. (2004). "Vocal individual discrimination in Japanese monkeys," *Primates* **45**, 119–128.

Charif, R. A., Clark, C. W., and Frisrup, K. M. (2008). "Raven Pro User's Manual."

Cheney, D. L., and Seyfarth, R. M. (1980). "Vocal recognition in free-ranging vervet monkeys," *Anim. Behav.* **28**, 362–367.

Cheney, D. L., and Seyfarth, R. M. (1999). "Recognition of other individuals' social relationships by female baboons," *Anim. Behav.* **58**, 67–75.

Cheney, D. L., and Seyfarth, R. M. (2005). "Constraints and preadaptations in the earliest stages of language evolution," *Linguist. Rev.* **22**, 135–159.

Clink, D. J., Bernard, H., Crofoot, M. C., and Marshall, A. J. (2017a). "Investigating individual vocal signatures and small-scale patterns of geographic variation in female Bornean gibbon (*Hylobates muelleri*) great calls," *Int. J. Primatol.* **38**, 656–671.

Clink, D. J., Charif, R. A., Crofoot, M. C., and Marshall, A. J. (2018). "Evidence of vocal performance constraints in a female non-human primate," *Anim. Behav.* **141**, 85–94.

Clink, D. J., Grote, M. N., Crofoot, M. C., and Marshall, A. J. (2017b). <https://github.com/DenaJGibbon/MANOVA.stan> (Last viewed December 12, 2017).

Dallmann, R., and Geissmann, T. (2009). "Individual and geographical variability in the songs of wild silvery gibbons (*Hylobates moloch*) on Java, Indonesia," in *The Gibbons*, edited by D. Whittaker and S. Lappan (Springer, New York), pp. 91–110.

de la Torre, S., and Snowdon, C. T. (2009). "Dialects in pygmy marmosets? Population variation in call structure," *Am. J. Primatol.* **71**, 333–342.

Delgado, R. A. (2007). "Geographic variation in the long calls of male orangutans (*Pongo spp.*)," *Ethology* **113**, 487–498.

Derryberry, E. P. (2009). "Ecology shapes birdsong evolution: Variation in morphology and habitat explains variation in white-crowned sparrow song," *Am. Nat.* **174**, 24–33.

DeVoogd, T. J. (2004). "Neural constraints on the complexity of avian song," *Brain. Behav. Evol.* **63**, 221–232.

Fedurek, P., Zuberbühler, K., and Semple, S. (2017). "Trade-offs in the production of animal vocal sequences: Insights from the structure of wild chimpanzee pant hoots," *Front. Zool.* **14**, 50–60.

Fischer, J., Hammerschmidt, K., Cheney, D. L., and Seyfarth, R. M. (2002). "Acoustic features of male baboon loud calls: Influences of context, age, and individuality," *J. Acoust. Soc. Am.* **111**, 1465–1474.

Geissmann, T. (1984). "Inheritance of song parameters in the gibbon song, analysed in 2 hybrid gibbons (*Hylobates pileatus* × *H lar*)," *Folia Primatol.* **42**, 216–235.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian Data Analysis* (CRC Press, Boca Raton, FL), Vol. 2.

Gelman, A., and Shirley, K. (2011). "Inference from simulations and monitoring convergence," in *Handbook of Markov Chain Monte Carlo*, edited by S. Brooks, A. Gelman, G. L. Jones, and X. L. Meng (Chapman & Hall/CRC, New York).

Goossens, B., Chikhi, L., Jalil, M. F., Ancrenaz, M., Lackman-Ancrenaz, I., Mohamed, M., Andau, P., and Bruford, M. W. (2005). "Patterns of genetic diversity and migration in increasingly fragmented and declining orangutan (*Pongo pygmaeus*) populations from Sabah, Malaysia," *Mol. Ecol.* **14**(2), 441–456.

Guo, J., Lee, D., Sakrejda, K., Gabry, J., and Goodrich, B. (2016). "rstan: R Interface to Stan," R Packaged version.

Gustison, M. L., Semple, S., Ferrer-i-Cancho, R., and Bergman, T. J. (2016). "Gelada vocal sequences follow Menzerath's linguistic law," *Proc. Natl. Acad. Sci. U.S.A.* **113**, E2750–E2758.

Haimoff, E., and Gittins, S. (1985). "Individuality in the songs of wild agile gibbons (*Hylobates agilis*) of peninsular Malaysia," *Am. J. Primatol.* **8**, 239–247.

Haimoff, E., and Tilson, R. (1985). "Individuality in the female songs of wild Kloss' gibbons (*Hylobates klossii*) on Siberut Island, Indonesia," *Folia Primatol.* **44**, 129–137.

Heller, R. A., Sander, A. F., Wang, C. W., Usman, F., and Dabelsteen, T. (2010). "Macrogeographical variability in the great call of *Hylobates agilis*: Assessing the applicability of vocal analysis in studies of fine-scale taxonomy of gibbons," *Am. J. Primatol.* **72**, 142–151.

Hofmann, D. A., and Gavin, M. B. (1998). "Centering decisions in hierarchical linear models: Implications for research in organizations," *J. Manage.* **24**, 623–641.

Illes, A. E., Hall, M. L., and Vehrencamp, S. L. (2006). "Vocal performance influences male receiver response in the banded wren," *Proc. R. Soc. B Biol. Sci.* **273**, 1907–1912.

Irwin, D. E., Thimban, M. P., and Irwin, J. H. (2008). "Call divergence is correlated with geographic and genetic distance in greenish warblers (*Phylloscopus trochiloides*): A strong role for stochasticity in signal evolution?," *J. Evol. Biol.* **21**, 435–448.

Koda, H., Lemasson, A., Oyakawa, C., Rizaldi, Pamungkas, J., and Masataka, N. (2013). "Possible role of mother-daughter vocal interactions on the development of species-specific song in gibbons," *PLoS One* **8**, e71432.

Kroodsma, D. (2017). "Birdsong performance studies: A contrary view," *Anim. Behav.* **125**, e1–e16.

- Lameira, A. R., and Wich, S. A. (2008). "Orangutan long call degradation and individuality over distance: A playback approach," *Int. J. Primatol.* **29**, 615–625.
- Leliveld, L. M. C., Scheumann, M., and Zimmermann, E. (2011). "Acoustic correlates of individuality in the vocal repertoire of a nocturnal primate (*Microcebus murinus*)," *J. Acoust. Soc. Am.* **129**, 2278–2288.
- Macedonia, J. M. (1986). "Individuality in a contact call of the ringtailed lemur (*Lemur catta*)," *Am. J. Primatol.* **11**, 163–179.
- Marten, K., and Marler, P. (1977). "Sound transmission and its significance for animal vocalization," *Behav. Ecol. Sociobiol.* **2**, 271–290.
- Merlo, J., Chaix, B., Yang, M., Lynch, J., and Råstam, L. (2005). "A brief conceptual tutorial of multilevel analysis in social epidemiology: Linking the statistical concept of clustering to the idea of contextual phenomenon," *J. Epidemiol.* **59**, 443–449.
- Mitani, J. C. (1985). "Responses of gibbons (*Hylobates muelleri*) to self, neighbor, and stranger song duets," *Int. J. Primatol.* **6**, 193–200.
- Mitani, J. C., Gros-Louis, J., and Macedonia, J. M. (1996). "Selection for acoustic individuality within the vocal repertoire of wild chimpanzees," *Int. J. Primatol.* **17**, 569–583.
- Mitani, J. C., Hunley, K. L., Murdoch, M. E., and Arbor, A. (1999). "Geographic variation in the calls of wild chimpanzees: A reassessment," *Am. J. Primatol.* **47**, 133–151.
- Morton, E. S. (1975). "Ecological selection on avian sounds," *Am. Nat.* **109**, 17–34.
- Morton, E. S. (1977). "On the occurrence and significance of motivation-structural rules in some bird and mammal sounds," *Am. Nat.* **111**, 855–869.
- Mundry, R., and Sommer, C. (2007). "Discriminant function analysis with nonindependent data: Consequences and an alternative," *Anim. Behav.* **74**, 965–976.
- Oyakawa, C., Koda, H., and Sugiura, H. (2007). "Acoustic features contributing to the individuality of wild agile gibbon (*Hylobates agilis agilis*) songs," *Am. J. Primatol.* **69**, 777–790.
- 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 signal structure in Darwin's finches," *Nature* **409**, 185–187.
- Podos, J. (2016). "Birdsong performance studies: Reports of their death have been greatly exaggerated," *Anim. Behav.* **125**, e17–e24.
- Podos, J., Huber, S. K., and Taft, B. (2004). "Bird song: The interface of evolution and mechanism," *Annu. Rev. Ecol. Evol. Syst.* **35**, 55–87.
- R Development Core Team (2017). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Řeháková-Petrů, M., Policht, R., and Peške, L. (2012). "Acoustic repertoire of the Philippine tarsier (*Tarsius syrichta fraterculus*) and individual variation of long-distance calls," *Int. J. Zool.* **2012**, 1–10.
- Rendall, D., Rodman, P. S., and Emond, R. E. (1996). "Vocal recognition of individuals and kin in free-ranging rhesus monkeys," *Anim. Behav.* **51**, 1007–1015.
- Rice, A. N., Palmer, K. J., Tielens, J. T., Muirhead, C. A., and Clark, C. W. (2014). "Potential Bryde's whale (*Balaenoptera edeni*) calls recorded in the northern Gulf of Mexico," *J. Acoust. Soc. Am.* **135**, 3066–3076.
- Roth, M. (2013). "On the multivariate t distribution," Technical Report from Automatic Control, Linköping University, Linköping, Sweden.
- Salmi, R., Hammerschmidt, K., and Doran-Sheehy, D. M. (2014). "Individual distinctiveness in call types of wild Western female gorillas," *PLoS One* **9**, e101940.
- Semple, S. (2001). "Individuality and male discrimination of female copulation calls in the yellow baboon," *Anim. Behav.* **61**, 1023–1028.
- Siebert, E. R., and Parr, L. A. (2003). "A structural and contextual analysis of chimpanzee screams," *Ann. N.Y. Acad. Sci.* **100**, 104–109.
- Slabbekoorn, H., and Smith, T. (2002). "Habitat-dependent song divergence in the little greenbul: An analysis of environmental selection pressures on acoustic signals," *Evolution* **56**, 1849–1858.
- Slocombe, K., and Zuberbühler, K. (2005). "Agonistic screams in wild chimpanzees (*Pan troglodytes schweinfurthii*) vary as a function of social role," *J. Comp. Psychol.* **119**, 67–77.
- Stan Development Team (2017). "Stan Modeling Language Users Guide and Reference Manual."
- Sueur, J., Aubin, T., and Simonis, C. (2008). "Seewave: A free modular tool for sound analysis and synthesis," *Bioacoustics* **18**, 213–226.
- Sun, K., Luo, L., Kimball, R. T., Wei, X., Jin, L., Jiang, T., Li, G., and Feng, J. (2013). "Geographic variation in the acoustic traits of greater horseshoe bats: Testing the importance of drift and ecological selection in evolutionary processes," *PLoS One* **8**, e70368.
- Symmes, D., Newman, J. D., Talmage-Riggs, G., and Lieblich, A. K. (1979). "Individuality and stability of isolation peeps in squirrel monkeys," *Anim. Behav.* **27**, 1142–1152.
- Tenaza, R. (1985). "Songs of hybrid gibbons (*Hylobates lar* × *H muelleri*)," *Am. J. Primatol.* **8**, 249–253.
- Terleph, T. A., Malaivijitnond, S., and Reichard, U. H. (2015). "Lar gibbon (*Hylobates lar*) great call reveals individual caller identity," *Am. J. Primatol.* **821**, 811–821.
- Terleph, T. A., Malaivijitnond, S., and Reichard, U. H. (2016). "Age related decline in female lar gibbon great call performance suggests that call features correlate with physical condition," *BMC Evol. Biol.* **16**, 4.
- Terleph, T. A., Malaivijitnond, S., and Reichard, U. H. (2018). "An analysis of white-handed gibbon male song reveals speech-like phrases," *Am. J. Phys. Anthropol.* **166**, 649–660.
- Tibbetts, E. A., and Dale, J. (2007). "Individual recognition: It is good to be different," *Trends Ecol. Evol.* **22**, 529–537.
- Tumer, E. C., and Brainard, M. S. (2007). "Performance variability enables adaptive plasticity of 'crystallized' adult birdsong," *Nature* **450**, 1240–1244.
- Walsh, R. P., and Newbery, D. M. (1999). "The ecoclimatology of Danum, Sabah, in the context of the world's rainforest regions, with particular reference to dry periods and their impact," *Philos. Trans. R. Soc. London, Ser. B* **354**, 1869–1883.
- Wich, S. A., Schel, A. M., and de Vries, H. (2008). "Geographic variation in Thomas langur (*Presbytis thomasi*) loud calls," *Am. J. Primatol.* **70**, 566–574.
- Wiley, R. (1991). "Associations of Song Properties with Habitats for Territorial Oscine Birds of Eastern," *Am. Nat.* **138**, 973–993.
- Wilkins, M. R., Seddon, N., and Safran, R. J. (2012). "Evolutionary divergence in acoustic signals: Causes and consequences," *Trends Ecol. Evol.* **28**, 156–166.
- Wilson, D. R., Bitton, P., Podos, J., and Mennill, D. J. (2014). "Uneven sampling and the analysis of vocal performance constraints," *Am. Nat.* **183**, 214–228.
- Wright, T. F., Wilkinson, G. S., Wright, T. F., and Wilkinson, G. S. (2001). "Population genetic structure and vocal dialects in an amazon parrot Population genetic structure and vocal dialects in an amazon parrot," *Proc. R. Soc. B Biol. Sci.* **268**, 609–616.
- Yurk, H., Barrett-Lennard, L., Ford, J. K., and Matkin, C. (2002). "Cultural transmission within maternal lineages: Vocal clans in resident killer whales in southern Alaska," *Anim. Behav.* **63**, 1103–1119.
- Zollinger, S. A., Podos, J., Nemeth, E., Goller, F., and Brumm, H. (2012). "On the relationship between, and measurement of, amplitude and frequency in birdsong," *Anim. Behav.* **84**, e1–e9.