

Machine learning and statistical classification of birdsong link vocal acoustic features with phylogeny

Moises Rivera, Jacob A. Edwards, Mark E. Hauber, and Sarah M. N. Woolley

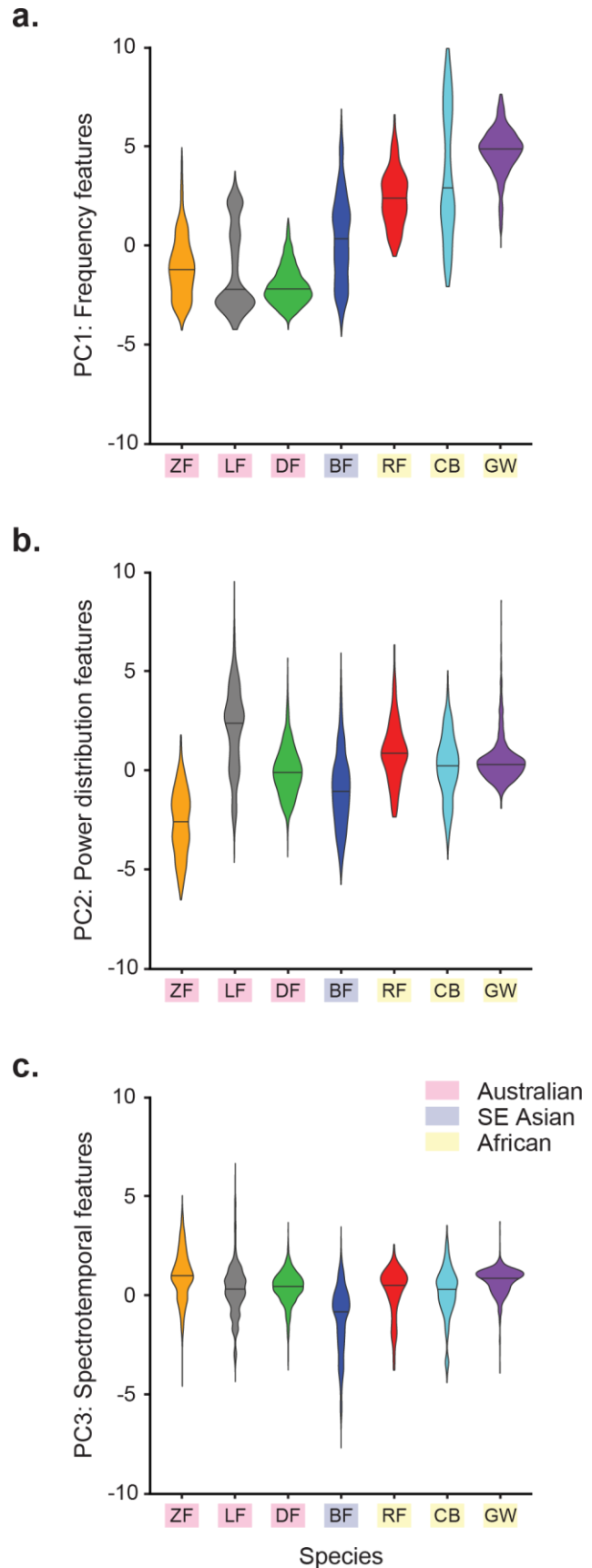
**Supplementary Table S1.** Features, descriptions (based on warbleR documentation; see Araya-Salas & Smith-Vidaurre, 2017), PC loadings, feature importance scores, Fowlkes-Mallows indices (*FMI*), EM-Mantel correlation coefficients (*r*), and phylogenetic inertia estimates (*R*<sup>2</sup>) for each of the 21 acoustic features measured from songs of the seven species. Superscripts denote features as components of frequency (<sup>1</sup>), power distribution (<sup>2</sup>), or spectrotemporal (<sup>3</sup>) features. Significant PC loadings ( $|\text{Loading}| > 0.218$ ) and feature importance scores (i.e., first six rank-ordered scores) appear in bold. Significant phylogenetic measures following false discovery rate p-value correction for multiple testing denoted by asterisks at corresponding alpha levels (\* < 0.05, \*\* < 0.01, \*\*\* < 0.001).

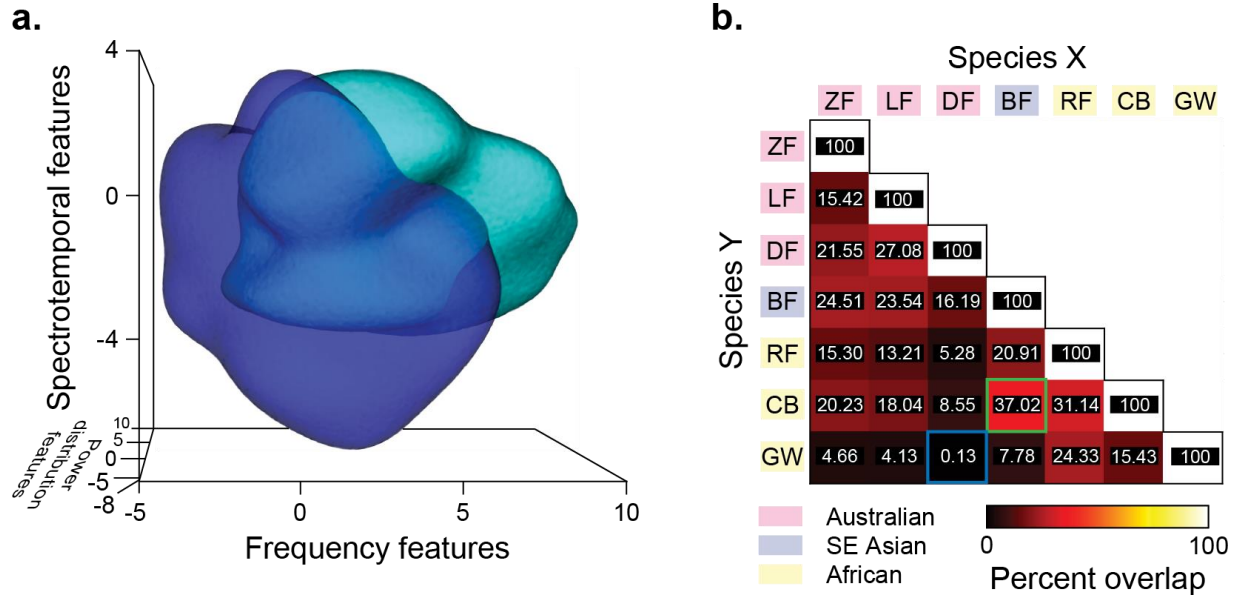
Acoustic Feature	Description	PC Loadings			Import. Score	Phylogen. Measures
		PC1	PC2	PC3		
Duration <sup>3</sup>	Duration (ms) of syllable.	0.05	-0.08	<b>0.41</b>	<b>5.12</b>	<i>FMI</i> = 0.31 <i>r</i> = 0.13 <i>R</i> <sup>2</sup> = 0.018
Mean Frequency <sup>1</sup>	Weighted average (by amplitude) of the syllable frequency power spectrum (FPS; Hz).	<b>0.30</b>	-0.14	0.02	<b>8.49</b>	<i>FMI</i> = 0.31 <i>r</i> = 0.26 <i>R</i> <sup>2</sup> = 0.066
Standard Deviation of Mean Frequency <sup>1,2</sup>	Standard deviation of the weighted mean frequency (Hz).	<b>-0.22</b>	<b>-0.23</b>	-0.12	2.77	<i>FMI</i> = 0.37 <i>r</i> = 0.50 <i>R</i> <sup>2</sup> = 0.25
Peak Mean Frequency <sup>1</sup>	Frequency (Hz) from the mean frequency spectrum with the highest amplitude.	<b>0.30</b>	-0.12	-0.05	1.49	<i>FMI</i> = 0.39 <i>r</i> = 0.42 <i>R</i> <sup>2</sup> = 0.18
Minimum Fundamental Frequency <sup>1</sup>	Minimum frequency (Hz) of the first harmonic band in the syllable.	<b>0.31</b>	0.06	-0.10	<b>14.25</b>	<i>FMI</i> = 0.62 <i>r</i> = 0.34 <i>R</i> <sup>2</sup> = 0.12
Maximum Fundamental Frequency <sup>1</sup>	Maximum frequency (Hz) of the first harmonic band in the syllable.	<b>0.30</b>	-0.02	-0.05	<b>7.44</b>	<i>FMI</i> = 0.39 <i>r</i> = 0.55 <i>R</i> <sup>2</sup> = 0.30
Mean Fundamental Frequency <sup>1</sup>	Average frequency (Hz) of the first harmonic band in the syllable.	<b>0.31</b>	0.02	-0.08	<b>5.33</b>	<i>FMI</i> = 0.39 <i>r</i> = 0.60 <i>R</i> <sup>2</sup> = 0.36
Minimum Dominant Frequency <sup>1</sup>	Minimum of dominant frequencies (Hz) across the syllable. Dominant frequency is defined as the frequency with peak amplitude in the FPS at each sampling of the syllable along the time wave.	<b>0.30</b>	0.06	0.01	1.78	<i>FMI</i> = 0.62 <i>r</i> = 0.80 <i>R</i> <sup>2</sup> = 0.64
Maximum Dominant Frequency <sup>1,2</sup>	Maximum of dominant frequencies (Hz) across the syllable.	<b>0.28</b>	<b>-0.23</b>	0.02	2.19	<i>FMI</i> = 0.37 <i>r</i> = 0.01 <i>R</i> <sup>2</sup> = 0.00015

Acoustic Feature	Description	PC Loadings			Import. Score	Phylogen. Measures
		PC1	PC2	PC3		
Mean Dominant Frequency <sup>1</sup>	Average of dominant frequencies (Hz) across the syllable.	<b>0.32</b>	-0.09	0.01	1.67	<i>FMI</i> = 0.37 <i>r</i> = 0.45 <i>R</i> <sup>2</sup> = 0.20
Dominant Frequency Range <sup>2</sup>	Range of dominant frequencies (Hz) across the syllable (MaxDom – MinDom).	0.04	<b>-0.34</b>	0.01	1.61	<i>FMI</i> = 0.15 <i>r</i> = - 0.36 <i>R</i> <sup>2</sup> = 0.13
Starting Dominant Frequency <sup>1</sup>	Dominant frequency (Hz) at the beginning of the syllable.	<b>0.28</b>	-0.10	-0.18	2.52	<i>FMI</i> = 0.73* <i>r</i> = 0.75 <i>R</i> <sup>2</sup> = 0.56
Ending Dominant Frequency <sup>1</sup>	Dominant frequency (Hz) at the end of the syllable.	<b>0.29</b>	0.00	0.11	1.37	<i>FMI</i> = 0.31 <i>r</i> = 0.51 <i>R</i> <sup>2</sup> = 0.26
Dominant Frequency Slope <sup>3</sup>	Change in dominant frequency across the syllable (Hz/ms). (EndDom – StartDom)/Duration.	0.01	0.11	<b>0.39</b>	3.30	<i>FMI</i> = 0.17 <i>r</i> = - 0.26 <i>R</i> <sup>2</sup> = 0.068
Skewness <sup>2</sup>	Asymmetry of the FPS. FPS may be left skewed ( <i>S</i> < 0), right skewed ( <i>S</i> > 0), or symmetrical ( <i>S</i> = 0).	0.01	<b>0.38</b>	0.14	1.45	<i>FMI</i> = 0.46 <i>r</i> = - 0.19 <i>R</i> <sup>2</sup> = 0.036
Kurtosis <sup>2</sup>	Peakedness of the FPS. FPS may be platikurtic ( <i>K</i> < 3), leptokurtic ( <i>K</i> > 3), or normal ( <i>K</i> = 3).	0.00	<b>0.33</b>	0.14	0.47	<i>FMI</i> = 0.26 <i>r</i> = - 0.26 <i>R</i> <sup>2</sup> = 0.068
Modulation Index <sup>3</sup>	Sum of absolute differences between adjacent dominant frequencies divided by the dominant frequency range. A modulation index of 1 means syllable is not modulated.	0.02	-0.21	<b>0.37</b>	3.55	<i>FMI</i> = 0.26 <i>r</i> = - 0.19 <i>R</i> <sup>2</sup> = 0.038
Spectral Entropy <sup>2</sup>	A measure of the energy distribution across frequencies. Noisy syllables (those whose energy is not "neatly" confound to frequency bands) will tend toward 1. Tonal syllables (those whose energy is confound to frequency bands) will tend toward 0. Described as "Wiener Entropy" in Tchernichovski et al. (2000).	-0.12	<b>-0.41</b>	-0.06	1.76	<i>FMI</i> = 0.46 <i>r</i> = 0.11 <i>R</i> <sup>2</sup> = 0.011

Acoustic Feature	Description	PC Loadings			Import. Score	Phylogen. Measures
		PC1	PC2	PC3		
Temporal Entropy <sup>3</sup>	A measure of the energy distribution across time. Noisy syllables (those with sustained amplitudes or high spectral continuity [Tchernichovski et al., 2000]) will tend toward 1. Quiet syllables (those with low spectral continuity) will tend toward 0.	0.04	-0.07	<b>0.55</b>	2.09	<i>FMI</i> = 0.37 <i>r</i> = 0.16 <i>R</i> <sup>2</sup> = 0.026
Entropy <sup>2,3</sup>	Measure of the two-dimensional spectrotemporal energy distribution of the syllable (Spectral Entropy · Temporal Entropy). Noisy syllables will tend toward 1. Tonal syllables will tend toward 0.	-0.08	<b>-0.37</b>	<b>0.31</b>	1.20	<i>FMI</i> = .67 <i>r</i> = .12 <i>R</i> <sup>2</sup> = 0.014
Spectral Flatness <sup>2</sup>	Measure of how closely the FPS energy distribution resembles a uniform distribution (i.e., white noise). Similar to SpectEnt. Noisy syllables will tend toward 1. Tonal syllables will tend toward 0.	-0.15	<b>-0.32</b>	-0.17	<b>10.57</b>	<i>FMI</i> = 0.46 <i>r</i> = 0.01 <i>R</i> <sup>2</sup> = 0.00017

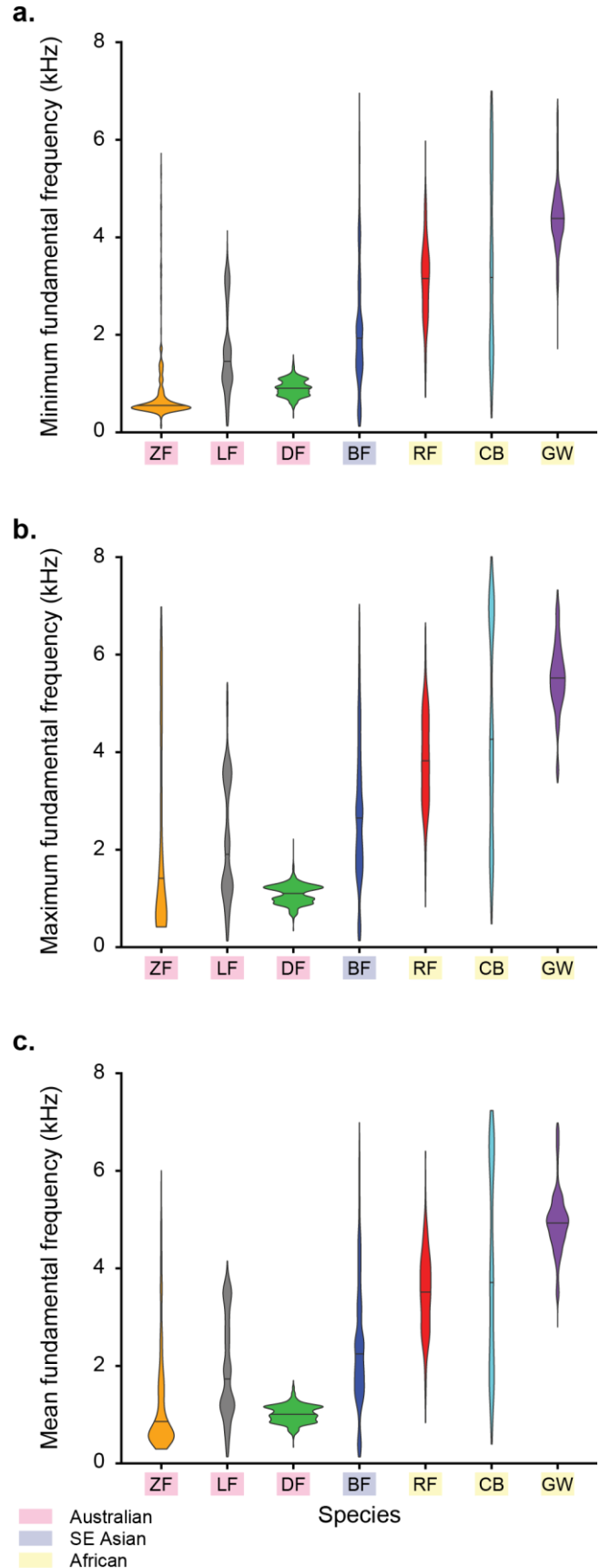
**Supplementary Figure S1.** Violin plots for species distributions along the 3 acoustic feature components. Horizontal line across each violin shows the distribution's median value. (a) Frequency features clustered species according to their biogeographic clades: Australian species (ZF, LF, DF) have low frequency values; African species (RF, CB, GW) have high frequency values range, and the Southeast Asian species (BF) spans the range between both clades. (b) Power distribution features do not cluster species according to biogeographic clades, on average: ZFs have the lowest values, consistent with the abundance of noisy (high entropy/spectral flatness), broadband harmonic syllables in their songs; LFs have the highest values, consistent with the abundance of low-harmonicity and low-noise (tonal) syllables in their song; all other species showed intermediate values (near PC2 = 0). (c) Spectrotemporal features do not cluster species according to biogeographic clades, on average. BFs had the lowest values, consistent with the abundance of very short durations and down-sweeping syllables in their songs. All other species showed similar values to each other.

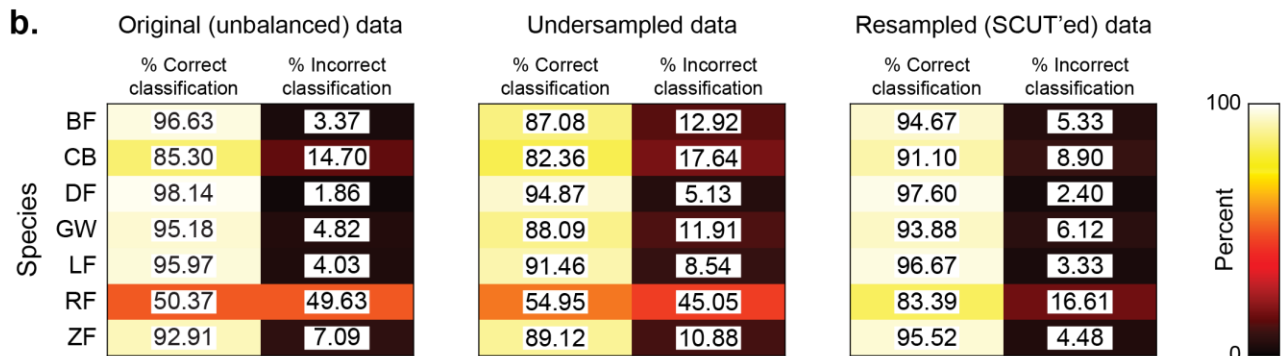
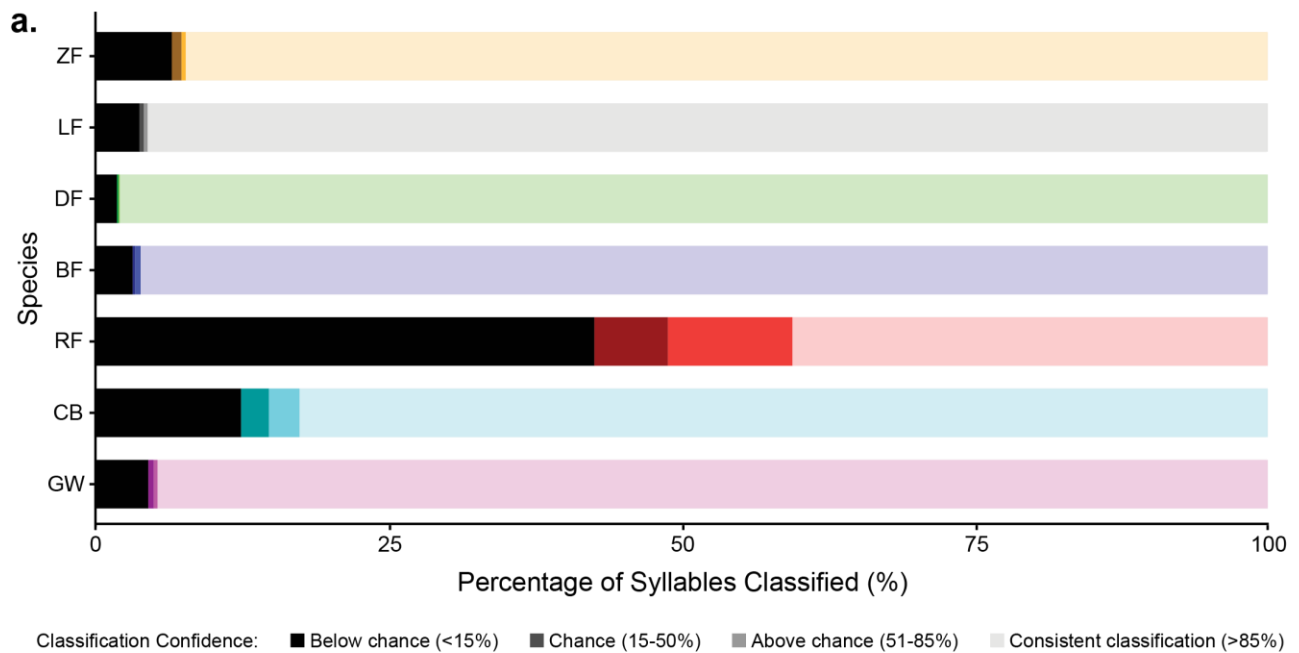




**Supplementary Figure S2.** Overlap of volumes in PC space differed in pairwise species comparisons. (a) The feature volumes of BF (blue) and CB (cyan) syllables show high overlap volume (> 37%). (b) Pair-wise Jaccard similarity indices between species. Jaccard similarity index provides a percentage of the total volume occupied by the two species that is shared between them. The highest overlap volume is between BF and CB (green square), while the lowest is between DF and GW (0.13%; blue square). High overlap between BF and CB is likely due to the moderate-to-high frequency harmonics and variable entropy in syllables – characteristics shared by both species. CB syllables tend to be higher in frequencies and have more upsweeps than BF syllables. Low overlap between DF and GW is due to the dissimilar syllable acoustics of these songs. DF syllables are harmonic with relatively low fundamental frequencies and slower spectrotemporal modulations, while GW syllables are tonal, with high fundamental frequencies and faster spectrotemporal modulations.

**Supplementary Figure S3.** Species distributions for syllable (a) minimum, (b) maximum, and (c) mean fundamental frequencies. Horizontal line across each violin shows the distribution's median value. Overall, Australian species (ZF, LF, DF) have lower fundamental frequencies than African species (RF, CB, GW), and BF spans the range between these two clades.

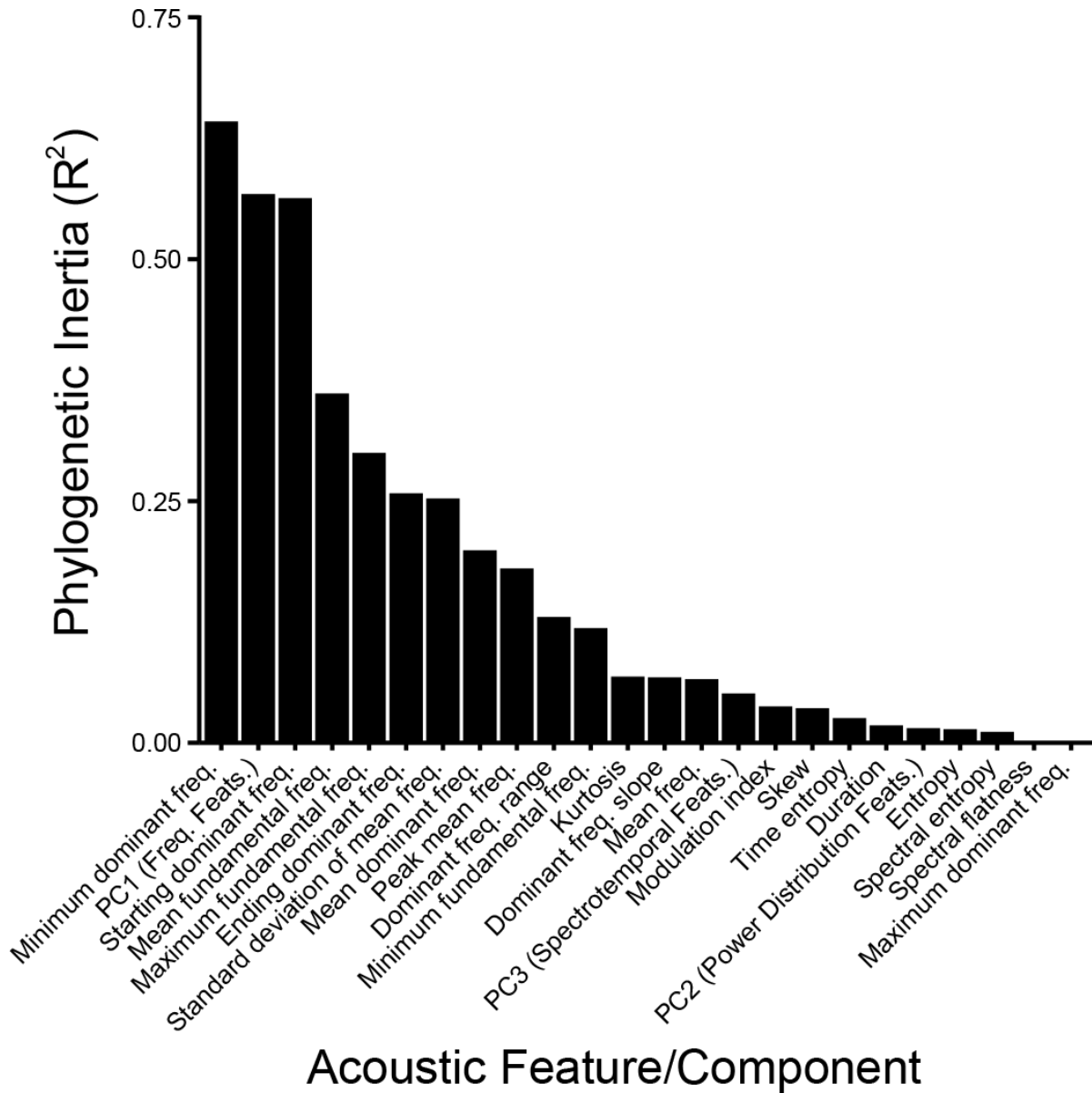




**Supplementary Figure S4.** Random forest model performance was robust across iterations and with unbalanced sample sizes. (a) Classification confidence for species syllables across 100 iterations of random forest in the 6-feature model. The percent of random forests that correctly classified each syllable across the 100 iterations (shading gradients in legend) represents the confidence (i.e., consistency) with which syllables are correctly labelled. We calculated the percentage of syllables (x-axis) that were classified at four levels of confidence within each species: Below Chance (<15%), Chance (15-50%), Above Chance (51-85%), Consistent Classification (>85%). 15% represents the minimum majority needed to assign a species label assuming uniform probabilities ( $1/(\text{no. species}) = 1/7 = 14.29 < 15$ ). 51% represents the minimum majority needed to assign a species label assuming dichotomous probability (i.e., 50% chance it will be correctly labelled, 50% chance it will not). 86% (i.e., >85%) represents the minimum majority needed to accept syllable labelling beyond significant consideration for any other single species (i.e., a syllable correctly labelled as belonging to *species X* 86% of the time could at most only be labelled as belonging to *species Y* 14% of the time, which is below our 15% threshold for chance). Overall, there was high consistency in the classification of syllables across random forest iterations for all species, with RF classifications being the least consistent, followed by CB. (b)



Percent correct and incorrect classifications for each species in the 6-feature model using the original (left), under-sampled (centered), and resampled (right) data sets. Classification accuracy using the under-sampled data set decreased for all species except RFs, which showed an increase from 50% to 55%. Classification accuracy using the resampled dataset showed changes in species consistent with their resampling: over-sampled species (CB, RF, ZF) increased in classifications, while under-sampled species (BF, DF) decreased in classification. Changes in LF (0.7% increase) and GW (1.3% decrease) classifications were inconsistent with their resampling but are minor compared to considerable increases in classifications of CB (6% increase) and RF (33% increase) syllables. Overall, our results suggest that although the random forest model is robust against unbalanced sample sizes among species, species-wise classifications could be improved with greater sample sizes.



**Supplementary Figure S5.** Rank-ordered phylogenetic inertia estimates for each of our acoustic features and components. The inertia estimate is obtained by calculating the coefficient of determination from the EM-Mantel test for each feature (i.e., the Mantel correlation statistic squared,  $R^2$ ). Acoustic features are plotted in order of decreasing phylogenetic inertia values, such that features toward the left of the plot have more influence of phylogeny in their evolutionary change (i.e., greater phylogenetic inertia), while features toward the right of the plot have less influence of phylogeny in their evolutionary change (i.e., lower phylogenetic inertia and greater plasticity for adaptation). Generally, we find that frequency features (including PC1) experience moderate-to-high phylogenetic constraint (e.g., phylogenetic inertia for PC1 = 57%), while power distribution and spectrotemporal features (including PCs 2 & 3) experience less phylogenetic constraint (i.e., phylogenetic inertia for PC2 = 2%, PC3 = 5%), suggesting these latter features may be suitable candidates for adaptation and the plasticity needed to avoid interspecific acoustic competition.