

Acoustic identification of Mexican bats based on taxonomic and ecological constraints on call design

Veronica Zamora-Gutierrez^{1,2*}, Celia Lopez-Gonzalez³, M. Cristina MacSwiney Gonzalez⁴, Brock Fenton⁵, Gareth Jones⁶, Elisabeth K. V. Kalko^{7,8}, Sebastien J. Puechmaille^{9,10}, Vassilios Stathopoulos¹¹ and Kate E. Jones^{2,12*}

¹Conservation Science Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK;

²Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK; ³Centro Interdisciplinario de Investigación para el Desarrollo Integral Regional (CIIDIR) Unidad Durango, Instituto Politécnico Nacional, Calle Sigma 119, Fraccionamiento 20 de Noviembre II, Durango 34220, Mexico; ⁴Centro de Investigaciones Tropicales, Universidad Veracruzana, Casco de la ExHacienda Lucas Martin, Privada de Araucarias Col. Periodistas, Xalapa 91019, Mexico; ⁵Department of Biology, Western University, London, ON N6A 5B7, Canada; ⁶School of Biological Sciences, University of Bristol, 24 Tyndall Avenue, Bristol BS8 1TQ, UK; ⁷Institute of Experimental Ecology, University of Ulm, Albert-Einstein-Allee 11, Ulm 89069, Germany; ⁸Smithsonian Tropical Research Institute, Balboa, Panama; ⁹Zoology Institute, Ernst-Moritz-Arndt University, Greifswald D-17489, Germany; ¹⁰School of Biology and Environmental Science, University College Dublin, Dublin 4, Ireland; ¹¹Department of Statistics, University of Warwick, Coventry CV4 7AL, UK; and ¹²Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK

Summary

1. Monitoring global biodiversity is critical for understanding responses to anthropogenic change, but biodiversity monitoring is often biased away from tropical, megadiverse areas that are experiencing more rapid environmental change. Acoustic surveys are increasingly used to monitor biodiversity change, especially for bats as they are important indicator species and most use sound to detect, localise and classify objects. However, using bat acoustic surveys for monitoring poses several challenges, particularly in megadiverse regions. Many species lack reference recordings, some species have high call similarity or differ in call detectability, and quantitative classification tools, such as machine learning algorithms, have rarely been applied to data from these areas.

2. Here, we collate a reference call library for bat species that occur in a megadiverse country, Mexico. We use 4685 search-phase calls from 1378 individual sequences of 59 bat species to create automatic species identification tools generated by machine learning algorithms (Random Forest). We evaluate the improvement in species-level classification rates gained by using hierarchical classifications, reflecting either taxonomic or ecological constraints (guilds) on call design, and examine how classification rate accuracy changes at different hierarchical levels (family, genus and guild).

3. Species-level classification of calls had a mean accuracy of 66%, and the use of hierarchies improved mean species-level classification accuracy by up to 6% (species within families 72%, species within genera 71.2% and species within guilds 69.1%). Classification accuracy to family, genus and guild-level was 91.7%, 77.8% and 82.5%, respectively.

4. The bioacoustic identification tools we have developed are accurate for rapid biodiversity assessments in a megadiverse region and can also be used effectively to classify species at broader taxonomic or ecological levels. This flexibility increases their usefulness when there are incomplete species reference recordings and also offers the opportunity to characterise and track changes in bat community structure. Our results show that bat bioacoustic surveys in megadiverse countries have more potential than previously thought to monitor biodiversity changes and can be used to direct further developments of bioacoustic monitoring programs in Mexico.

Key-words: acoustic identification, guild, hierarchical classification, machine learning, Neotropical, random forest, whispering bats

Introduction

Effective conservation depends on our ability to define, measure and track ecological communities through time and

space (Magurran *et al.* 2010). Although biodiversity monitoring programmes are critical to assess the impact of anthropogenic change, many are biased towards high latitude, temperate countries (Collen *et al.* 2009). Megadiverse countries (e.g. Indonesia, Mexico, Zaire) cover only 34% of the Earth surface, yet they harbour 70% of the world's biodiversity and are undergoing rapid environmental degradation

*Correspondence authors. E-mails: zamora.gtz@gmail.com and kate.e.jones@ucl.ac.uk

(Mittermeier, Mittermeier & Robles-Gil 1997). In spite of the great conservation opportunity these hot spot regions offer, biodiversity monitoring programmes are often lacking, causing considerable knowledge gaps.

Bioacoustic surveys, especially for bats, are increasingly used to survey and monitor biodiversity responses to anthropogenic change (Jones *et al.* 2013; Amorim *et al.* 2014). Echolocating bats use sound to detect, localise and classify objects (Schnitzler, Moss & Denzinger 2003), making them detectable both remotely and non-invasively. Bats are also ideal biodiversity indicators since they have a wide range of ecological traits, different tolerances to environmental variables and play key roles in ecosystems (Jones *et al.* 2009; Russo & Jones 2015). However, using bat acoustics as a monitoring tool poses several challenges, especially in megadiverse and tropical regions (Walters *et al.* 2013). First, in spite of the growing efforts to create more bat call reference recording libraries, tropical and megadiverse regions have rarely been included in such initiatives. This is compounded by recording method heterogeneity (e.g. full spectrum, frequency division, heterodyne), which makes compiling comprehensive libraries difficult (Walters *et al.* 2013). Such poor and uneven coverage of intra- and interspecific variation makes identification of bat calls for these regions challenging.

Secondly, although it is possible to identify many bat species based on their calls, phylogenetic relatedness, ecological similarities and call plasticity have led to overlapping structures and high call similarity among and within species in some groups (Obrist 1995; Jones & Teeling 2006). For example, species may have similar calls within families and genera (Jung, Kalko & von Helversen 2007; Jung, Molinari & Kalko 2014), and ecological guild membership may also reflect foraging and echolocation behaviour (e.g. aerial insectivores, gleaners) (Denzinger & Schnitzler 2013). An additional challenge is that bat species differ in detectability of their calls. Aerial insectivores typically produce loud calls of high intensity and low frequency, whereas 'whispering' bats (including many bats in the families Phyllostomidae, Natalidae, and Thyropteridae) often produce low-intensity, high-frequency calls (Griffin 1958). However, recent findings suggest that some 'whispering' bat calls are more detectable than previously thought. For example, *Macrophyllum macrophyllum* and *Artibeus jamaicensis* can emit calls as loud as those of many aerial insectivores (Brinkløv, Kalko & Surlykke 2009). *Otonycteris hemprichii*, a passive gleaner, can also operate as an aerial hawk and can adjust its call intensity depending on foraging mode even while flying in the same habitat type (Hackett, Korine & Holderied 2014).

Thirdly, although acoustic species identification tools for different species are developing rapidly (e.g. European bats Walters *et al.* 2012; birds Stowell & Plumbley 2014), they remain rare for megadiverse regions. The immense amount of data obtained from acoustic monitoring can be daunting, and automatic analytical tools are extremely useful in analysing such data (Walters *et al.* 2013). Bat call identification tools have been mainly developed using multivariate statistical techniques such as discriminant function analysis (e.g. Vaughan, Jones & Harris 1997; Russo & Jones 2002;

Avila-Flores & Fenton 2005; MacSwiney, Clarke & Racey 2008) or machine learning algorithms (e.g. Skowronski & Harris 2006; Walters *et al.* 2012), the latter generally providing higher species-level classification accuracy (Armitage & Ober 2010; Britzke *et al.* 2011; Keen *et al.* 2014). Machine learning algorithms have mostly been applied to classify data at one level of categorisation (e.g. species) and have rarely incorporated hierarchical information to aid classification accuracy (e.g. species within families or orders). Hierarchical classification approaches have been shown to improve general species classification accuracy for European bat calls up to 13% (Parsons & Jones 2000; Walters *et al.* 2012). Assigning taxa to classes within a hierarchy may reduce model complexity and minimise misclassifications outside their hierarchy (Vens *et al.* 2008). However, if an erroneous hierarchy is applied, then classification errors are added cumulatively across different levels, leading to a reduction in classification accuracy.

A hierarchical classification approach may be useful to classify calls to broader classes (e.g. genera, families or guilds) when reference material is missing for species, or where discrimination at species level is difficult, for example where there is high call variability within species or a high overlap of call parameters between species. Although identification to species is most desirable, monitoring the status of the same recognisable signal over time without specific identification may be sufficient in some situations (Redgwell *et al.* 2009; Armitage & Ober 2010). Finding alternatives to species-level studies is needed in megadiverse areas, which usually face considerable financial and data constraints but are a priority for rapid conservation assessments.

Here, we collate a reference call library for bat species that occur in a megadiverse country to create acoustic identification tools using machine learning algorithms. We focus on Mexico because it contains one of the highest number of species in the world and has one of the highest rates of species extinction and habitat loss (Myers *et al.* 2000; Brooks *et al.* 2002). We also evaluate the improvement in species-level classification rates gained by using hierarchical classifications reflecting either taxonomic or ecological constraints on call design. Our results show that accurate bioacoustic identification tools can be developed for rapid biodiversity assessments in megadiverse regions where hierarchies generally improve species-level classifications. These tools can also be used effectively to classify calls at broader levels, so increasing the usefulness of the tool when there are incomplete species reference recordings.

Materials and methods

REFERENCE CALL LIBRARY

We collated reference search-phase echolocation calls for bat species that occur in Mexico through a combination of fieldwork and donated material. Fieldwork was conducted in central and northern Mexico from June 2012 to May 2013 at 35 sites (Fig. 1a). Bats were caught with mist nets and identified to species level using field keys (Reid 1997;

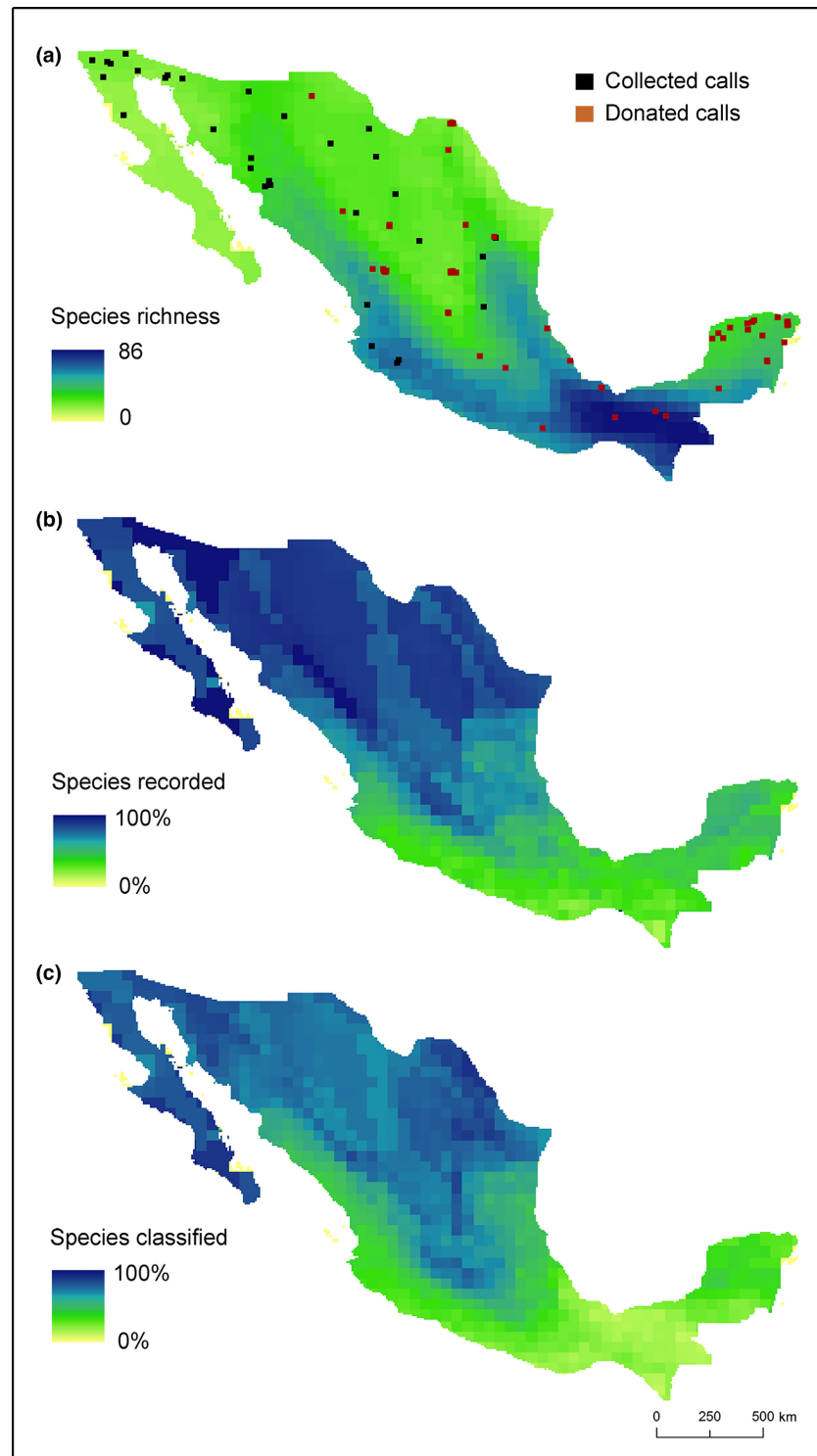


Fig. 1. Spatial coverage of the number of species recorded in Mexico using a grid size of 50 km², where (a) shows recording locations in solid squares ($n = 91$) overlaid with bat species richness, (b) proportion of species recorded compared to potential species richness in each grid, and (c) proportion of species used in the classifiers compared to potential species richness in each grid. A gradient of light green to dark blue indicates higher number of species and higher percentages. Black solid squares represent collection sites which were sampled in this study, and red solid squares represent collection sites of donated material.

Medellín, Arita & Sánchez 2008), before being released. Full spectrum, real-time recordings were made from all individuals in the habitat in which they were captured using a Pettersson D1000x detector, sampling rate 500 kHz, high-pass filter off (Pettersson Elektronik AB, Uppsala, Sweden). Files were saved in WAV format on a flash card. We obtained 907 recordings of 39 species from six families (see Table S1). Additionally, 1403 full spectrum recordings of bat calls from 87 species that occur in Mexico were donated by colleagues, giving a total of 2310 recordings (each recording was assumed to contain one

individual call sequence) from 92 species in eight families (68% of species and 100% of families of bats occurring in Mexico). These recordings were obtained from bats released in different ways using several different real-time or time-expanded full-spectrum detectors, and in a range of habitats across species' distributions (including localities outside Mexico) (Table S2). The inclusion of call variation in the data set avoids generating biases for any particular recording situation or method (Walters *et al.* 2013) and provides the acoustic identification tools with more flexibility and generality (see Walters *et al.* 2012).

Taxonomy followed Simmons (2005), but because of taxonomic changes since 2005, we assume that Natalidae contains only one species, *Natalus stramineus* (López-Wilchis *et al.* 2012). Data from *Molossus sinaloae* and the new species *M. alvarezii* (González-Ruiz, Ramírez-Pulido & Arroyo-Cabrales 2011) were analysed together as *M. sinaloae* because most of the material was recorded prior to the description of the new species. As some species are hard to identify in the field, we only used the material which were confidently identified. To examine the taxonomic and geographic coverage of the reference call library within Mexico, distribution maps were downloaded from the IUCN mammal assessments (IUCN 2012) and species richness within each 50-km² grid cell was estimated by overlaying and counting how many of those range maps overlap in each grid cell (Hawths Tools, Beyer 2004). We then calculated the proportion of species both recorded and used in our classifiers from out of those potentially distributed in each cell.

ACOUSTIC IDENTIFICATION TOOLS

We visually inspected all recorded sequences using the sound analysis software BATSOUND PRO v.3.31b (Pettersson Elektronik AB, Uppsala, Sweden) to remove non-search-phase calls. We distinguished search-phase calls from approach-phase and terminal-phase calls as these phase shifts are characterised by a decrease in call duration and interval, and increase in repetition rate (Schnitzler & Kalko 2001). Social calls were distinguished from echolocation calls by their duration, frequency and pattern of change over time, with social calls being more sporadic and often of a lower frequency range (Fenton 2003). In addition, bats were recorded in situations that significantly minimised the presence of social calls and approach and end-phase echolocation calls (e.g. recorded in open spaces upon release). We then automatically extracted and parameterised search-phase calls using the in-built algorithms in Sonobat v.3 (Szewczak 2010) (following methods in Walters *et al.* 2012). For species which used harmonics, we used measurements from the call used as the main harmonic. We measured a total of 21 064 search-phase echolocation calls from 1692 sequences and 85 species in eight families, with each sequence assumed to be from a different individual. Material recorded in Mexico contained 16 344 calls, 1187 sequences from 65 species in seven families across 91 different localities (Fig. 1a).

We used Random Forest (RF) models (RANDOMFOREST package, Liaw & Wiener 2002) to train the classifiers, rejecting species that had less than five sequences. RF models consist of a collection or ensemble of decision tree classifiers where each classifier is randomly built using a bootstrapped sample of the training data set (Breiman 2001). Each classifier is estimated based on probabilities using a selection of the predictor variables (in our case call parameters) that best separate the classes of interest (e.g. species, families) at different branching splits or nodes in the tree. RF model classifications are then derived from averages of the tree ensembles. RF models possess several advantages over other machine learning algorithms as they are not affected by heteroscedasticity or distributional errors in the data, are not sensitive to outliers or irrelevant variables, can deal with mixed data and missing variables and are relatively simple to train using reasonable computational resources (Olden, Lawler & Poff 2008). We selected 27 of the relevant call parameter variables (following methods in Walters *et al.* 2012) extracted and parameterised by Sonobat (Table S3), and ran a grid search to find the mtry value (optimal number of variables to be randomly sampled at each node). This value was allowed to range from 2 to 10, in steps of one. Each forest was grown to 2000 trees, and the final mtry value and number of trees were selected for their highest accuracy. The final set of parameters used was 1000 trees and an mtry value of

three. We used the coefficient of the Gini impurity index (used by the RF models to select the most informative variables at nodes during training), as an indicator of call parameter variable importance (Breiman 2001).

We trained four different RF model classifiers: Classifier 1 – species level without a hierarchy; Classifier 2 – species level within a family hierarchy (see call examples in Fig. S1a–f); Classifier 3 – species level within a genus hierarchy (see call examples in Fig. S1g); and Classifier 4 – species level within a guild hierarchy, following definitions of guilds from Denzinger & Schnitzler (2013) (see call examples in Fig. S2a–e): Guild 1 represented open-space aerial foragers; Guild 2 – edge-space aerial foragers; Guild 3 – edge-space trawling foragers; Guild 4 – narrow-space flutter detecting foragers; and Guild 5 – narrow-space passive gleaning foragers; and Guild 6 – narrow-space passive/active gleaning foragers. Guild 7 – narrow-space active gleaning foragers was not included in the study because of the lack of reference material.

We used fivefold cross-validation to assess the accuracy of all four RF classifiers and assigned the individual calls into the fivefolds by sequence rather than individual calls (Stathopoulos *et al.* 2014). This procedure ensured that calls from the same individual (i.e. sequence) were not used in the same training and testing run of the cross-validation to avoid overfitting. We set a maximum of 100 calls per species for Classifier 1 and a minimum of 20 calls per species for Classifiers 2, 3 and 4, as a compromise between maximising the number of calls and balancing the data sets, since RF classifiers tend to be biased towards the majority class (species, genus, family or guild with the highest number of training calls) (Chen, Liaw & Breiman 2004). Only the highest quality calls were selected from each sequence (determined by the signal to noise ratio given by Sonobat), until the selected number of calls was reached. However, for some species with smaller sample sizes, we continued selecting calls from sequences in descending order of quality until we had used all available data or reached the number of calls allowed (Table S4). The number of calls selected per sequence was a compromise between maximising the number of calls and avoiding overfitting the RF models. Sample sizes after this selection process were 4685 calls and 1378 sequences from eight families, 32 genera and 59 species that occur in Mexico. See Fig. S3 for an outline of the analytical procedure.

As we used recordings from locations from both inside and outside of Mexico, we checked that the variation in call parameters recorded in locations outside of Mexico did not impact species classification accuracy. To investigate this, we compared model accuracy using the four classifiers of two data sets consisting of 47 species recorded from locations inside Mexico and the same species recorded from all locations. We found very little difference in classification accuracy between the two data sets. Classifier 1 had the biggest difference in classification accuracy, albeit with only 1.5% reduction in correct classification rates (67.1% and 65.6% for inside Mexico and for all locations, respectively). We therefore used recordings from outside Mexico to complement species with less than five Mexican sequences. All analyses were performed in R version 3.0.2 (R Development Core Team 2013).

Results

DATA BASE COVERAGE

Our collated library of echolocation call recordings covered 69% of the species, 79% of the genera and 100% of the families occurring in Mexico. Data of high enough quality to build the automatic identification tools covered 43% of the species, 51% of the genera and 100% of the families (Table S5). There

was generally a good representation of species for the identification tools within genera and families (>50%), except for Phyllostomidae, where only 19% of the species were represented. Species coverage was more comprehensive within the central and northern parts of Mexico for both the library and identification tools (Fig. 1b,c).

ACOUSTIC IDENTIFICATION TOOLS

Overall 16 of the 27 parameters used to train the models contributed most to all classifiers (based on a score >30 for the Gini Coefficient from the RF models) (Table S6, Fig. 2a–d). Although different parameters were important for each hierarchy, the most important overall were Fc Characteristic call frequency (kHz), Fctr Frequency at the centre of the call duration (kHz), FLed Frequency of the ledge (kHz), StartF Frequency at the start of a call (kHz), HFreq Highest call frequency (kHz) and FMPwr Frequency of the maximum call amplitude (kHz) (Fig. 2a–d, see Table S3 for further variable definitions).

Overall mean species-level classification accuracies for Mexican bat species varied across the four classifiers between 66.0% (Classifier 1: species level without a hierarchy) and 72.0% (Classifier 2: species level within a family hierarchy), with Classifiers 3 (species level with a genus hierarchy) and 4 (species level with a guild hierarchy) having accuracies of 71.2% and 69.1%, respectively (Table 1). Across all classifiers, on average

the highest classification accuracies were found for species within families Natalidae (100%), Mormoopidae (94.6%), Thyropteridae (81.5%), and Emballonuridae (77.7%), with the lowest found within Noctilionidae (70.4%), Molossidae (67%), Vespertilionidae (51.5%) and Phyllostomidae (51.4%) (Fig. 3). Phyllostomid species were mostly misclassified with other phyllostomids or with vespertilionids, whereas vespertilionids were commonly misclassified with other vespertilionids or with molossids (Table 1). For the ecological classifier, species within Guild 4 (narrow-space flutter detecting foragers) (100%), Guild 3 (edge-space trawling foragers) (74.6%) and Guild 1 (open-space aerial foragers) (63.8%) had on average the highest classification rates. The lowest average classification rates for species were found within in the gleaners (Guild 5 58.5% and Guild 6 57.7%) and Guild 2 (edge-space aerial foragers) (54.5%) (Fig. 4).

Classification accuracy at different hierarchical levels was highest at family level with a mean of 91.7% across all families (Table 1, Fig. 3), where Natalidae and Mormoopidae had the highest classification accuracies (100% and 97.3%, respectively). Noctilionidae had the lowest classification accuracy (72.8%) and was frequently misclassified as Molossidae (17% of the calls). Genus-level mean classification accuracy was 77.8% across all genera (Table 1), *Natalus* (Natalidae) and *Rhynchonycteris* (Emballonuridae) had the highest classification accuracies (100%), and 18 genera had accuracies >80% (Fig. 3). The genus *Myotis* yielded a classification accuracy of

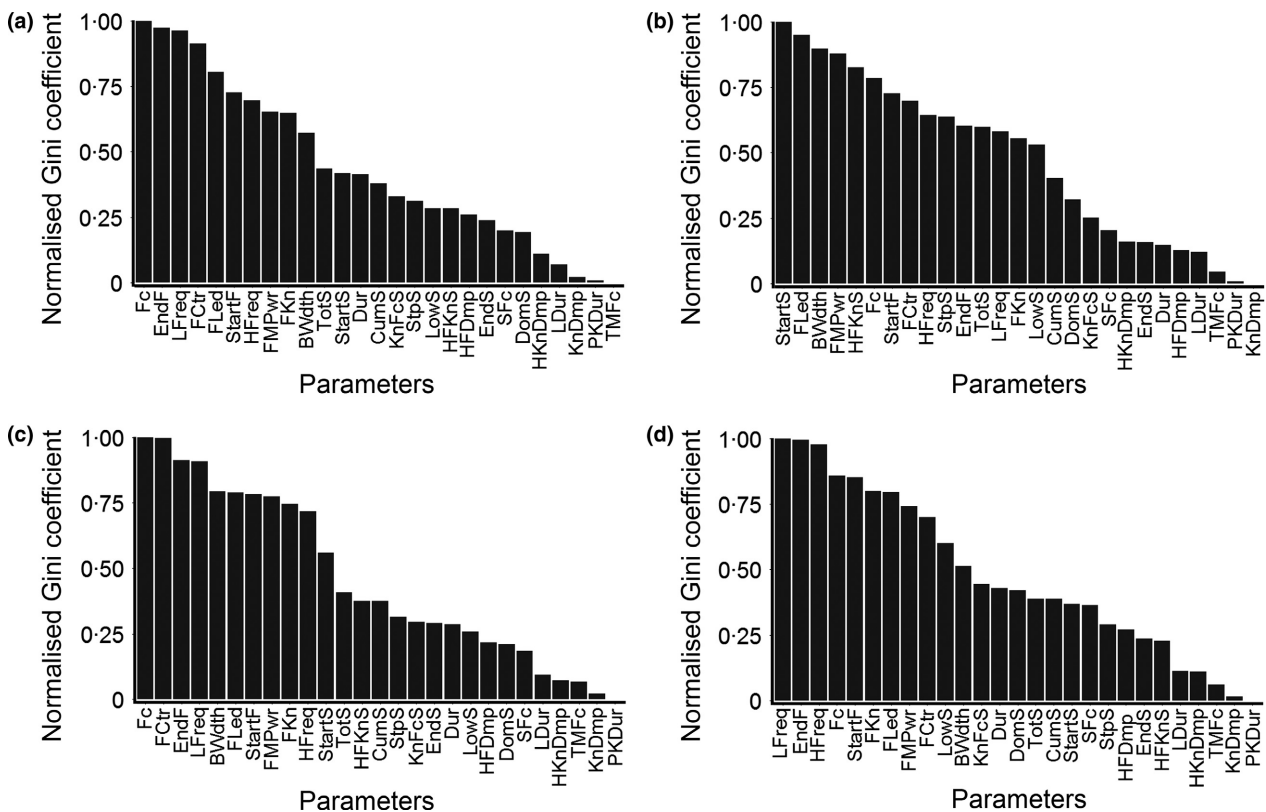


Fig. 2. Echolocation call parameters ($n = 27$) selected to build each Random Forest classifier ranked by Gini Coefficient where (a) Classifier 1: species level without a hierarchy; (b) Classifier 2: species level within a family hierarchy; (c) Classifier 3: species level within a genus hierarchy; and (d) Classifier 4: species level within a guild hierarchy. See Table S3 for parameter definitions.

Table 1. Comparison of classification accuracies of four acoustic classifiers for Mexican bat species ($n = 59$ species). Where Classifier 1 represents species level without a hierarchy; Classifier 2: species level within families; Classifier 3: species level within genera; Classifier 4: species level within guild. Misclassification represents those classes commonly misclassified with each other for each classifier and level, where Phyllo Phyllostomidae; Vesp Vespertilionidae; Molo Molossidae; and Noct Noctilionidae.

Classifier	Level	Mean accuracy %	Accuracy range %	% of classes \geq 80% accuracy	% of classes \leq 60% accuracy	Misclassifications
1	Species	66	4.2–100	29	41	Species of Phyllo with themselves or Vesp; Vesp with themselves or Molo
2	Species	72	0–100	32	44	Species within families
	Family	91.7	72.8–100	88	0	Noct with Molo
3	Species	71.2	0–100	36	37	Species within genera
	Genus	77.8	0–100	56	16	Phyllo with other Phyllo and Vesp genera; and Vesp with Phyllo and Molo genera
4	Species	69.1	4.5–100	25	44	Species within guilds
	Guild	82.5	68–100	50	0	Guild 5 with Guild 6; Guild 6 with Guild 5

73.8%, with two species over >80% (*Myotis thysanodes* and *Myotis keaysi*) and only four with <50%. Genera with the lowest classification accuracies (<50%) were in the Phyllostomidae and Vespertilionidae (Fig. 3). Phyllostomids were mostly misclassified as other phyllostomids, while vespertilionids were misclassified as other vespertilionids and molossids. Mean guild-level classification accuracy was 82.5% across all guilds (Table 1, Fig. 4). Guild 4 (narrow-space flutter detecting foragers) had the highest classification accuracy (100%), followed by Guild 6 (88.3%) although 6% of these calls were misclassified with Guild 5. Guild 5 had the worst classification accuracy (68%), and 18% of calls were misclassified as Guild 6 (Fig. 4).

Discussion

We have collated the most extensive bat acoustic library for a megadiverse region (all families and over half of the species occurring in Mexico) and developed the most comprehensive bat acoustic automated species-level classifiers to date. The mean species-level classification accuracy rate of 66–72% (depending on which hierarchy is chosen) is reasonable given the high level of call similarity of the bat species in this area (Walters *et al.* 2013). The species-level classifiers also contain a large variation in accuracy rates, where some species are classified to >80% accuracy (species of Emballonuridae, Mormoopidae, Natalidae and Thyropteridae), with the poorest results overall from species of Vespertilionidae and Phyllostomidae. This suggests that acoustic monitoring may be more feasible focusing on a few species whose calls can be reliably classified.

The bat call library and classifiers incorporate both extensive geographic (from 9 countries within the species range of Mexican bats) and intraspecific variation in call types (e.g. the classifiers were trained on the different search-phase echolocation call types found within molossid species, Jung, Molinari & Kalko 2014). However, the species-level classifiers have a very low coverage of Phyllostomidae and results should be interpreted with caution. It has been traditionally assumed that whispering bats, which include all phyllostomids, echolocate at intensities that were too low for the inclusion of these species in acoustic studies. However, recent field studies of their echolocation behaviour challenged these assumptions about their

echolocation characteristics (Brinkløv, Kalko & Surlykke 2009; Hackett, Korine & Holderied 2014). Future work should focus on collecting more reference material for the family, to better assess its potential for acoustic monitoring programmes.

Our classifiers will be the most accurate in regions where there is a higher coverage of the species present, such as the less species-rich arid and semi-arid regions of Mexico. These ecosystems (e.g. xerophytic scrubland and grasslands) cover at least 40% of the territory (Rzedowski 2006), and together with other North American dry lands, support some of the biggest concentrations of mammalian abundance, because bats can form colonies of several millions of individuals (O'Shea & Bogan 2003). These bat populations can provide important ecosystem services such as pollination and control of insect populations (Cleveland *et al.* 2006; Munguía-Rosas *et al.* 2009). These important arid and semi-arid environments are increasingly threatened by environmental changes (Villers-Ruiz & Trejo-Vázquez 2003; Rodríguez-Estrella 2007), and future efforts should focus on these arid areas where there are considerable information gaps.

Our species-level classifier mean accuracy was similar to that of previous studies of bats for species shared with this study (Mexico – MacSwiney, Clarke & Racey 2008; Stathopoulos *et al.* 2014; West Indies – Pio *et al.* 2010; United States – Skowronski & Harris 2006; Britzke *et al.* 2011). However, our classification accuracies were slightly lower for some species compared with previous work. This is a consequence of the higher number of classes (species) included in our classifiers compared to all previous studies. Higher numbers of species increase the similarity in the call parameters of several species. For example, we included 26 vespertilionids and 8 *Myotis* species, compared to six vespertilionids and one *Myotis* in MacSwiney, Clarke & Racey (2008) and nine vespertilionids and two *Myotis* in Stathopoulos *et al.* (2014). Our study nearly triples the number of species used compared to any other quantitative bat call classification study in the Americas or any other megadiverse tropical region in the world. We also included a wide range of ecological, technological and methodological variation in the training data set, which on one hand increases the classification challenge, but on the other makes the classifiers more robust to real-world recording situations. In spite of

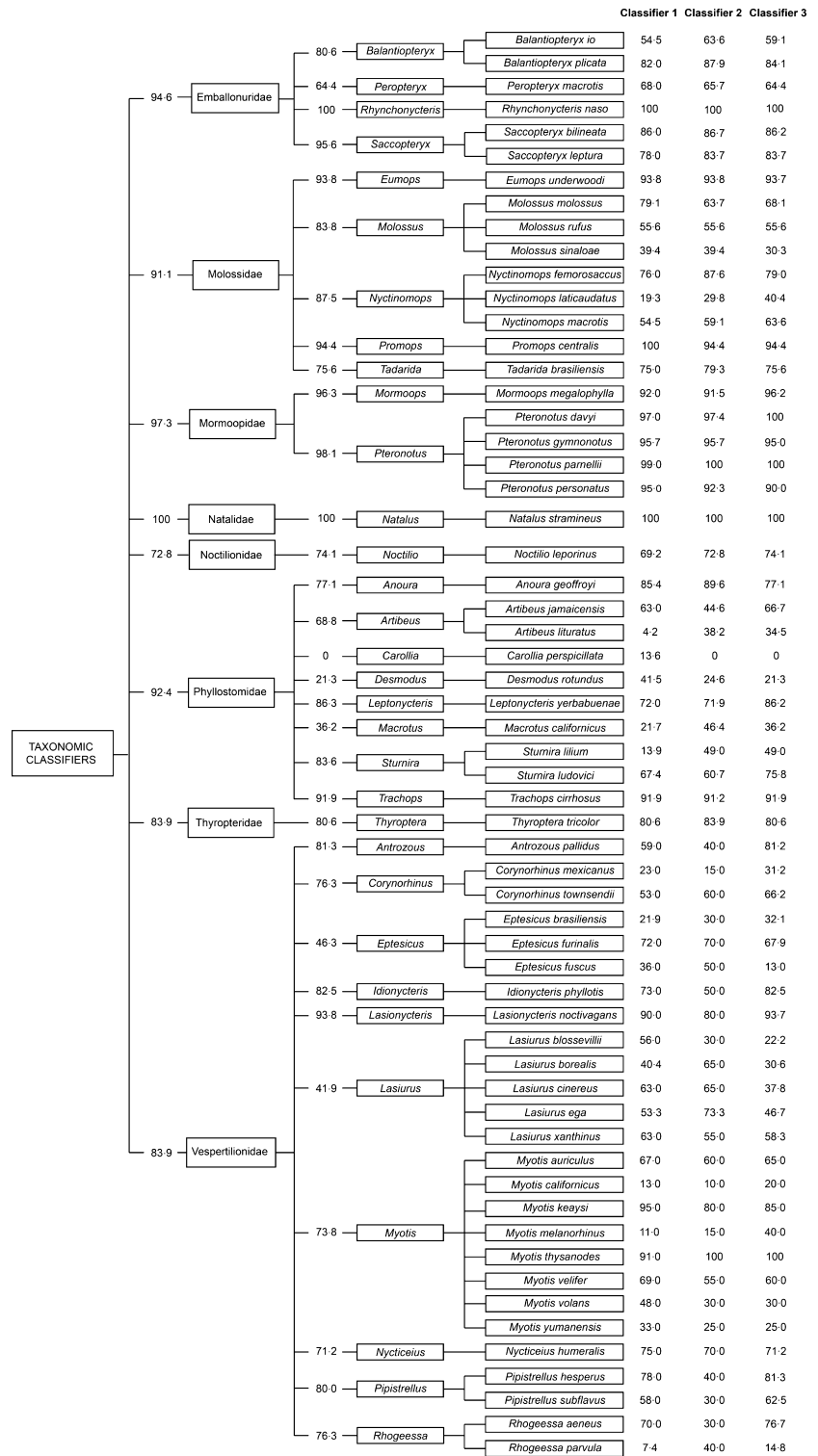


Fig. 3. Random Forest percentage classification accuracies obtained for the taxonomic classifiers (Classifiers 1–3). Species-level accuracies are shown at the end of each branch for Classifiers 1, 2 and 3. Classification accuracies per family and genus are shown in the middle of each branch ($n = 59$ species). See Table S4 for species acronym definitions.

the great difference in the number of species used here, we also obtained higher classification accuracies to species level and better mean accuracies than previous studies (e.g. Pio *et al.* 2010; Stathopoulos *et al.* 2014).

Our use of taxonomic and ecological guild hierarchies improved mean species-level classification rates. By using hierarchical classification approaches, the number of final classes is considerably reduced and misclassifications are

limited to classes within the respective hierarchy (Vens *et al.* 2008). Mean species-level classification accuracies were most improved using a family hierarchy, closely followed by genera (72% and 71%, respectively), although not all species improved their accuracies (contrary to other studies - Parsons & Jones 2000; Walters *et al.* 2012). The genus-level hierarchy produced the highest number of species-level classifications with >80% accuracy, but for many genera, not

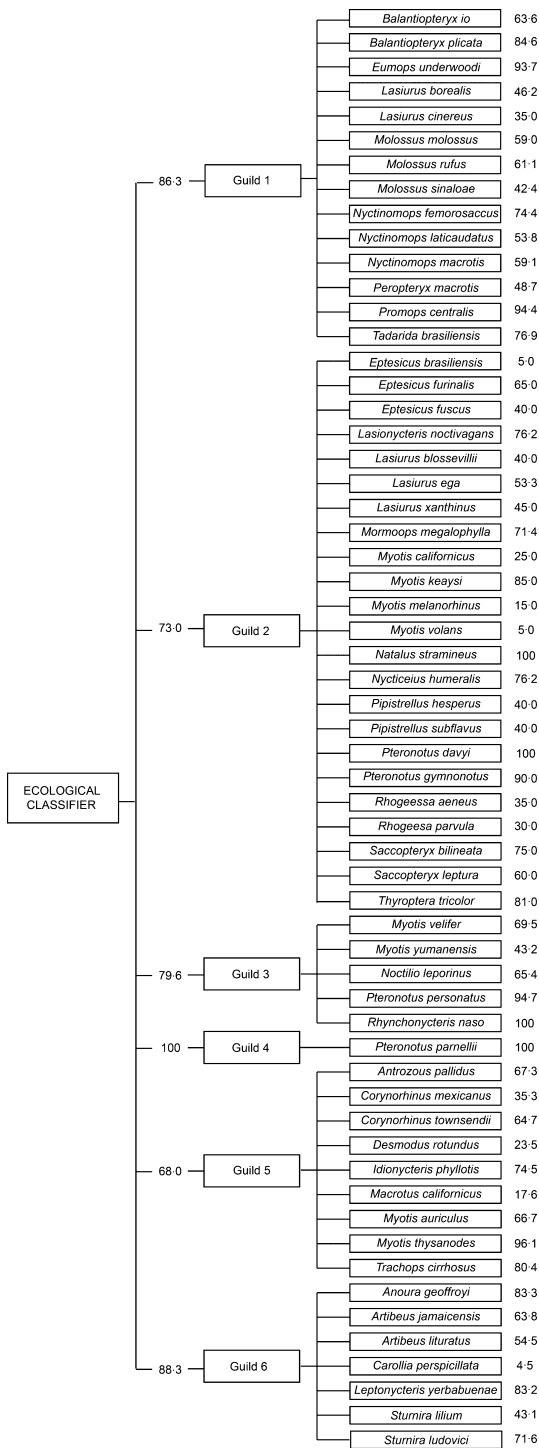


Fig. 4. Random Forest percentage classification accuracies obtained for the ecological guild classifier (Classifier 4). Species-level accuracies are shown at the end of each branch. Classification accuracies per guild are shown in the middle of each branch ($n = 59$ species). Guild 1 – open-space aerial foragers; Guild 2 – edge-space aerial foragers; Guild 3 – edge-space trawling foragers; Guild 4 – narrow-space flutter detecting foragers; Guild 5 – narrow-space passive gleaning foragers and Guild 6 – narrow-space passive/active gleaning foragers. See Table S4 for species acronym definitions.

all species were included in the analysis and genus-level taxonomic names can be subject to rapid changes (Simmons 2005). This may suggest that using a genus-level hierarchy

may be more problematic than a family hierarchy, especially with incomplete reference material. For example, accuracy may decrease as more species are included, whereas variation within a family may be already adequately represented. In contrast, classification to genus level may be more helpful to reduce the number of options of possible misclassifications inside the hierarchy and further methods for call identification could then be applied (e.g. visual inspection).

Although we found species-level classification rates within an ecological guild-level hierarchy were worse than species-level classification rates within either taxonomic hierarchy, classification of calls to guild-level performed well and could provide a useful alternative to taxonomic-level classifiers. Gleaners, in particular the speciose family Phyllostomidae, are the most abundant and diverse in bat communities in the Americas, yet poorly represented in acoustic libraries. Our results at family and guild-level suggest that there is a good potential for accurate acoustic identification of gleaners. As more sensitive microphones with better signal/noise ratios become available, the detectability of these species will improve, as will the potential for monitoring them acoustically. Guilds 5 and 6, representing gleaning foragers, were frequently confused with each other, so these should be grouped into one class, since the main difference among them is how they use other non-echolocation cues to forage (Denzinger & Schnitzler 2013).

Acoustic analysis techniques are evolving rapidly, and there is a growing tendency to replace classifications based on parameters extraction with those of whole signal analysis. However, applications of these approaches have mainly focused on bird and marine mammal acoustics (e.g. Ren *et al.* 2009; Damoulas *et al.* 2010) and most bat acoustic classification tasks still represent classifications with a few parameters and further classify them using manual or nonparametric techniques. Such whole signal analyses in bat acoustics are growing (Obriest, Boesch & Flückiger 2004; Skowronski & Harris 2006; Stathopoulos *et al.* 2014) but should be further explored. Such exploration of new approaches requires adequate reference material collected in a systematic way, controlling for variation introduced by the use of different methods, and we strongly encourage further efforts to collect comprehensive reference bat call libraries.

APPLICATIONS

Standardised identification tools such as these offer the opportunity for objective and repeatable identifications of monitoring ‘units’ to identify changes in populations, distributions or community structures through time and space. Furthermore, hierarchical approaches offer the flexibility to adapt the identification tools to the purpose of the study or monitoring programme and the geographic and taxonomic coverage of the reference material available. Although the accuracy reached for some groups might not be sufficient for studies targeting their particular species (e.g. *Myotis* spp.), the hierarchical classifiers can act as filters for large amounts of data. The use of hierarchies considerably reduces the list of species to which an

unknown call could belong, thus making detailed inspections and further validations more feasible.

Hierarchical classifications, in particular at family level, could help reduce the costs of monitoring tropical bat communities, which is crucial due to the limited funding these regions often devote to conservation efforts. Despite the relatively poor classification accuracies to species level within the guild-level hierarchy, classification to guild-level could be used to rapidly characterise ensemble/environment associations or to track changes in community structure. The hierarchical approach may be improved through the use of regional classifiers which allow the reduction of the number of classes and the improvement of classification accuracy. However, such an approach should be used with caution as least known species or those with expanding ranges could be ignored.

Conclusions

Our study shows that there is more potential for bat acoustic monitoring in megadiverse countries than previously considered. Hierarchies considerably reduced the complexity of call identification at different levels and provided sufficient confidence in the classification of unknown calls into higher taxonomic levels and ecological guilds. While the classifiers did not provide high classification accuracies for several species, they did offer the opportunity to have objective and repeatable identification of monitoring 'units' to implement in national acoustic monitoring programmes.

Acknowledgements

This study was financially supported by CONACYT (No. 310731), Cambridge Commonwealth European and International Trust (No. 301879989), The Rufford Foundation (No. 12059-1), American Society of Mammalogists, Bat Conservation International, Idea Wild and The Whitmore Trust to V.Z.G and Engineering and Physical Sciences Research Council (EPSRC) Grant EP/K015664/1 to K.E.J. and V.S. We are deeply grateful to all the people that provided their help with logistical support in the field, and we also thank Juan Cruzado Cortes, Michael Barataud and David Jacobs for donated material. A collecting permit was granted by SEMARNAT, Mexico (No. 03374).

Data accessibility

Complete call measurements for each of the 4685 search-phase calls from 1378 individual sequences of 59 bat species used to create the classification tools and R scripts have been uploaded to DataDryad (<http://datadryad.org/resource/doi:10.5061/dryad.760r8>).

References

Amorim, F., Carvalho, S.B., Honrado, J. & Rebelo, H. (2014) Designing optimized multi-species monitoring networks to detect range shifts driven by climate change: a case study with bats in the north of Portugal. *PLoS ONE*, **9**, e87291.

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.

Avila-Flores, R. & Fenton, M.B. (2005) Use of spatial features by foraging insectivorous bats in a large urban landscape. *Journal of Mammalogy*, **86**, 1193–1204.

Beyer, H.L. (2004) Hawth's analysis tools for ArcGIS. URL <http://www.spatialecology.com/htools> [accessed 1 July 2015]

Breiman, L. (2001) Random forests. *Journal of Machine Learning Research*, **45**, 5–32.

Brinklov, S., Kalko, E.K.V. & Surlykke, A. (2009) Intense echolocation calls from two 'whispering' bats, *Artibeus jamaicensis* and *Macrophyllum macrophyllum* (Phyllostomidae). *The Journal of Experimental Biology*, **212**, 11–20.

Britzke, E.R., Duchamp, J.E., Murray, K.L., Swihart, R.K. & Robbins, L.W. (2011) Acoustic identification of bats in the eastern United States: a comparison of parametric and nonparametric methods. *Journal of Wildlife Management*, **75**, 660–667.

Brooks, T.M., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Rylands, A.B., Konstant, W.R. *et al.* (2002) Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology*, **16**, 909–923.

Chen, C., Liaw, A. & Breiman, L. (2004) Using random forest to learn imbalanced data. Dept. Statistics, Univ. California, Berkeley, CA, Tech. Rep.666.

Cleveland, C.J., Betke, M., Federico, P., Frank, J.D., Hallam, T.G., Horn, J.D.L. Jr *et al.* (2006) Economic value of the pest control service provided by Brazilian free-tailed bats in south-central Texas. *Frontiers in Ecology and the Environment*, **4**, 238–243.

Collen, B., Loh, J., Whitmee, S., McRae, L., Amin, R. & Baillie, J.E.M. (2009) Monitoring change in vertebrate abundance: the living planet index. *Conservation Biology*, **23**, 317–327.

Damoulas, T., Henry, S., Farnsworth, A., Lanzone, M. & Gomes, C. (2010) Bayesian classification of flight calls with a novel dynamic time warping kernel. *Proceedings of the 2010 Ninth International Conference on Machine Learning and Applications*, pp. 424–429. IEEE Computer Society, Washington, DC.

Denzinger, A. & Schnitzler, H.-U. (2013) Bat guilds, a concept to classify the highly diverse foraging and echolocation behaviors of microchiropteran bats. *Frontiers in Physiology*, **4**, 1–15.

Fenton, M.B. (2003) Eavesdropping on the echolocation and social calls of bats. *Mammal Review*, **33**, 193–204.

González-Ruiz, N., Ramírez-Pulido, J. & Arroyo-Cabrales, J. (2011) A new species of mastiff bat (Chiroptera: Molossidae: *Molossus*) from Mexico. *Mammalian Biology*, **76**, 461–469.

Griffin, D.R. (1958) *Listening in the Dark*. Yale University Press, New Haven, Connecticut.

Hackett, T.D., Korine, C. & Holderied, M.W. (2014) A whispering bat that screams: bimodal switch of foraging guild from gleaning to aerial hawking in the desert long-eared bat. *The Journal of Experimental Biology*, **217**, 3028–3032.

IUCN. (2012) IUCN Red List of Threatened Species version 2012.14. URL <http://www.iucnredlist.org> [accessed 1 July 2015]

Jones, G. & Teeling, E.C. (2006) The evolution of echolocation in bats. *Trends in Ecology and Evolution*, **21**, 149–156.

Jones, G., Jacobs, D.S., Kunz, T.H., Willig, M.R. & Racey, P.A. (2009) Carpe noctem: the importance of bats as bioindicators. *Endangered Species Research*, **8**, 93–115.

Jones, K.E., Russ, J., Bashta, A.-T., Bilhari, Z., Catto, C., Csoz, I. *et al.* (2013) Indicator Bats Program: a system for the global acoustic monitoring of bats. *Biodiversity Monitoring and Conservation: Bridging the Gaps between Global Commitment and Local Action* (eds B. Collen, N. Pettorell, J.E.M. Baillie & S. Durant), pp. 211–247. Wiley-Blackwell, London.

Jung, K., Kalko, E.K.V. & von Helversen, O. (2007) Echolocation calls in Central American emballonurid bats: signal design and call frequency alternation. *Journal of Zoology*, **272**, 125–137.

Jung, K., Molinari, J. & Kalko, E.K.V. (2014) Driving factors for the evolution of species-specific echolocation call design in new world free-tailed bats (Molossidae). *PLoS ONE*, **9**, e85279.

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.

Liaw, A. & Wiener, M. (2002) Classification and regression by random forest. *R News*, **2**, 18–22.

López-Wilchis, R., Guevara-Chumacero, L.M., Pérez, N.Á., Juste, J., Ibáñez, C. & Barriga-Sosa, I.D.L.A. (2012) Taxonomic status assessment of the Mexican populations of funnel-eared bats, genus *Natalus* (Chiroptera: Natalidae). *Acta Chiropterologica*, **14**, 305–316.

MacSwiney, M.C., Clarke, F.M. & Racey, P.A. (2008) What you see is not what you get: the role of ultrasonic detectors in increasing inventory completeness in Neotropical bat assemblages. *Journal of Applied Ecology*, **45**, 1364–1371.

Magurran, A.E., Baillie, S.R., Buckland, S.T., Dick, J.M., Elston, D.A., Scott, E.M., Smith, R.I., Somerfield, P.J. & Watt, A.D. (2010) Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. *Trends in Ecology and Evolution*, **25**, 574–582.

- Medellín, R.A., Arita, H.T. & Sánchez, O. (2008) *Identificación de los Murciélagos de México, Clave de Campo*, 2nd edn. Instituto de Ecología, Universidad Nacional Autónoma de México-CONABIO, Mexico D.F.
- Mittermeier, R.A., Mittermeier, C.G. & Robles-Gil, P. (1997) *Megadiversity: Earth's Biologically Wealthiest Nations*. Cemex, Mexico, D.F.
- Munguía-Rosas, M.A., Sosa, V.J., Ojeda, M.M. & De-Nova, J.A. (2009) Specialization clines in the pollination systems of agaves (Agavaceae) and columnar cacti (Cactaceae): a phylogenetically controlled meta-analysis. *American Journal of Botany*, **96**, 1887–1895.
- Myers, N., Mittermeier, R.A., da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.
- Obrist, M.K. (1995) Flexible bat echolocation: the influence of individual, habitat and conspecifics on sonar signal design. *Behavioral Ecology and Sociobiology*, **36**, 207–219.
- Obrist, M.K., Boesch, R. & Flückiger, P.F. (2004) Variability in echolocation call design of 26 Swiss bat species: consequences, limits and options for automated field identification with a synergetic pattern recognition approach. *Mammalia*, **68**, 307–322.
- Olden, J.D., Lawler, J.J. & Poff, N.L. (2008) Machine learning methods without tears: a primer for ecologists. *The Quarterly Review of Biology*, **83**, 171–193.
- O'Shea, T.J. & Bogan, M.A. (2003) *Monitoring trends in bat populations of the United States and territories: Problems and prospects*.
- Parsons, S. & Jones, G. (2000) Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. *The Journal of Experimental Biology*, **203**, 2641–2656.
- Pio, D.V.V., Clarke, F.M., MacKie, I. & Racey, P.A. (2010) Echolocation calls of the bats of Trinidad, West Indies: is guild membership reflected in echolocation signal design? *Acta Chiropterologica*, **12**, 217–229.
- R Development Core Team. (2013) R: a language and environment for statistical computing. Vienna. URL <http://www.R-project.org> [accessed 1 July 2015]
- Redgwell, R.D., Szwczak, J.M., Jones, G. & Parsons, S. (2009) Classification of echolocation calls from 14 species of bat by support vector machines and ensembles of neural networks. *Algorithms*, **2**, 907–924.
- Reid, F. (1997) *A Field Guide to the Mammals of Central America and Southeast Mexico*. Oxford University Press, New York.
- Ren, Y., Johnson, M.T., Clemins, P., 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.
- Rodríguez-Estrella, R. (2007) Land use changes affect distributional patterns of desert birds in the Baja California peninsula, Mexico. *Diversity and Distributions*, **13**, 877–889.
- Russo, D. & Jones, G. (2002) Identification of twenty-two bat species (Mammalia: Chiroptera) from Italy by analysis of time-expanded recordings of echolocation calls. *Journal of Zoology*, **258**, 91–103.
- Russo, D. & Jones, G. (2015) Bats as bioindicators: an introduction. *Mammalian Biology*, **80**, 157–158.
- Rzedowski, J. (2006) *Vegetación de México*. CONABIO, Mexico, D.F.
- Schnitzler, H.-U. & Kalko, E.K.V. (2001) Echolocation by insect-eating bats. *BioScience*, **51**, 557–569.
- Schnitzler, H.-U., Moss, C.F. & Denzinger, A. (2003) From spatial orientation to food acquisition in echolocating bats. *Trends in Ecology and Evolution*, **18**, 386–394.
- Simmons, N.B. (2005) Order Chiroptera. *Mammal Species of the World* (eds D.E. Wilson & D.M. Reeder), pp. 312–529. John Hopkins University Press, Baltimore.
- Skowronski, M.D. & Harris, J.G. (2006) Acoustic detection and classification of microchiroptera using machine learning: lessons learned from automatic speech recognition. *The Journal of the Acoustical Society of America*, **119**, 1817–1833.
- Stathopoulos, V., Zamora-Gutierrez, V., Jones, K.E. & Girolami, M. (2014) Bat call identification with gaussian process multinomial probit regression and a dynamic time warping kernel. *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics*, **33**, 913–921.
- Stowell, D. & Plumbley, M.D. (2014) Large-scale analysis of frequency modulation in birdsong data bases. *Methods in Ecology and Evolution*, **5**, 901–912.
- Szwczak, J.M. (2010) Sonobat v.3. URL www.sonobat.com [accessed 1 July 2015]
- Vaughan, N., Jones, G. & Harris, S. (1997) Identification of British bat species by multivariate analysis of echolocation call parameters. *Bioacoustics*, **7**, 189–207.
- Vens, C., Struyf, J., Schietgat, L., Dzeroski, S. & Blockeel, H. (2008) Decision trees for hierarchical multi-label classification. *Machine Learning*, **73**, 185–214.
- Villers-Ruiz, L. & Trejo-Vázquez, I. (2003) El cambio climático y la vegetación en México. *México: una Visión Hacia el Siglo XXI. El Cambio Climático en México* (ed. G. Garcia-Carlos), pp. 57–72. INE, UNAM, US Country Studies Program, México, D.F.
- Walters, C.L., Freeman, R., Collen, A., Dietz, C., Fenton, M.B., Jones, G. et al. (2012) A continental-scale tool for acoustic identification of European bats. *Journal of Applied Ecology*, **49**, 1064–1074.
- Walters, C., Collen, A., Lucas, T., Mroz, K., Sayer, C. & Jones, K.E. (2013) Challenges of using bioacoustics to globally monitor bats. *Bat Evolution, Ecology, and Conservation* (eds R.A. Adams & S.C. Pedersen), pp. 479–500. Springer, New York.

Received 10 January 2016; accepted 22 February 2016
Handling Editor: Robert Freckleton

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Fig. S1. Spectrograms showing the interspecific variability of representative search-phase echolocation calls within taxonomic groups used for the classifiers.

Fig. S2. Spectrograms showing interspecific variability of representative search-phase echolocation calls within ecological guilds used for the classifiers.

Fig. S3. Schematic representation of the protocol used to build the classifiers.

Table S1. Metadata for the search-phase echolocation calls collected during field work in Mexico from June 2012 to May 2013.

Table S2. Metadata for the search-phase echolocation calls donated for this study.

Table S3. Definitions of the 27 call parameters extracted by Sonobat v.3 used for training the Random Forest classifiers.

Table S4. Number of classes included in each hierarchy of the classifiers and number of calls used in the training process.

Table S5. Taxonomic coverage of the bat call library within each family for the number of genera and species recorded / used in the classifiers.

Table S6. Descriptive statistics (mean and standard deviation) for the 16 most important call parameters ranked by Random Forest Gini Coefficient measured by Sonobat for the 59 species.