



University of  
**Salford**  
MANCHESTER

# Species undersampling in tropical bat surveys : effects on emerging biodiversity patterns

Meyer, CFJ, Aguiar, LMS, Aguirre, JF, Baumgarten, J, Clarke, FM, Cosson, J, Estrada Villegas, S, Fahr, J, Faria, D, Furey, N, Henry, M, Jenkins, RKB, Kunz, TH, Cristina MacSwiney González, M, Moya, I, Pons, J, Racey, PA, Rex, K, Sampaio, EM, Stoner, KE, Voigt, CC, von Staden, D, Weise, CD and Kalko, EKV

<http://dx.doi.org/10.1111/1365-2656.12261>

<b>Title</b>	Species undersampling in tropical bat surveys : effects on emerging biodiversity patterns
<b>Authors</b>	Meyer, CFJ, Aguiar, LMS, Aguirre, JF, Baumgarten, J, Clarke, FM, Cosson, J, Estrada Villegas, S, Fahr, J, Faria, D, Furey, N, Henry, M, Jenkins, RKB, Kunz, TH, Cristina MacSwiney González, M, Moya, I, Pons, J, Racey, PA, Rex, K, Sampaio, EM, Stoner, KE, Voigt, CC, von Staden, D, Weise, CD and Kalko, EKV
<b>Type</b>	Article
<b>URL</b>	This version is available at: <a href="http://usir.salford.ac.uk/id/eprint/37806/">http://usir.salford.ac.uk/id/eprint/37806/</a>
<b>Published Date</b>	2014

USIR is a digital collection of the research output of the University of Salford. Where copyright permits, full text material held in the repository is made freely available online and can be read, downloaded and copied for non-commercial private study or research purposes. Please check the manuscript for any further copyright restrictions.

For more information, including our policy and submission procedure, please contact the Repository Team at: [usir@salford.ac.uk](mailto:usir@salford.ac.uk).

# **Species undersampling in tropical bat surveys: effects on emerging**

## **biodiversity patterns**

Christoph F. J. Meyer<sup>1,2\*</sup>, Ludmilla M. S. Aguiar<sup>3</sup>, Luis F. Aguirre<sup>4,5</sup>, Julio Baumgarten<sup>6</sup>, Frank M. Clarke<sup>7</sup>, Jean-François Cosson<sup>8</sup>, Sergio Estrada Villegas<sup>9</sup>, Jakob Fahr<sup>2,10,11</sup>, Deborah Faria<sup>6</sup>, Neil Furey<sup>7</sup>, Mickaël Henry<sup>12</sup>, Richard K. B. Jenkins<sup>7,13,14,15</sup>, Thomas H. Kunz<sup>16</sup>, M. Cristina MacSwiney Gonzalez<sup>7,17</sup>, Isabel Moya<sup>5</sup>, Jean-Marc Pons<sup>18</sup>, Paul A. Racey<sup>19</sup>, Katja Rex<sup>20</sup>, Erica M. Sampaio<sup>2</sup>, Kathryn E. Stoner<sup>21</sup>, Christian C. Voigt<sup>20</sup>, Dietrich von Staden<sup>2</sup>, Christa D. Weise<sup>22</sup>, Elisabeth K. V. Kalko<sup>2,23</sup>

<sup>1</sup>Centro de Biologia Ambiental, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

<sup>2</sup>Institute of Experimental Ecology, University of Ulm, Albert-Einstein-Allee 11, 89069 Ulm, Germany

<sup>3</sup>Departamento de Zoologia, Universidade de Brasília, Brasília, Distrito Federal 70910-900, Brazil

<sup>4</sup>Centro de Biodiversidad y Genética, Universidad Mayor de San Simón, Casilla Postal 538, Cochabamba, Bolivia

<sup>5</sup>Centro de Estudios en Biología Teórica y Aplicada, Programa para la Conservación de los Murciélagos de Bolivia, La Paz, Bolivia

<sup>6</sup>Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz, CEP 45650-000 Ilhéus, Bahia, Brazil

<sup>7</sup>Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen, AB24 2TZ, UK

<sup>8</sup>INRA, UMR CBGP, Campus International de Baillarguet CS 30016, F-34988 Montferrier-sur-Lez, France

<sup>9</sup>Department of Biology, McGill University, 1205 Docteur Penfield, Montréal, Canada H3A 1B1

<sup>10</sup>Department of Migration and Immuno-ecology, Vogelwarte Radolfzell, Max Planck Institute for Ornithology, Am Obstberg 1, D-78315 Radolfzell, Germany

<sup>11</sup>Division of Evolutionary Biology, Zoological Institute, TU Braunschweig, D-38106 Braunschweig, Germany

<sup>12</sup>INRA, UMR 406 Abeilles & Environnement, Site Agroparc, F-84914 Avignon, France

<sup>13</sup>Madagasikara Voakajy, B. P. 5181, Antananarivo, Madagascar

<sup>14</sup>DICE, School of Anthropology and Conservation, The University of Kent, Canterbury, Kent, CT2 7NR, UK

<sup>15</sup>School of Environment, Natural Resources and Geography, Bangor University, Bangor, LL57 2UW, UK

<sup>16</sup>Department of Biology, Center for Ecology and Conservation Biology, Boston University, Boston, MA 02215, USA

<sup>17</sup>Centro de Investigaciones Tropicales, Universidad Veracruzana, Veracruz C.P. 91019, Mexico

<sup>18</sup>UMR 7205, Department Systématique et Evolution, Muséum National d'Histoire Naturelle, 55 Rue Buffon, BP51, F-75005 Paris, France

<sup>19</sup>Centre for Ecology and Conservation, University of Exeter in Cornwall, Penryn, TR10 9EZ, UK

<sup>20</sup>Leibniz-Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, 10315 Berlin, Germany

<sup>21</sup>Department of Biological and Health Sciences, Texas A&M University-Kingsville, MSC 158,

700 University Boulevard, Kingsville, TX 78363, USA

<sup>22</sup>Bat Conservation International, P.O. Box 162603, Austin, TX 78716, USA

<sup>23</sup>Smithsonian Tropical Research Institute, P.O. Box 0843-03092, Balboa, Panama

\*Corresponding author.

Postal address: Centro de Biologia Ambiental, Faculdade de Ciências, Universidade de Lisboa,

Campo Grande C2, 1749-016 Lisboa, Portugal

E-mail: [cmeyer@fc.ul.pt](mailto:cmeyer@fc.ul.pt), Phone: 00351-217-500000 ext. 22516

**Running title:** Sample representativeness in tropical bat surveys

**Article type:** Standard paper

**Number of words:** 8.329

**Number of tables:** 3

**Number of figures:** 3

## 1 **Summary**

2 **1.** Undersampling is commonplace in biodiversity surveys of species-rich tropical assemblages in  
3 which rare taxa abound, with possible repercussions for our ability to implement surveys and  
4 monitoring programs in a cost-effective way.

5 **2.** We investigated the consequences of information loss due to species undersampling (missing  
6 subsets of species from the full species pool) in tropical bat surveys for the emerging patterns of  
7 species richness and compositional variation across sites.

8 **3.** For 27 bat assemblage datasets from across the tropics, we used correlations between original  
9 datasets and subsets with different numbers of species deleted either at random, or according to  
10 their rarity in the assemblage, to assess to what extent patterns in species richness and  
11 composition in data subsets are congruent with those in the initial dataset. We then examined to  
12 what degree high sample representativeness ( $r \geq 0.8$ ) was influenced by biogeographic region,  
13 sampling method, sampling effort, or structural assemblage characteristics.

14 **4.** For species richness, correlations between random subsets and original datasets were strong ( $r$   
15  $\geq 0.8$ ) with moderate (ca. 20%) species loss. Bias associated with information loss was greater  
16 for species composition; on average ca. 90% of species in random subsets had to be retained to  
17 adequately capture among-site variation. For non-random subsets, removing only the rarest  
18 species (on average ~10% of the full dataset) yielded strong correlations ( $r > 0.95$ ) for both  
19 species richness and composition. Eliminating greater proportions of rare species resulted in  
20 weaker correlations and large variation in the magnitude of observed correlations among  
21 datasets.

22 **5.** Species subsets that comprised ca. 85% of the original set can be considered reliable  
23 surrogates, capable of adequately revealing patterns of species richness and temporal or spatial

24 turnover in many tropical bat assemblages. Our analyses thus demonstrate the potential as well  
25 as limitations for reducing survey effort and streamlining sampling protocols, and consequently  
26 for increasing the cost-effectiveness in tropical bat surveys or monitoring programs. The  
27 dependence of the performance of species subsets on structural assemblage characteristics (total  
28 assemblage abundance, proportion of rare species), however, underscores the importance of  
29 adaptive monitoring schemes and of establishing surrogate performance on a site-by-site basis  
30 based on pilot surveys.

31

32 **Key-words:** biodiversity surveys; Chiroptera; cost-effectiveness; representative sampling;  
33 species rarity; species subsamples

34

## 35 **Introduction**

36 Recent studies suggest that the indicator potential and surrogacy value of single taxa is  
37 usually poor (Kessler *et al.* 2011; Larsen *et al.* 2012) and that tropical biodiversity surveys  
38 should aim to include as many different taxa as possible under given financial and logistical  
39 constraints. Selection of ‘high-performance indicator taxa’ for monitoring purposes requires  
40 consideration not only of the ecological value of a taxon, but also of the practical feasibility and  
41 cost-effectiveness with which it can be surveyed (Gardner *et al.* 2008; Kessler *et al.* 2011). The  
42 monetary cost and time allocation necessary to survey a given taxon, undoubtedly, are two of the  
43 main constraints faced in monitoring programs, which therefore typically seek to obtain the  
44 information required for the least cost and within the shortest time (Gardner *et al.* 2008;  
45 McDonald-Madden *et al.* 2010).

46           Designing a survey program that is at the same time statistically robust and cost-effective  
47 requires balancing opposing limitations – maximizing sample representativeness (i.e. trying to  
48 enumerate all or most species in an assemblage), versus maximizing statistical power by  
49 increasing the number of sites surveyed at the expense of survey comprehensiveness. The effects  
50 of reducing cost and sampling effort may be particularly significant when those species that are  
51 most difficult to sample are also the rare ones. Species that are locally rare abound in species-  
52 rich assemblages in the humid tropics (e.g. Coddington *et al.* 2009), usually rendering attempts at  
53 achieving sampling completeness in biodiversity surveys or monitoring programs cost-  
54 ineffective. Apart from species that are genuinely rare as a result of small geographic ranges,  
55 limited habitat breadth, or low local population density (Rabinowitz 1981), in many cases  
56 apparent rarity may simply reflect a sampling artifact linked to sampling effort, methodology, or  
57 differential species detectability (Kéry & Schmid 2008; Meyer *et al.* 2011; van der Burg *et al.*  
58 2011). As a recent study suggests, the explicit inclusion or exclusion of rare species can  
59 profoundly affect estimates of the relative conservation value of different land-uses (Barlow *et*  
60 *al.* 2010), and can be thought to generally influence comparisons of biodiversity survey or  
61 monitoring data among habitat or land-use types.

62           Bats are considered potentially valuable indicators of biodiversity and ecosystem health  
63 and there is now increased momentum for establishing a global bat monitoring network (Jones *et*  
64 *al.* 2009; Flaquer & Puig-Montserrat 2012; KE Jones *et al.* 2013). The value of bats as  
65 bioindicators stems from their high taxonomic and functional diversity, widespread geographic  
66 distribution, their documented sensitivity to a host of anthropogenic alterations in habitat quality,  
67 and to changes in environmental conditions associated with climate change (Jones *et al.* 2009;  
68 Sherwin, Montgomery & Lundy 2013). Moreover, response patterns of bats to habitat



69 deterioration may be congruent with those of other taxa, such as insects (Jones *et al.* 2009), an  
70 important attribute with respect to the surrogacy value of a particular taxon (Moreno *et al.* 2007).  
71 Especially in the tropics where bats reach peak species richness and comprise a large fraction of  
72 local mammal faunas, they are providers of key ecosystem services and as such are integral to  
73 ecosystem functioning (Kalka, Smith & Kalko 2008; Lobova, Geiselman & Mori 2009; Kunz *et*  
74 *al.* 2011). Single localities in Neotropical lowland forests may support more than 100 sympatric  
75 bat species (Rex *et al.* 2008) and highly species-rich assemblages are also known from both  
76 tropical Asia (Kingston, Boo Liat & Zubaid 2006) and Africa (Fahr & Kalko 2011). Despite the  
77 fact that most bat biodiversity is concentrated in the tropics, current systematic monitoring  
78 efforts focus on bats in temperate regions (Battersby 2010).

79 Adequate sample representativeness is a fundamental tenet of any monitoring program or  
80 biodiversity study, as replicate surveys should adequately reflect the underlying assemblage at a  
81 site. As we have previously shown, in tropical bat surveys a certain number of repeat visits is  
82 indispensable for reliable estimation of species detectability (Meyer *et al.* 2011) and the  
83 detection of population trends (Meyer *et al.* 2010). However, as for other highly diverse tropical  
84 taxa, aiming to capture the whole spectrum of diversity at a site may not be feasible in practice as  
85 it would require a disproportionate and usually prohibitively large amount of resources within a  
86 project's given budgetary and time constraints. On the other hand, such efforts might not even be  
87 necessary when the primary objective is to characterize assemblage-environment associations or  
88 track changes in species richness or turnover rather than an in-depth enumeration of all species  
89 present at a site.

90 The effects of excluding rare species on assemblage comparisons have been well studied  
91 and have been the subject of controversial debate for aquatic macroinvertebrate and fish

92 assemblages (Cao, Williams & Williams 1998; Marchant 2002; Holtrop, Cao & Dolan 2010;  
93 Wan *et al.* 2010). A recent study by Vellend, Lilley & Starzomski (2008) addressed this topic  
94 also for several terrestrial taxa, including plants, reptiles, birds, and alpine mammals. However,  
95 for species-rich tropical bat assemblages inferential biases associated with information loss due  
96 to species undersampling have not been systematically assessed and remain poorly understood.

97         Drawing from a unique suite of some of the most extensive tropical bat assemblage  
98 datasets available, pantropical in extent, our aim was to evaluate the effectiveness of species  
99 subsets in representing among-site variation in species richness and composition. To this end, we  
100 assessed the magnitude of correlations for bat species richness and species composition,  
101 respectively, between each full dataset that included all species sampled vs. species subsets with  
102 different numbers of species deleted either at random, or according to their rarity in the  
103 respective assemblage. We predicted that species subsets would be less effective at describing  
104 among-site variation in species composition compared to species richness, as found for other  
105 taxa (Magierowski & Johnson 2006; Vellend *et al.* 2008). Further, we expected subset  
106 performance to be dependent on (i) sampling effort and sampling method, due to their influence  
107 on species detectability (Meyer *et al.* 2011) and consequently on patterns of species rarity; (ii)  
108 structural assemblage characteristics, particularly the proportion of rare species, whereby subset  
109 performance should decrease with increasing proportions of rare species in assemblages; and (iii)  
110 biogeographic region, considering that bat assemblages in the Neotropics and Paleotropics are  
111 structured differently (e.g. Struebig *et al.* 2013).

112         If species subsets retained sufficient information relative to full species sets and, for  
113 instance rare species that would be time-intensive to survey could be ignored with little loss of  
114 information, survey costs may be considerably reduced, as fewer repeat visits per sampling site

115 would be required. Our assessment therefore is of immediate relevance in the context of  
116 evaluating the feasibility and cost-effectiveness of a potential future monitoring program for  
117 tropical bats. Moreover, our analysis is timely in view of the recently revived interest in the role  
118 of bats as bioindicators (Flaquer & Puig-Montserrat 2012).

119

## 120 **Materials and methods**

### 121 DATASETS

122 We focused on tropical bat assemblages as this study was conceived as part of an  
123 evaluation of the suitability of tropical bats for long-term monitoring within Conservation  
124 International's Tropical Ecology, Assessment and Monitoring (TEAM) network  
125 (<http://www.teamnetwork.org>). Following a call for data among tropical bat ecologists, a total of  
126 27 datasets were provided by colleagues and included in the study (Table S1, Supporting  
127 Information). In all cases, datasets consisted of species abundance data collected at multiple  
128 sampling sites. For datasets originating from fragmented or otherwise disturbed areas, data only  
129 from control plots in continuous or mostly undisturbed forest were used for analysis. Nineteen  
130 datasets were based on ground-level mist netting (GN), six on canopy-level mist netting (CN),  
131 and two on acoustic sampling (AS). The majority of datasets were from the Neotropics (21  
132 compared to six from the Paleotropics). Disparities in the datasets' coverage reflect general  
133 differential research efforts in terms of sampling method and geographic region (e.g. Kingston  
134 2013). Although the datasets analysed cannot be regarded as representing 100% sampled  
135 assemblages from which to subsample, they were comparable in that they comprised bat  
136 assemblages that in each case were thoroughly sampled to similarly high levels of completeness

137 (mean inventory completeness  $81.3 \pm 6.6\%$  SD [range 67-92%] as assessed with the Jackknife1  
138 species richness estimator; Gotelli & Colwell 2010).

139

#### 140 GENERATION OF RANDOM AND NON-RANDOM SPECIES SUBSETS

141 For each dataset, we calculated species richness (SR) for each site. In addition, we  
142 performed a detrended correspondence analysis (DCA) based on the species-by-site matrix of  
143 each dataset and extracted the site scores of the first DCA axis (DCA1), which represents the  
144 dominant gradient in species composition (Legendre & Legendre 1998; see Vellend *et al.* 2008  
145 for details about the rationale for choosing this eigenanalysis-based ordination method in the  
146 context of the present analysis). As a measure of species composition complementary to DCA1,  
147 we calculated Jaccard's dissimilarity index ( $J$ ) for each pair of sites, one of the most widely used  
148 dissimilarity indices for species presence-absence data (Jost, Chao & Chazdon 2011).

149 For all datasets, we calculated SR, DCA1 and  $J$  for the full dataset, i.e. using the full  
150 species pool (hereafter denoted  $SR_{FULL}$ ,  $DCA1_{FULL}$ , and  $J_{FULL}$ ). We then calculated SR, DCA1,  
151 and  $J$  for different subsets of species per dataset ( $SR_{SUB}$ ,  $DCA1_{SUB}$ , and  $J_{SUB}$ ), whereby two  
152 different approaches were taken. In the first case, species were randomly drawn from the original  
153 species pool at each of five species pool sizes (i.e. producing five species subsets of a variable  
154 number of species). The latter ranged from 90% down to 50% of the full species pool. Following  
155 Vellend *et al.* (2008), at each species pool size, 100 subsets of species were randomly chosen  
156 with replacement. For each random draw of species, Pearson product-moment correlations  
157 between the full vs. randomly generated reduced datasets [ $r(SUB \times FULL)$ ] were subsequently  
158 used to characterize the degree to which patterns of species richness and composition in the data  
159 subsets reflect those in the complete dataset. For each dataset, we plotted the median and 95<sup>th</sup>

160 percentile correlations for the 100 random subsets against species pool size. We based  
161 assessments of the effect of randomly subsampling the entire species pool on the lower 95<sup>th</sup>  
162 percentile of the 100 correlations at each species pool size, which can be regarded as a  
163 conservative estimate of the information loss as a consequence of surveying less than the full set  
164 of species (see Vellend *et al.* 2008).

165         As a second approach, in addition to evaluating the consequences of random species  
166 subsampling, we explored the effects of removing species from the full species pool in a non-  
167 random fashion, based on their rarity in the local assemblage. To this end, we calculated the  
168 relative abundance (RA) (%) for each species per dataset and selected species subsets by  
169 eliminating increasingly larger proportions of rare species. Rare species were defined as those  
170 with a relative abundance less than 1% of total relative abundance (Maurer & McGill 2011). The  
171 average proportion of rare species across datasets was  $0.54 \pm 0.15$  (range 0.12-0.71). Depending  
172 on the species-abundance distribution of the respective assemblage, we evaluated effects with  
173 respect to up to three different rarity thresholds (whenever applicable), successively removing all  
174 species with  $RA < 0.1\%$ ,  $< 0.5\%$ , and  $< 1\%$ , i.e. always starting with the rarest species. As with  
175 random species subsets, we assessed correlations between full datasets and non-random subsets  
176 for SR, DCA1, and  $J$ .

177         We then calculated for each dataset the minimum proportion of species from the full set  
178 of species that would be required to achieve lower 95<sup>th</sup> percentile correlations  $r(\text{SUB} \times \text{FULL})$  of  
179  $\geq 0.8$ . We considered a correlation of 0.8 as an appropriate threshold as the effectiveness of using  
180 a species subset as a surrogate at lower correlation levels is questionable and may provide  
181 misleading statistical results (Vellend *et al.* 2008).

182 Analyses were conducted in R (R Development Core Team 2010), mostly using package  
183 *vegan* (Oksanen *et al.* 2008) and code adapted from Vellend *et al.* (2008) for generating random  
184 species subsets. Differences between response metrics, non-random data subsets, and sampling  
185 methods in the magnitude of achieved correlations  $r(\text{SUB} \times \text{FULL})$  were tested in a linear mixed  
186 model framework in the R package ‘lme4’ (Bates & Maechler 2010), using likelihood ratio tests  
187 to assess significance (Zuur *et al.* 2009).

188

## 189 CORRELATES OF HIGH SAMPLE REPRESENTATIVENESS

190 We assessed whether the surrogate effectiveness of species subsets, defined as the  
191 proportion of species necessary to reach lower 95<sup>th</sup> percentile correlations  $r \geq 0.8$ , was influenced  
192 by factors related to biogeographic region (Neotropics, Palaeotropics), sampling method (GN,  
193 CN [AS was not considered as there were only two datasets]), sampling effort (number of  
194 sampling plots, mean number of surveys per sampling plot), or structural characteristics of the  
195 respective assemblage (total assemblage abundance, proportion of rare species [those  
196 representing < 1% of total RA], and the reciprocal form of Simpson’s diversity index  $1/D$  (e.g.  
197 Maurer & McGill 2011)). Similarly, for non-random species subsets with rare species removed,  
198 we modeled the probability of achieving a correlation of 0.8 between full and reduced datasets  
199 (binary response variable) as a function of those same covariates.

200 Analyses were performed as generalized linear mixed-effects models (GLMMs; Zuur *et*  
201 *al.* 2009), with ‘location’ specified as random factor. Models were fitted using the ‘glmer’  
202 function in the R package ‘lme4’ (Bates & Maechler 2010), assuming a binomial error  
203 distribution and logit link function. To account for the variation in inventory completeness  
204 among datasets (see above), we included this variable as an offset. Continuous predictor

205 variables were standardized to facilitate comparison of parameter estimates (Schielzeth 2010).  
206 We conducted AIC<sub>c</sub>-based model selection and multi-model inference (Burnham & Anderson  
207 2002) using the R package ‘AICcmodavg’ (Mazerolle 2010). We chose AIC over other model  
208 selection criteria such as BIC as it is not only by far the most widely used in ecological studies,  
209 but also the best suited in the context of our application based on a recently developed decision  
210 framework (Aho, Derryberry & Peterson 2014).

211

## 212 **Results**

### 213 RANDOM SPECIES SUBSETS

214 Median correlations between full datasets and random subsets in general showed  
215 relatively little variation across datasets, and irrespective of the response metric, strong  
216 correlations were observed even with a large fraction of species removed (Figs 1 & 2, Figs S1 &  
217 S2). For species subsets representing 50% of the initial species pool, median correlations  
218 averaged  $0.87 \pm 0.10$  SD (range 0.64-0.98) for SR,  $0.83 \pm 0.12$  (range 0.44-0.99) for DCA1, and  
219  $0.76 \pm 0.09$  (range 0.56-0.94) for *J* across all datasets analyzed.

220 On the other hand, lower 95<sup>th</sup> percentile correlations, which represent a more  
221 conservative estimate, suggest that a great deal of information may be lost unless most of the  
222 original species set is retained. The magnitude of lower 95<sup>th</sup> percentile correlations differed  
223 significantly according to sampling method (LMM,  $\chi^2 = 7.35$ ,  $df = 2$ ,  $P = 0.025$ ), being slightly  
224 higher for datasets based on canopy versus ground mist netting (Tukey contrasts,  $Z = -2.51$ ,  $P_{adj.}$   
225  $= 0.028$ ). Moreover, it differed significantly with respect to the response metric considered ( $\chi^2 =$   
226  $29.04$ ,  $df = 2$ ,  $P < 0.001$ ), with generally weaker correlations for DCA1 compared to SR ( $Z =$   
227  $4.31$ ,  $P_{adj.} < 0.001$ ) and *J* ( $Z = 5.06$ ,  $P_{adj.} < 0.001$ ). Correlations were weak with large proportions

228 of the species pool dropped and highly variable among datasets, especially for DCA1 (Fig. 2);  
229 average correlations at  $\leq 70\%$  of the original species pool were  $< 0.56$  for SR,  $< 0.27$  for DCA1,  
230 and  $< 0.55$  for *J*. Strong lower 95<sup>th</sup> percentile correlations were only found with 90% of species  
231 retained, averaging  $0.89 \pm 0.12$  (range 0.53-0.99) for SR,  $0.64 \pm 0.30$  (range 0.11-0.99) for  
232 DCA1, and  $0.80 \pm 0.18$  (range 0.05-0.96) for *J* (Fig. 2). In other words, across datasets, for SR  
233 on average  $79.8 \pm 15.6\%$  (range 40.9 - 100%) of the species from the initial set were necessary to  
234 achieve lower 95<sup>th</sup> percentile correlations  $r \geq 0.8$  between original datasets and random subsets.  
235 Effective surrogates for species composition would require that on average roughly 90% of the  
236 original species pool be retained (DCA1:  $88.8 \pm 15.8\%$ , range 51.4 - 100%; *J*:  $90.0 \pm 8.3\%$ ,  
237 range 68.2 - 100%).

238

#### 239 NON-RANDOM SPECIES SUBSETS

240 As with random species subsampling, eliminating species from the original species pool  
241 in a non-random fashion based on their rarity in the respective assemblage yielded correlations  
242 that were highly variable across datasets (Figs 1 & 2, Figs S1 & S2). The magnitude of  
243 correlations between original and reduced datasets did not vary significantly among response  
244 metrics (LMM,  $\chi^2 = 1.75$ ,  $df = 2$ ,  $P = 0.418$ ) or sampling methods ( $\chi^2 = 2.06$ ,  $df = 2$ ,  $P = 0.356$ ).  
245 It did, however, differ significantly among data subsets, i.e. depending on what fraction of rare  
246 species was trimmed off from the tail of the species-abundance distribution ( $\chi^2 = 20.46$ ,  $df = 2$ ,  $P$   
247  $< 0.001$ ). Irrespective of the response metric, correlations across datasets were on average  $> 0.8$   
248 when only the rarest species ( $< 0.1\%$  of total RA, corresponding to  $9.4 \pm 11.2\%$  of the initial  
249 species pool) were eliminated, with little variation among datasets for SR and *J* compared to  
250 DCA1 (SR:  $0.95 \pm 0.05$ , range 0.84-0.99; DCA1:  $0.82 \pm 0.29$ , range 0.10-0.99; *J*:  $0.94 \pm 0.05$ ,



251 range 0.85-0.99). Additionally removing the species in the next higher rarity categories ( $< 0.5$   
252 and  $< 1\%$  of total RA, corresponding to  $39.9 \pm 19.4\%$  and  $54.4 \pm 15.3\%$ , respectively, of the  
253 initial species pool) resulted in significantly lower correlations (Tukey contrasts,  $P_{adj.} < 0.001$ )  
254 and increased variability in the magnitude of correlations among datasets also for SR and  $J$  (Fig.  
255 2).

256

## 257 CORRELATES OF HIGH SAMPLE REPRESENTATIVENESS

258 For random species subsampling,  $AIC_c$ -model selection revealed strong support for an  
259 effect of sampling effort, particularly the number of repeat visits per plot, on the proportion of  
260 species required to yield lower 95<sup>th</sup> percentile correlations  $r(\text{SUB} \times \text{FULL}) \geq 0.8$  (Table 1 & 3).  
261 Number of surveys or the composite model ‘sampling effort’, which considered the number of  
262 sampling plots and the number of visits/plot, were the top-ranked or second-ranked model in the  
263 candidate set, irrespective of the response metric chosen.

264 In the case of non-random species subsets (Table 2 & 3), for SR and  $J$  as response  
265 metrics, high sample representativeness was most strongly correlated with the proportion of rare  
266 species in the assemblage ( $\omega = 0.63$  and  $\omega = 0.88$ , respectively). For DCA, there was  
267 considerable evidence for an overall effect of structural assemblage characteristics, although  
268  $AIC_c$ -differences and model weights suggested considerable model selection uncertainty. Total  
269 assemblage abundance was the top-ranked model ( $\omega = 0.31$ ), followed by Simpson’s diversity  
270 index ( $\omega = 0.19$ ) and the proportion of rare species ( $\omega = 0.14$ ).

271

## 272 Discussion

273 We quantified inferential biases associated with species subsampling in tropical bat  
274 assemblages and demonstrate that moderately undersampled species subsets may in many cases  
275 be sufficient to enable reliable comparisons of species richness and compositional variation  
276 across sites. As expected, species subsets performed better at retaining information on inter-site  
277 variation in species richness than species composition. Moreover, in line with our predictions, we  
278 found that sampling effort and structural assemblage characteristics, specifically the proportion  
279 of rare species in an assemblage, were important predictors of subset performance. In contrast,  
280 there was no significant effect of either sampling method or geographic region on high sample  
281 representativeness.

282

#### 283 SURROGATE PERFORMANCE OF RANDOM AND NON-RANDOM SUBSETS

284 Similar assessments for a diverse array of other taxa, including plants, invertebrates, fish,  
285 reptiles, birds, and non-volant mammals (Vellend *et al.* 2008; Molloy *et al.* 2010; Bried *et al.*  
286 2012) also found high levels of congruence between full and reduced datasets when ignoring  
287 10%, and often larger proportions, of the original species pool. Our results show that information  
288 loss was equally low with similar levels of species' exclusion (ca. 15%). An important caveat to  
289 note is that our analysis was based on datasets that had an average of 81% completeness.

290 Although we statistically controlled for variation in inventory completeness among datasets in  
291 modeling correlates of high sample representativeness, it remains unknown to what degree our  
292 results might have been different if we had subsampled fully inventoried assemblages, i.e.  
293 datasets that had near 100% completeness.

294 In many instances we found that correlations for non-random subsets mirrored those  
295 based on random subsets reasonably well; however, for certain datasets correlations deviated

296 considerably from median correlations for randomly chosen subsets. Strong differences were  
297 particularly apparent with DCA1 correlations for some datasets (e.g. Comoé, Yungas, Victoria-  
298 Mayaro, Fig. S1b), whereas agreement between correlations for random and non-random subsets  
299 was in most cases much better for SR and *J*. Our findings concur with those of Vellend *et al.*  
300 (2008) in that correlations for DCA1 often showed greater variability across datasets than for *J*.  
301 This indicates that species subsets may often be less effective at capturing the same maximum  
302 possible amount of compositional variation among sites (as given by DCA1) than the full set. In  
303 contrast, pairwise site differences (Jaccard dissimilarities) in species composition may be more  
304 consistently revealed with a reasonably large subsample of the entire species set.

305         While part of our analyses focused on random species subsets, our findings concerning  
306 the effects of undersampling due to species rarity for predicting diversity patterns may be more  
307 revealing and of greater general relevance. Corroborating previous studies on invertebrates  
308 (Heino & Soininen 2010; Franklin *et al.* 2013), our results suggest that patterns of spatial  
309 turnover in tropical bat assemblages are to a large extent driven by the more common species and  
310 for the accurate description of assemblage similarity-environment relationships, rare species may  
311 often be of limited importance. Removing only the least abundant species from an assemblage  
312 (those with < 0.1% of total RA, comprising on average ~10% of the original species pool)  
313 yielded strong correlations (> 0.8) across nearly all datasets (Fig. 2). This indicates that if only  
314 the rarest species in an assemblage were missed during a survey, information loss would be  
315 tolerable in most cases and that the species subset sampled can serve as a good surrogate for the  
316 full suite of species actually present in the assemblage. Limiting surveys to sampling only the  
317 more common species and ignoring the rarest ones therefore seems a reasonable shortcut for  
318 reducing costs in tropical bat monitoring programs. It is important, however, to emphasize that

319 our findings in this regard do not apply to situations where the objective is the detailed  
320 population monitoring of rare species. Although inherently of greater conservation interest than  
321 common ones, rare species in tropical bat assemblages are difficult to monitor and will always  
322 require a high-effort sampling design for reliable trend detection, as we have previously  
323 demonstrated (Meyer *et al.* 2010). Common species have variously been shown to contribute  
324 disproportionately to species richness patterns (Pearman & Weber 2007; Gaston 2008; Šizling *et*  
325 *al.* 2009; Lennon *et al.* 2011). Our results are in line with these findings and point towards a  
326 considerable degree of structural redundancy in species composition (*sensu* Clarke & Warwick  
327 1998) in tropical bat assemblages, which may in fact be a general feature of many biological  
328 communities (Cayuela, De La Cruz & Ruokolainen 2011).

329

#### 330 FACTORS AFFECTING SPECIES SUBSET PERFORMANCE

331       Contrary to expectations, subset performance was not dependent on geographic region.  
332 However, this finding should be interpreted with some caution since our study included far more  
333 datasets from the New World than from the Old World tropics, which may have reduced  
334 statistical power to detect significant differences. Only increased research efforts underway in  
335 Asia (Kingston 2013) and, hopefully, in the future also in Africa, can help to substantiate this  
336 finding based on a geographically more balanced set of studies.

337       While sampling effort was the best correlate of high sample representativeness with  
338 random subsampling, structural assemblage characteristics, most notably the proportion of rare  
339 species in an assemblage, was the best predictor of surrogate performance when datasets were  
340 subsampled according to rarity. Tropical bat assemblages typically comprise many rare species,  
341 yet vary substantially with respect to the number of rare species they contain, as evidenced by

342 our datasets (Fig. 3). Our results imply that whenever assemblages are comprised of a large  
343 number of rare species, relatively larger fractions of these will need to be sampled to adequately  
344 capture among-site variation in species richness and composition, essentially requiring increased  
345 sampling effort and more comprehensive surveys. Trimming off progressively greater  
346 proportions of species (i.e. species representing  $< 0.5$  and  $< 1\%$  of total RA, encompassing on  
347 average 40% and 54%, respectively of the full set) resulted in correlations often lower than 0.8  
348 (Fig. 2). Correlations  $< 0.7$  may greatly reduce the statistical power for testing relationships  
349 between species diversity or composition and environmental covariates, and in fact only strong  
350 relationships may be detectable using a surrogate in such cases (Vellend *et al.* 2008). Thus, in the  
351 search for suitable surrogates, correlations  $> 0.7$  should be aimed for to guarantee that  
352 assemblage-environment relationships can be reliably assessed.

353         That inferential biases associated with undersampling increase for species-rich  
354 assemblages that are made up of a large number of rare species can clearly be seen in the case of  
355 the assemblages from Comoé (Ivory Coast), Victoria-Mayaro (Trinidad) or Barro Colorado  
356 Nature Monument (Panama), each characterized by a high proportion of rare species (60-70%).  
357 Congruence in multivariate response patterns between original data and subsets with all of those  
358 rare species removed was generally very low (correlations  $\ll 0.8$ ), particularly for DCA1. For  
359 these assemblages, subsets containing only the more abundant species would fail to capture the  
360 same dominant gradient in species composition as in the initial dataset. This was most prominent  
361 in the Comoé ground-net assemblage where even removing only the rarest few species yielded a  
362 correlation of less than 0.4 (Fig. S1b). Such apparent failure to capture among-site patterns in  
363 species composition with species subsets may reflect the major role of high habitat heterogeneity  
364 in shaping diversity patterns in this particular assemblage. The Comoé assemblage had the

365 largest proportion of rare species of all datasets examined (71%) and is characterized by high  
366 species richness, a pattern largely attributable to its geographical position in a biome transition  
367 zone between forest and savanna, where habitat heterogeneity is sharply elevated (Fahr & Kalko  
368 2011). In contrast, the bat assemblage at Tiputini (Ecuador), although one of the most species-  
369 rich known (Rex *et al.* 2008), is characterized by comparatively higher evenness of its abundance  
370 distribution, which may explain the generally high correlations found with non-random species  
371 removal. These findings indicate that the trade-off between number of sites surveyed and survey  
372 comprehensiveness is system-specific. This in turn implies that the investment required for  
373 capturing a representative sample of the whole assemblage varies across geographic locations,  
374 reflecting spatial variation in the number of rare species and ultimately in mean species detection  
375 probabilities (Meyer *et al.* 2011), both of which are intuitively closely linked (McCarthy *et al.*  
376 2013). Gauging the relationship between species abundance and detectability is important as it  
377 can help to determine adequate sampling effort. However, in general how exactly detection  
378 probabilities scale with abundance remains little explored (McCarthy *et al.* 2013), an aspect  
379 which provides an interesting avenue for future research in the context of bat biodiversity  
380 surveys.

381         The majority of assemblages analyzed in this study were sampled using a single method,  
382 ground-level mist nets, reflecting the general fact that tropical bat assemblage inventories that  
383 use a combination of different survey methods remain scarce. Our low sample size for datasets  
384 not based on ground-level mist netting may in part explain why, opposite to what we expected,  
385 sampling method was not found to be an important predictor of species subset performance.  
386 Sampling method influences species detectability in tropical bats (Meyer *et al.* 2011) and hence  
387 is an important determinant of local-scale patterns of species rarity. To accurately infer which

388 species in an assemblage are truly rare therefore requires comprehensive surveys employing a  
389 combination of active (i.e. mist netting at ground- and canopy level, harp traps) and passive  
390 survey methods (acoustic sampling) to maximize inventory completeness (MacSwiney, Clarke &  
391 Racey 2008; Kunz, Hodgkison & Weise 2009; Kingston 2013). Consequently, assessments of  
392 the surrogate effectiveness of species subsets should ideally be based on assemblages that have  
393 been surveyed with multiple complementary methods to properly account for confounding  
394 effects of sampling method on patterns of species rarity. If we had had such data available, this  
395 would no doubt have strengthened the robustness of our inferences drawn about how the  
396 exclusion of rare species influences surrogate effectiveness (see above). We therefore consider  
397 this an important aspect that merits attention in similar future evaluations.

398

#### 399 SURROGATE EFFECTIVENESS OF SPECIES SUBSETS: SPECIES RICHNESS VS. 400 COMPOSITION

401 As predicted and corroborating previous work on other taxa (Magierowski & Johnson  
402 2006; Vellend *et al.* 2008), we found that partial species sets generally are robust surrogates of  
403 total species richness, however, they perform less well in uncovering compositional patterns.  
404 Although species richness is a state variable commonly used in monitoring programs (JPG Jones  
405 *et al.* 2013), its usefulness in environmental impact assessments has recently been questioned as  
406 measures of assemblage composition and turnover have been found to be more informative and  
407 sensitive to change (Barlow *et al.* 2007; Magurran & Henderson 2010; Banks-Leite, Ewers &  
408 Metzger 2012; Dornelas *et al.* 2014). Undersampling bias is a key challenge not only with regard  
409 to biodiversity assessment and monitoring, as examined here, but also constitutes an active area  
410 of research in many other fields of ecological research, including species distribution modeling

411 (Kramer-Schadt *et al.* 2013; Syfert, Smith & Coomes 2013) or the analysis of plant-animal  
412 interaction networks (Nielsen & Bascompte 2007; Rivera-Hutinel *et al.* 2012), where equivalents  
413 of species richness (e.g. interaction richness) have also been found to be less robust than  
414 alternative metrics (e.g. Tylianakis *et al.* 2010). This highlights the general need for ecologists  
415 and conservation biologists to move beyond mere species numbers and to focus on more  
416 informative assemblage metrics, capable of adequately capturing changes in relation to  
417 environmental impacts or monitoring alterations in ecological network structure. We argue that  
418 in the context of monitoring for environmental impact assessment researchers should give greater  
419 consideration to measures of species composition and turnover to increase the validity of  
420 inferences made from evaluations of the suitability and performance of species subsets as  
421 surrogates of total taxon richness. More specifically, we advocate a wider application of metrics  
422 suitable for quantifying biodiversity change, for instance commonly applied similarity or  
423 distance measures (e.g. the Morisita-Horn index) and specialized turnover indices (Magurran &  
424 Henderson 2010; Jost *et al.* 2011; Magurran 2011) or rank abundance statistics such as mean  
425 rank shift (Collins *et al.* 2008). The merits and necessity of a shift of focus towards such  
426 measures are well illustrated by the recent finding of a global analysis of long-term assemblage  
427 time series, which detected no systematic temporal change in alpha diversity, but consistent  
428 compositional change and turnover (Dornelas *et al.* 2014).

429

## 430 **Conclusions**

431 Our analyses stress that there is potential for reducing costs in tropical bat monitoring by  
432 streamlining sampling activities if the focus is on assessing assemblage-environment  
433 relationships or changes in species richness or turnover. Protocols that consider reasonably high



434 but not exhaustive sampling, which may equate to fewer surveys, seem to be sufficiently  
435 sensitive to allow reliable inferences regarding among-site variation in bat species richness and  
436 assemblage composition. This suggests that survey efficiency may be maximized by ignoring  
437 those species that are most time-consuming to sample, i.e. those that make up the far end of the  
438 extended rare-species tail of the relative species-abundance distribution.

439         Our analyses demonstrate, however, that a one-size-fits-all approach to surrogate  
440 selection based on species subsets may be inappropriate, but will have to be tailored to site-  
441 specific circumstances and consider the structural idiosyncrasies of local assemblages. In  
442 essence, monitoring programs will have to establish site-specific performance levels for  
443 biodiversity surrogates based on pilot data. In practice, this will require relatively detailed  
444 surveys at the beginning of a survey or monitoring program, which should entail the use of  
445 multiple sampling methods to accurately establish true patterns of species rarity. Such pilot  
446 surveys should be combined with the application of robust statistical approaches to assess survey  
447 completeness based on the species richness estimator most appropriate for a given dataset (see  
448 Reese, Wilson & Flather 2014 for a recent framework concerning estimator selection) to  
449 determine to what extent the use of species subsets is justifiable (Franklin *et al.* 2013).  
450 Implementing adaptive sampling schemes that avoid oversampling at some sites and  
451 undersampling at others (cf. Holtrop *et al.* 2010), i.e. aim to spatially prioritize sampling effort,  
452 may ultimately be key to maximizing cost-effectiveness in tropical bat surveys. Finally, when  
453 adopting a surrogate as part of a bat monitoring program or in environmental impact assessments  
454 it will be essential to assess its robustness across relevant spatial and also temporal scales, and to  
455 determine its performance prior to and after environmental impact as disturbance may alter the

456 relationship between the species subset and total biodiversity (Magierowski & Johnson 2006;  
457 Sebek *et al.* 2012).

458

#### 459 **Acknowledgements**

460 We thank Conservation International (CI) for financial support. The authors further  
461 acknowledge support from the following organizations for funding of the studies included in this  
462 work: The MacArthur Foundation (LFA, IM), German Academic Exchange Service (CFJM, JF),  
463 German Science Foundation (CFJM, EKVK [KA 1241/6-1], CCV [Vo 890/7]), The Leverhulme  
464 Trust (FMC, PAR), The National Geographic Society (RKBJ), German Federal Ministry of  
465 Education and Research (JF, EKVK [BIOTA program, Project 01LC0017, 01LC0411 &  
466 01LC0617E1]), The Center for Ecology and Conservation Biology, Boston University (THK,  
467 KR, CCV), Lube Bat Conservancy (THK), US National Science Foundation (THK),  
468 Landesgraduiertenförderung Baden-Württemberg (JF), The Darwin Initiative (NF, RKBJ), The  
469 Rufford Foundation (NF), Consejo Nacional de Ciencia y Tecnología (MCMG [No. 168990]),  
470 The University of Aberdeen (MCMG, PAR), Conselho Nacional de Pesquisa – CNPq (LMSA),  
471 Electricité de France (MH, JFC, JMP [Convention Muséum/EDF CQZH 1294]), and Centro de  
472 Investigaciones en Ecosistemas, Universidad Nacional Autónoma de México (KES). We are  
473 grateful to Brock Fenton, two anonymous reviewers, and the Associate Editor for valuable  
474 comments, which improved the manuscript. This research evolved from a bat monitoring  
475 workshop hosted by the TEAM network at CI. Special thanks go to S. Andelman, K.E. Jones,  
476 T.H. Kunz, and M.R. Willig who were instrumental in pursuing the idea of this meeting. We  
477 dedicate this research to our mentor, friend, and colleague, Elisabeth Kalko, who sadly died  
478 before completion of this manuscript.

479

480 **References**

481 Aho, K., Derryberry, D. & Peterson, T. (2014) Model selection for ecologists: the worldviews of  
482 AIC and BIC. *Ecology*, **95**, 631-636.

483 Banks-Leite, C., Ewers, R.M. & Metzger, J.P. (2012) Unraveling the drivers of community  
484 dissimilarity and species extinction in fragmented landscapes. *Ecology*, **93**, 2560-2569.

485 Barlow, J., Gardner, T.A., Araujo, I.S., Avila-Pires, T.C., Bonaldo, A.B., Costa, J.E. *et al.* (2007)  
486 Quantifying the biodiversity value of tropical primary, secondary and plantation forest.  
487 *Proceedings of the National Academy of Sciences USA*, **104**, 18555-18560.

488 Barlow, J., Gardner, T.A., Louzada, J. & Peres, C.A. (2010) Measuring the conservation value of  
489 tropical primary forests: the effect of occasional species on estimates of biodiversity  
490 uniqueness. *PLoS ONE*, **5**, e9609.

491 Bates, D. & Maechler, M. (2010) lme4: Linear mixed-effects models using S4 classes. R  
492 package version 0.999375-999337. <http://CRAN.R-project.org/package=lme999374>.

493 Battersby, J. (2010) *Guidelines for surveillance and monitoring of European bats*.  
494 UNEP/EUROBATS Secretariat, Bonn, Germany.

495 Bried, J.T., Hager, B.J., Hunt, P.D., Fox, J.N., Jensen, H.J. & Vowels, K.M. (2012) Bias of  
496 reduced-effort community surveys for adult Odonata of lentic waters. *Insect Conservation  
497 and Diversity*, **5**, 213-222.

498 Burnham, K.P. & Anderson, D.R. (2002) *Model selection and inference: a practical  
499 information-theoretic approach*, 2nd edn. Springer, New York.

500 Cao, Y., Williams, D.D. & Williams, N.E. (1998) How important are rare species in aquatic  
501 community ecology and bioassessment? *Limnology and Oceanography*, **43**, 1403-1409.

502 Cayuela, L., De La Cruz, M. & Ruokolainen, K. (2011) A method to incorporate the effect of  
503 taxonomic uncertainty on multivariate analyses of ecological data. *Ecography*, **34**, 94-102.

504 Clarke, K.R. & Warwick, R.M. (1998) Quantifying structural redundancy in ecological  
505 communities. *Oecologia*, **113**, 278-289.

506 Coddington, J.A., Agnarsson, I., Miller, J.A., Kuntner, M. & Hormiga, G. (2009) Undersampling  
507 bias: the null hypothesis for singleton species in tropical arthropod surveys. *Journal of*  
508 *Animal Ecology*, **78**, 573-584.

509 Collins, S.L., Suding, K.N., Cleland, E.E., Batty, M., Pennings, S.C., Gross, K.L. *et al.* (2008)  
510 Rank clocks and plant community dynamics. *Ecology*, **89**, 3534-3541.

511 Dornelas, M., Gotelli, N.J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C. *et al.* (2014)  
512 Assemblage time series reveal biodiversity change but not systematic loss. *Science*, **344**, 296-  
513 299.

514 Fahr, J. & Kalko, E.K.V. (2011) Biome transitions as centres of diversity: habitat heterogeneity  
515 and diversity patterns of West African bat assemblages across spatial scales. *Ecography*, **34**,  
516 177-195.

517 Flaquer, C. & Puig-Montserrat, X. (2012) Proceedings of the International Symposium on the  
518 Importance of Bats as Bioindicators. Museum of Natural Sciences Edicions, Granollers,  
519 Catalonia, Spain.

520 Franklin, E., de Moraes, J., Landeiro, V.L., de Souza, J.L.P., Pequeno, P.A.I.C.L., Magnusson,  
521 W.E. *et al.* (2013) Geographic position of sample grid and removal of uncommon species  
522 affect multivariate analyses of diverse assemblages: The case of oribatid mites (Acari:  
523 Oribatida). *Ecological Indicators*, **34**, 172-180.

524 Gardner, T.A., Barlow, J., Araujo, I.S., Avila-Pires, T.C., Bonaldo, A.B., Costa, J.E. *et al.* (2008)  
525 The cost effectiveness of biodiversity surveys in tropical forests. *Ecology Letters*, **11**, 139-  
526 150.

527 Gaston, K.J. (2008) Biodiversity and extinction: the importance of being common. *Progress in*  
528 *Physical Geography*, **32**, 73-79.

529 Gotelli, N.J. & Colwell, R.K. (2010) Estimating species richness. *Biological diversity: Frontiers*  
530 *in measurement and assessment* (eds A.E. Magurran & B.J. McGill), pp. 39-54. Oxford  
531 University Press.

532 Heino, J. & Soininen, J. (2010) Are common species sufficient in describing turnover in aquatic  
533 metacommunities along environmental and spatial gradients? *Limnology and Oceanography*,  
534 **55**, 2397-2402.

535 Holtrop, A.M., Cao, Y. & Dolan, C.R. (2010) Estimating sampling effort required for  
536 characterizing species richness and site-to-site similarity in fish assemblage surveys of  
537 wadeable Illinois streams. *Transactions of the American Fisheries Society*, **139**, 1421-1435.

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

540 Jones, J.P.G., Asner, G.P., Butchart, S.H.M. & Karanth, K.U. (2013) The ‘why’, ‘what’ and  
541 ‘how’ of monitoring for conservation. *Key Topics in Conservation Biology 2* (eds D.W.  
542 Macdonald & K.J. Willis), pp. 327-343. John Wiley & Sons, Oxford, UK.

543 Jones, K.E., Russ, J.A., Bashta, A.-T., Bilhari, Z., Catto, C., Csösz, I. *et al.* (2013) Indicator Bats  
544 Program: a system for the global acoustic monitoring of bats. *Biodiversity Monitoring and*  
545 *Conservation* (eds B. Collen, N. Pettorelli, J.E.M. Baillie & S.M. Durant), pp. 211-247.  
546 Wiley-Blackwell, Oxford, UK.

547 Jost, L., Chao, A. & Chazdon, R.L. (2011) Compositional similarity and beta diversity.  
548 *Biological diversity: frontiers in measurement and assessment* (eds A.E. Magurran & B.J.  
549 McGill), pp. 66-84. Oxford University Press, New York.

550 Kalka, M.B., Smith, A.R. & Kalko, E.K.V. (2008) Bats limit arthropods and herbivory in a  
551 tropical forest. *Science*, **320**, 71.

552 Kéry, M. & Schmid, B.R. (2008) Imperfect detection and its consequences for monitoring for  
553 conservation. *Community Ecology*, **9**, 207-216.

554 Kessler, M., Abrahamczyk, S., Bos, M., Buchori, D., Putra, D.D., Robbert Gradstein, S. *et al.*  
555 (2011) Cost-effectiveness of plant and animal biodiversity indicators in tropical forest and  
556 agroforest habitats. *Journal of Applied Ecology*, **48**, 330-339.

557 Kingston, T., Boo Liat, L. & Zubaid, A. (2006) *Bats of Krau Wildlife Reserve*. Penerbit  
558 Universiti Kebangsaan Malaysia Press, Bangi.

559 Kingston, T. (2013) Response of bat diversity to forest disturbance in Southeast Asia: insights  
560 from long-term research in Malaysia. *Bat Evolution, Ecology, and Conservation* (eds R.A.  
561 Adams & S.C. Pedersen), pp. 169-185. Springer New York.

562 Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schröder, B., Lindenborn, J., Reinfelder, V. *et al.*  
563 (2013) The importance of correcting for sampling bias in MaxEnt species distribution  
564 models. *Diversity and Distributions*, **19**, 1366-1379.

565 Kunz, T.H., Hodgkison, R. & Weise, C.D. (2009) Methods of capturing and handling bats.  
566 *Ecological and behavioral methods for the study of bats* (eds T.H. Kunz & S. Parsons), pp. 3-  
567 35. Johns Hopkins University Press, Baltimore, Maryland.

568 Kunz, T.H., de Torrez, E.B., Bauer, D., Lobova, T. & Fleming, T.H. (2011) Ecosystem services  
569 provided by bats. *Annals of the New York Academy of Sciences*, **1223**, 1-38.

570 Larsen, F.W., Bladt, J., Balmford, A. & Rahbek, C. (2012) Birds as biodiversity surrogates: will  
571 supplementing birds with other taxa improve effectiveness? *Journal of Applied Ecology*, **49**,  
572 349-356.

573 Legendre, P. & Legendre, L. (1998) *Numerical Ecology*, 2nd edn. Elsevier, Amsterdam.

574 Lennon, J.J., Beale, C.M., Reid, C.L., Kent, M. & Pakeman, R.J. (2011) Are richness patterns of  
575 common and rare species equally well explained by environmental variables? *Ecography*, **34**,  
576 529-539.

577 Lobova, T.A., Geiselman, C.K. & Mori, S.A. (2009) *Seed dispersal by bats in the Neotropics*.  
578 New York Botanical Garden Press, New York.

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

582 Magierowski, R.H. & Johnson, C.R. (2006) Robustness of surrogates of biodiversity in marine  
583 benthic communities. *Ecological Applications*, **16**, 2264-2275.

584 Magurran, A.E. & Henderson, P.A. (2010) Temporal turnover and the maintenance of diversity  
585 in ecological assemblages. *Philosophical Transactions of the Royal Society B: Biological*  
586 *Sciences*, **365**, 3611-3620.

587 Magurran, A.E. (2011) Measuring biological diversity in time (and space). *Biological diversity:*  
588 *frontiers in measurement and assessment* (eds A.E. Magurran & B.J. McGill), pp. 85-93.  
589 Oxford University Press, New York.

590 Marchant, R. (2002) Do rare species have any place in multivariate analysis for bioassessment?  
591 *Journal of the North American Benthological Society*, **21**, 311-313.

592 Maurer, B.A. & McGill, B.J. (2011) Measurement of species diversity. *Biological diversity:*  
593 *frontiers in measurement and assessment* (eds A.E. Magurran & B.J. McGill), pp. 55-65.  
594 Oxford University Press, New York.

595 Mazerolle, M.J. (2010) AICcmodavg: Model selection and multimodel inference based on  
596 (Q)AIC(c). R package version 1.13. <http://CRAN.R-project.org/package=AICcmodavg>.

597 McCarthy, M.A., Moore, J.L., Morris, W.K., Parris, K.M., Garrard, G.E., Vesk, P.A. *et al.*  
598 (2013) The influence of abundance on detectability. *Oikos*, **122**, 717-726.

599 McDonald-Madden, E., Baxter, P.W.J., Fuller, R.A., Martin, T.G., Game, E.T., Montambault, J.  
600 *et al.* (2010) Monitoring does not always count. *Trends in Ecology & Evolution*, **25**, 547-550.

601 Meyer, C.F.J., Aguiar, L.M.S., Aguirre, L.F., Baumgarten, J., Clarke, F.M., Cosson, J.-F. *et al.*  
602 (2010) Long-term monitoring of tropical bats for anthropogenic impact assessment: gauging  
603 the statistical power to detect population change. *Biological Conservation*, **143**, 2797-2807.

604 Meyer, C.F.J., Aguiar, L.M.S., Aguirre, L.F., Baumgarten, J., Clarke, F.M., Cosson, J.-F. *et al.*  
605 (2011) Accounting for detectability improves estimates of species richness in tropical bat  
606 surveys. *Journal of Applied Ecology*, **48**, 777-787.

607 Molloy, P.P., Anticamara, J.A., Rist, J.L. & Vincent, A.C.J. (2010) Frugal conservation: What  
608 does it take to detect changes in fish populations? *Biological Conservation*, **143**, 2532-2542.

609 Moreno, C.E., Pineda, E., Escobar, F. & Sánchez-Rojas, G. (2007) Shortcuts for biodiversity  
610 evaluation: a review of terminology and recommendations for the use of target groups,  
611 bioindicators and surrogates. *International Journal of Environment and Health*, **1**, 71-86.

612 Nielsen, A. & Bascompte, J. (2007) Ecological networks, nestedness and sampling effort.  
613 *Journal of Ecology*, **95**, 1134-1141.



614 Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G.L., Stevens, M.H.H. *et al.* (2008)  
615 vegan: Community Ecology Package. R package version 1.13-11. [http://CRAN.R-](http://CRAN.R-project.org/package=vegan)  
616 [project.org/package=vegan](http://CRAN.R-project.org/package=vegan).

617 Pearman, P.B. & Weber, D. (2007) Common species determine richness patterns in biodiversity  
618 indicator taxa. *Biological Conservation*, **138**, 109-119.

619 R Development Core Team (2010) R: A language and environment for statistical computing. R  
620 Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-900007-900050,  
621 <http://www.R-project.org>.

622 Rabinowitz, D. (1981) Seven forms of rarity. *The biological aspects of rare plant conservation*  
623 (ed. H. Synge), pp. 205-217. Wiley.

624 Reese, G.C., Wilson, K.R. & Flather, C.H. (2014) Performance of species richness estimators  
625 across assemblage types and survey parameters. *Global Ecology and Biogeography*, **23**, 585-  
626 594.

627 Rex, K., Kelm, D.H., Wiesner, K., Kunz, T.H. & Voigt, C.C. (2008) Species richness and  
628 structure of three Neotropical bat assemblages. *Biological Journal of the Linnean Society*, **94**,  
629 617-629.

630 Rivera-Hutinel, A., Bustamante, R.O., Marín, V.H. & Medel, R. (2012) Effects of sampling  
631 completeness on the structure of plant-pollinator networks. *Ecology*, **93**, 1593-1603.

632 Schielzeth, H. (2010) Simple means to improve the interpretability of regression coefficients.  
633 *Methods in Ecology and Evolution*, **1**, 103-113.

634 Sebek, P., Barnouin, T., Brin, A., Brustel, H., Dufrêne, M., Gosselin, F. *et al.* (2012) A test for  
635 assessment of saproxylic beetle biodiversity using subsets of "monitoring species".  
636 *Ecological Indicators*, **20**, 304-315.

637 Sherwin, H.A., Montgomery, W.I. & Lundy, M.G. (2013) The impact and implications of  
638 climate change for bats. *Mammal Review*, **43**, 171-182.

639 Šizling, A., Šizlingová, E., Storch, D., Reif, J. & Gaston, K.J. (2009) Rarity, commonness, and  
640 the contribution of individual species to species richness patterns. *The American Naturalist*,  
641 **174**, 82-93.

642 Struebig, M.J., Turner, A., Giles, E., Lasmana, F., Tollington, S., Bernard, H. *et al.* (2013)  
643 Quantifying the biodiversity value of repeatedly logged rainforests: gradient and comparative  
644 approaches from Borneo. *Advances in Ecological Research* (eds W. Guy & J.O.G. Eoin), pp.  
645 183-224. Academic Press.

646 Syfert, M.M., Smith, M.J. & Coomes, D.A. (2013) The effects of sampling bias and model  
647 complexity on the predictive performance of MaxEnt species distribution models. *PLoS*  
648 *ONE*, **8**, e55158.

649 Tylianakis, J.M., Laliberté, E., Nielsen, A. & Bascompte, J. (2010) Conservation of species  
650 interaction networks. *Biological Conservation*, **143**, 2270-2279.

651 van der Burg, M.P., Bly, B., VerCauteren, T. & Tyre, A.J. (2011) Making better sense of  
652 monitoring data from low density species using a spatially explicit modelling approach.  
653 *Journal of Applied Ecology*, **48**, 47-55.

654 Vellend, M., Lilley, P.L. & Starzomski, B.M. (2008) Using subsets of species in biodiversity  
655 surveys. *Journal of Applied Ecology*, **45**, 161-169.

656 Wan, H., Chizinski, C.J., Dolph, C.L., Vondracek, B. & Wilson, B.N. (2010) The impact of rare  
657 taxa on a fish index of biotic integrity. *Ecological Indicators*, **10**, 781-788.

658 Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. & Smith, G.M. (2009) *Mixed effects models*  
659 *and extensions in ecology with R*. Springer, New York.

660

662 **Supporting information**

663 Additional Supporting Information may be found in the online version of this article.

664

665 **Table S1.** List of datasets used in this study.

666 **Fig. S1.** Plots showing correlations between full and reduced datasets based on ground-level mist  
667 netting.

668 **Fig. S2.** Plots showing correlations between full and reduced datasets based on canopy-level mist  
669 netting and acoustic sampling.

670 **Table 1** Comparative performance of GLMMs at predicting high sample representativeness for  
671 *random species subsets* based on AIC<sub>c</sub> model selection. Models are shown up to 95% of  
672 cumulative Akaike weights ( $\omega$ ), with the ones receiving the strongest support ( $\Delta\text{AIC}_c < 2$ ) shown  
673 in bold.

Subset	Model	K	AIC <sub>c</sub>	$\Delta\text{AIC}_c$	$\omega$
Random – SR	<b>Mean no. of surveys</b>	<b>3</b>	<b>127.22</b>	<b>0</b>	<b>0.47</b>
	‘Sampling effort’	4	129.91	2.69	0.12
	No. of plots	3	130.17	2.95	0.11
	Total assemblage abundance	3	130.48	3.26	0.09
	Simpson’s diversity $1/D$	3	130.87	3.65	0.08
	Prop. of rare species	3	131.79	4.57	0.05
	‘Assemblage properties’	5	132.50	5.28	0.03
	Region	3	132.79	5.57	0.03
Random – DCA1	<b>‘Sampling effort’</b>	<b>4</b>	<b>118.64</b>	<b>0.00</b>	<b>0.83</b>
	Mean no. of surveys	3	122.02	3.37	0.15
Random – $J$	<b>Mean no. of surveys</b>	<b>3</b>	<b>91.28</b>	<b>0.00</b>	<b>0.70</b>
	<b>‘Sampling effort’</b>	<b>4</b>	<b>93.40</b>	<b>1.95</b>	<b>0.26</b>

674 **Table 2** Comparative performance of GLMMs at predicting high sample representativeness for  
675 *non-random species subsets* based on AIC<sub>c</sub> model selection. Models are given up to 95% of  
676 cumulative Akaike weights ( $\omega$ ), with the ones receiving the strongest support ( $\Delta\text{AIC}_c < 2$ ) shown  
677 in bold.

Subset	Model	K	AIC <sub>c</sub>	$\Delta\text{AIC}_c$	$\omega$
Nonrandom – SR	<b>Prop. of rare species</b>	<b>3</b>	<b>27.07</b>	<b>0</b>	<b>0.63</b>
	Simpson’s diversity $1/D$	3	30.67	3.60	0.10
	Mean no. of surveys	3	31.43	4.36	0.07
	Total assemblage abundance	3	32.55	5.48	0.04
	Region	3	32.66	5.59	0.04
	Method	3	32.69	5.62	0.04
	No. of plots	3	32.76	5.69	0.04
Nonrandom – DCA1	<b>Total assemblage abundance</b>	<b>3</b>	<b>34.05</b>	<b>0</b>	<b>0.31</b>
	<b>Simpson’s diversity <math>1/D</math></b>	<b>3</b>	<b>35.03</b>	<b>0.98</b>	<b>0.19</b>
	<b>Prop. of rare species</b>	<b>3</b>	<b>35.56</b>	<b>1.51</b>	<b>0.14</b>
	Region	3	36.61	2.56	0.09
	No. of plots	3	36.61	2.56	0.09
	Mean no. of surveys	3	36.77	2.72	0.08
	Method	3	36.79	2.74	0.08
Nonrandom – $J$	<b>Prop. of rare species</b>	<b>3</b>	<b>19.07</b>	<b>0.00</b>	<b>0.88</b>
	‘Assemblage properties’	5	23.17	4.11	0.11

678

679

680 **Table 3** Model-averaged parameter estimates, unconditional standard errors and 95% confidence  
 681 intervals for the best-selected GLMM models ( $\Delta AIC_c < 2$ ) assessing correlates of high sample  
 682 representativeness for random and non-random species subsets.

Subset	Model	Estimate	SE	95% unconditional CI	
				Lower	Upper
Random – SR	Mean no. of surveys	0.58	0.26	0.07	1.10
Random – DCA1	Mean no. of surveys	1.55	0.49	0.59	2.51
	No. of plots	1.65	0.64	0.40	2.89
Random – <i>J</i>	Mean no. of surveys	0.60	0.20	0.21	0.98
	No. of plots	0.16	0.17	-0.18	0.49
Nonrandom – SR	Prop. of rare species	-1.74	0.92	-3.54	0.05
Nonrandom – DCA1	Total assemblage abundance	-0.83	0.57	-1.96	0.29
	Simpson's $1/D$	0.65	0.57	-0.47	1.76
	Prop. of rare species	-0.69	0.76	-2.17	0.79
Nonrandom – <i>J</i>	Prop. of rare species	-6.76	3.90	-14.4	0.89

683

684

685

686 **Figure captions**

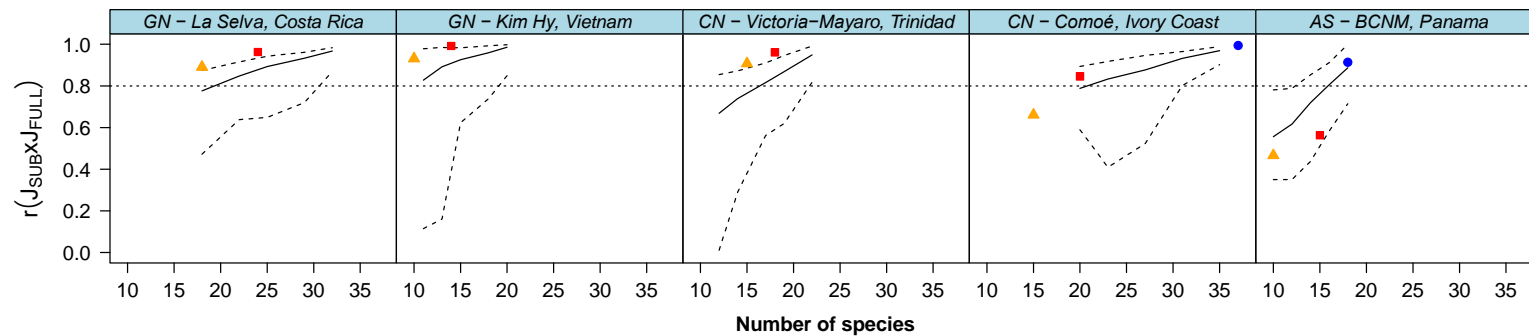
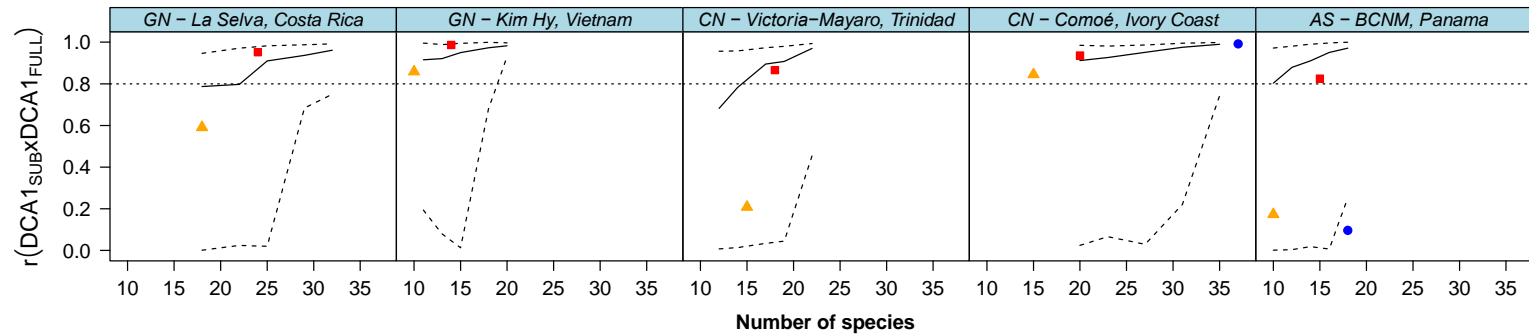
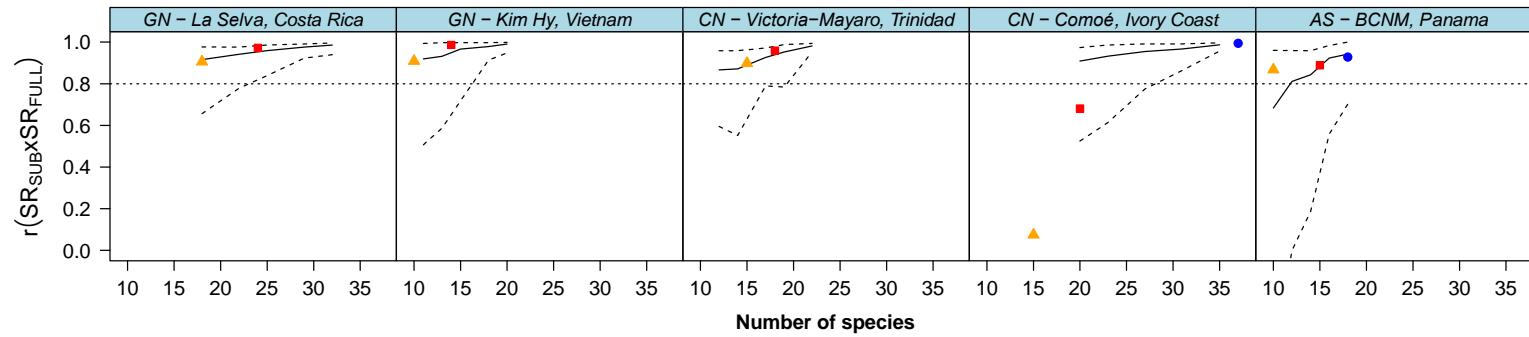
687 **Fig. 1.** The magnitude of correlations across sites between the original dataset and subsets of  
688 data for species richness (SR), and for species composition as represented by the ordination  
689 scores of the first axis of a detrended correspondence analysis (DCA1) and the Jaccard index (*J*).  
690 Shown are examples for bat assemblages from both the New and Old World tropics based on  
691 ground-level mist netting (GN), canopy-level mist netting (CN), and acoustic sampling (AS). See  
692 Figs S1 and S2 for plots for all datasets included in the study. Subsets were generated by deleting  
693 different numbers of species either at random or based on species rarity. The bold line connects  
694 median correlations  $r(\text{SUB} \times \text{FULL})$  for 100 randomly chosen subsets at each of five species pool  
695 sizes; the broken lines indicate upper and lower 95th percentile correlations. Open circles denote  
696 non-random subset correlations, with species eliminated based on up to three abundance  
697 thresholds, whenever applicable (RA < 0.1% (●), < 0.5% (■), < 1% (▲)).

698

699 **Fig. 2.** Boxplots summarizing Pearson product-moment correlations between reduced and full  
700 species sets across all datasets examined in this study. For random species subsets, provided are  
701 both median correlations as well as lower 95<sup>th</sup> percentile correlations at five different species  
702 pool sizes. For non-random subsets, correlations are shown for each of the three threshold levels  
703 of relative abundance (RA) based on which rare species were eliminated from the full species  
704 pool.

705

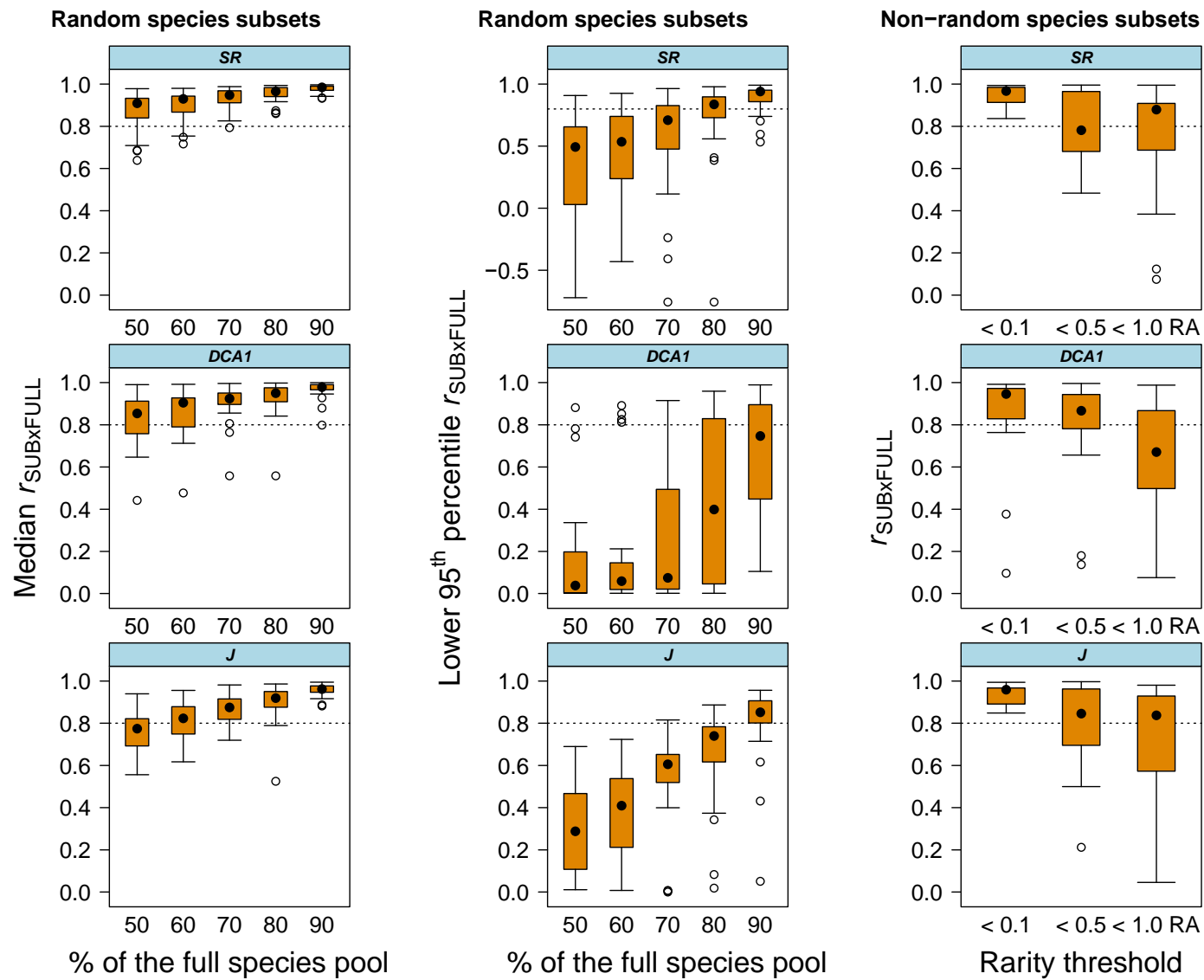
706 **Fig. 3.** Frequency histogram of the proportion of rare species across the datasets analyzed.



707

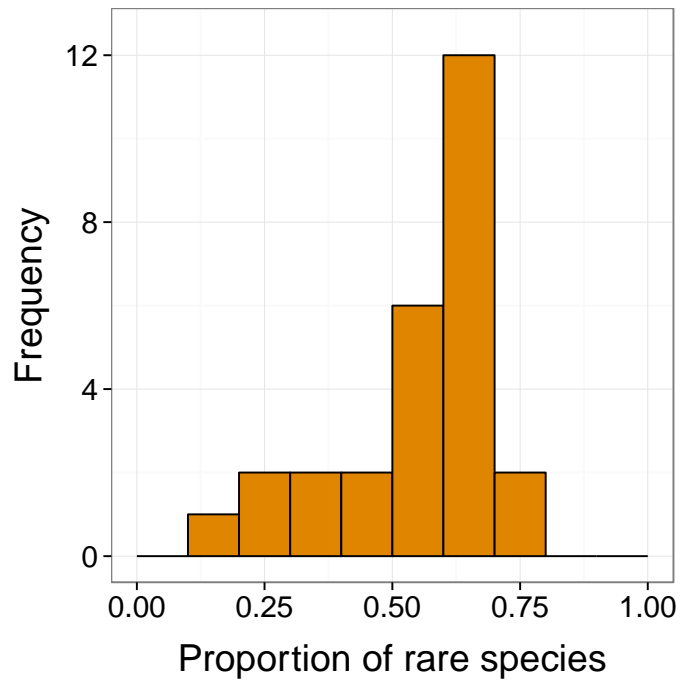
708 **Fig. 1**





709

710 **Fig. 2.**



711

712 **Fig. 3.**