Untested assumptions about within-species sample size and missing data in interspecific studies

László Zsolt Garamszegi & Anders Pape Møller

Abstract Phylogenetic comparative studies rely on species-specific data that often contain missing values and/or differ in sample size among species. These phenomena may violate statistical assumptions about the non-random variance component in sampling effort. A major reason why this assumption is often not fulfilled is because the probability of being sampled (i.e., being captured or observed) may depend on species-specific characteristics. Here, we test this assumption by using information on within-species sample sizes and missing data from five independent comparative datasets of European birds. First, we show that the two estimates of data availability (missing values and within-species sample size) are positively correlated and are associated with research effort in general (the number of papers published). Second, we demonstrate biologically meaningful relationships between data availability and phenotypic traits. For example, population size, risk-taking, and habitat specialization independently predicted within-species sample size. The key determinants of missing data were population size and distribution range. However, data availability was not structured by phylogenetic relationships. These results indicate that the accuracy of sampling is repeatable and distributed non-randomly among species, as several

species-specific attributes determined the probability of observation. Therefore, data availability seems to be a species-specific trait that can be shaped by ecology, life history, and behavior. Such relationships raise issues about non-random sampling, which requires attention in comparative studies.

Introduction

In a recent theoretical paper (Garamszegi and Møller 2011), we emphasized that comparative studies may be biased due to the non-random occurrence of missing data, which can occur if certain species are not represented in the study or if unequal numbers of individuals are sampled for different species. This is because to derive species-specific traits, individual animals must be detected, encountered, and/or captured for sampling, and the likelihood of such events vastly differs among species causing considerable interspecific variation in sampling effort. The probability of sampling may depend on the ecology, life-history, phylogenetic position, and behavior of species, and these characteristics can be hypothesized to determine data availability.

The aim of this study was to assess this hypothetical relationship between within-species sample size, the occurrence of missing data and certain biological predictors. For this evaluation, we chose to focus on European birds, which are frequently studied in a comparative context (Bennett and Owens 2002). We could obtain data on sampling effort from five independent comparative projects that included a large number of species. By comparing within-species sample sizes and the probability of missing data to occur across

L. Z. Garamszegi (✉)
Department of Evolutionary Ecology,
Estacion Biologica de Donana-CSIC,
c/ Americo Vespucio, s/n,
41092 Seville, Spain
e-mail: laszlo.garamszegi@ebd.csic.es

A. P. Møller
Laboratoire d'Ecologie, Systématique et Evolution,
CNRS UMR 8079, Université Paris-Sud,
Bâtiment 362,
91405 Orsay Cedex, France

these studies, we tested if sampling effort varies consistently within-species. We also assessed whether sampling effort is independent of phylogenetic relatedness of species and explored whether different biological predictors can mediate interspecific variation in sampling effort. Particularly, we related within-species sample size and missing data to population size, distribution range, body size, habitat generalism, the degree of adaptation to urban environments and an estimate of risk-taking.

We generally relied on the null hypothesis of random distribution of data availability, and deviations from this hypothesis would yield correlations between the above variables. Particularly, a positive correlation between within-species sample size and instances of missing data would imply that certain species are systematically better studied than others indicating that sampling effort should be considered a species-specific characteristic. Moreover, if sampling effort is dependent on phylogeny, we should detect evidence for certain taxa being better studied than others or for a phylogenetic signal in within-species sample size. Finally, we need to reject the hypotheses that data are available at random with respect to life history, ecology, and behavior of different species if we found evidence for any relationship between the predictors and sampling effort. If these alternative scenarios are supported it would undermine one of the key assumptions of interspecific studies, which requires missing information to occur randomly.

Materials and methods

Estimates of research effort

We searched the literature for comparative studies that focused on bird species with a breeding distribution within the Western Palearctic. This group of animals is ideal for our purposes, because it is often used in comparative tests potentially providing independent data on research effort for a large number of species. Additionally, we could obtain information on predictor variables in a consistent way (see below), allowing for powerful interspecific comparisons. In our literature survey, large interspecific sample size was of great concern, as we attempted to obtain a balanced dataset on missing data, for which we considered the whole European avifauna (526 species). We continued with extracting information from comparative datasets that provided data on at least 100 species. The following sources were included.

In a series of evolutionary studies, one of us (APM) investigated the interspecific correlates of senescence (Møller 2006a, b, 2008d), in which the author used an estimate of sampling effort. This estimate relied on the total number of recoveries and recaptures of banded birds across Europe ranging from 106 to 187,764 among species (CV 0 212.8 %), with a total of 2,640,601 records (http://www.euring.org). These data were provided for 257 species of common birds in Europe, for which longevity records were available. Such information was missing for 269 species.

Another set of studies used a large collection of dead birds (e.g., Garamszegi et al. 2002, 2005; Møller et al. 2003, 2005), which is managed by an expert taxidermist (Johannes Erritzøe) based on casualties brought to him over decades (http://www.birdresearch.dk/). This dataset included 164 species from Europe, which were represented by 1 to 140 individuals that have been processed (1820 records in total, CV 0 191.6 %). For the remaining 362 species, no records were available.

We also obtained data from multiple sources that screened blood of captured adult birds of any sex that breed in Europe for haemoparasite prevalence (Bennett et al. 1982; Haberkorn 1984; Merino et al. 1997; Valkiūnas et al. 1999; Krone et al. 2001; Shurulinkov and Golemansky 2003; Mendes et al. 2005; Palinauskas et al. 2005; Wiersch et al. 2007; Ishak et al. 2008; Krone et al. 2008; Valkiunas et al. 2008). This compiled dataset included 150 bird species that had at least one individual sampled, while the maximum within-species sample size was 370 (CV 0 176.8 %). This resulted in information for the remaining 376 birds to be missing. In total, the provided data included 5,867 individuals.

Iwaniuk and Nelson (2002, 2003) measured endo-cranial volume of a total of 10,364 individuals representing 1,482 species from museum specimens and carcasses. This dataset included 159 European species, with 2–37 specimens measured providing a sample of 1,454 individual birds altogether. Therefore, data were missing for 367 species in this study. Note that the variation in within-species sample size in this dataset was quite low when compared to the other datasets, as the majority of species were represented by ten observations (CV 0 44.3 %, whereas CV > 100 % in the other datasets).

The above datasets correspond to the capture or physical encounter with the subject animals. We also used a set of studies that sampled interspecific variation based on observations of visually detected birds. One of us (APM) initiated a project on flight initiation distances, in which he collected extensive behavioural data in France and Denmark during February–September 2006–2009 (Møller 2008a, b, c; Møller et al. 2008b). So far, 153 species were seen and scored for flight initiation distance (FID) corresponding to 4,126 individual records in total with 1–349 observations per species (CV 0 157.8 %). Only individuals that could be detected with a pair of binoculars from distances exceeding 30 m were subsequently scored for FID to avoid problems of distance affecting detectability and thereby biasing the estimates. The researcher used binoculars in each case to ensure that he did not miss small species at that distance. In

this sampling scheme, 373 bird species from Europe remained undetected. Thus, this last study avoided problems of capture probability affecting sample sizes, although detection probability may have been affected.

From the above five large datasets (termed "senescence", "taxidermist", "malaria", "brain size" and "behavioural" dataset, respectively), we extracted within-species sample sizes in the form of the number of individuals studied (captured or observed), as measures of sampling effort. We used these data on their original but $\log_{10}$-transformed scale to assess the correlation between within-species sample sizes in different datasets. To test the effect of different predictors on species-specific data availability in general, we created a combined estimate across the five datasets. For this purpose, given that the scale of the source studies was quite different, we first standardized sampling effort by adjusting species-specific estimates of sampling effort to a common zero mean and unit variance by using Z scores of the original variables. Then, we could combine within-species sample sizes across the five sources by calculating the mean of the standardised scores. Although we carried out this combination to allow generalizations, we must acknowledge that the importance of within-species data availability may vary depending on interspecific sample size, which we thus disregard in the following analyses with the combined data. However, in principle, we are interested in universal patterns regarding relative differences in sampling effort and not to explain differences between studies in terms of the variance in within-species sample size (see Table 1). We found that non-standardised estimates are strongly correlating, thus combined estimates contain species-specific information and making generalizations for the former phenomenon is meaningful.

We also classified each species in each source with regard to the occurrence of missing data. We first created a complete list European species, and checked if species-specific data were available for each of them in the five datasets. If data were not reported for a species in a particular dataset, it was considered to have missing data in that source. We summed this variable across the five sources by counting how many times species were found with data in different sources. Accordingly, if a species was listed in all sources, it received a score of five, while a species missing in all studies received a score of zero. This summed estimate of missing data was treated as an ordinal variable.

The probability of having any biological information on a given species may depend on the intensity of studies targeting the species in general. Therefore, we also estimated sampling effort in a broader sense. We counted the number of papers published since 1972 on each species as cited in the ISI Web of Science (http://www.isiknowledge.com/). If we repeated this procedure separately for two 15-year periods (1975–1990 and 1991–2006) based on random selection

of 20 species, we obtained highly repeatable estimates (r 0.801, N 20, P<0.001). Similarly, the number of hits counted in a Google Scholar search also revealed strong correlations with the estimates based on the Web of Science hits (r 0.821, N 20, P<0.001). Therefore, our measure of sampling effort in a broader sense can be considered as reliable and not confounded by variations between search engines and publication years.

Phylogenetic and taxonomic information

To test whether any taxon is systematically more studied than others, we categorized each species according to its taxonomical position, for which we relied on the Howard and Moore classification (Dickinson 2003). To this end, we placed species into the following taxonomic groups: genus, family, and order.

Moreover, we also applied phylogenetic tests to estimate evidence for phylogenetic signal in the research effort data. For this purpose, we created a composite phylogeny by using information from Sibley and Ahlquist (1990), which were combined with resolutions from other sources (Sheldon et al. 1992; Seibold et al. 1993; Suhonen et al. 1994; Blondel et al. 1996; Badyaev 1997; Leisler et al. 1997; Cibois and Pasquet 1999; Møller et al. 2001). We applied branch lengths from Sibley and Ahlquist (1990) for higher taxonomic levels. Within families the distance between different genera was set to 3.4 $\Delta T_{50}H$ units and between species within genera to 1.1 $\Delta T_{50}H$ units. Using equal branch lengths, which is equivalent to an assumption of a punctuated model of evolution, gave very similar results.

Biological predictors

Abundance and population density may determine the likelihood of sampling because rare species should be more difficult to observe or capture due to their lower encounter rate than common species. Data on population sizes were obtained from Burfield and van Bommel (2004), who reported the total number of breeding pairs in the Western Palearctic west of the Ural Mountains, obtained in a consistent way from national bird census programs in all countries. We used the mean of the minimum and maximum estimates in that source.

Distribution range may also constitute important constraints for sampling probability, as more people have a higher chance to observe widely distributed species than species with narrow ranges. We estimated the total world geographical range of each species by using the boundaries of their distribution. This was approximated as the area of the shape bounded by the greatest span of latitude and longitude of each species' breeding range, as published in Cramp and Perrins (1977–1994), which was corrected for

the curvature of the earth (see Møller et al. 2008a). In widespread species, Old and New World ranges were calculated separately and subsequently summed to exclude the area of the North Atlantic. This estimate was strongly correlated with estimates originating from direct image analyses, with calculations based on the counts of one-degree grid cells overlain on published breeding distribution maps, and with range size as reported for a subsample of threatened species (Møller et al. 2008a).

Body size can also have direct implications for a suitable within-species sample size, if small-bodied subjects are practically more difficult to detect, collect, or capture than large subjects. Moreover, body size differences can indicate differences in life history (see Vitone et al. 2004; Webster et al. 2004; Møller 2006b; Kamilar et al. 2010), which may also involve factors that influence the probability of sampling. Accordingly, body mass was recorded as the mean mass of males and females from the breeding season, as reported by Cramp and Perrins (1977-1994). If more than a single estimate was reported in that source, we used the one with the largest sample size.

Habitat generalism may also play a role, because species occupying several different habitats are likely to have higher chance to become detected by human observers than specialist species exploiting a single environment. Following the approach of Belliure et al. (2000), we counted the number of different breeding habitats listed in the habitat sections of Cramp and Perrins (1977–1994). Because these habitat categories were defined before the data were collated for this handbook, and because all habitat descriptions were made blindly with respect to the predictions being tested here, these habitat descriptions cannot have caused any bias in the analyses. Several studies have shown that this estimate of habitat specialization provides scientifically useful information (Belliure et al. 2000; Cardillo 2002; Møller et al. 2004, 2011; Phillimore et al. 2007).

Species inhabiting certain environments such as urban habitats may also be easier to collect than specialists of specific habitats that are less accessible for human observers. We defined a species as being urbanized based on information in Cramp and Perrins (1977–1994) and Glutz von Blotzheim and Bauer (1966–1997), as described in detail in Møller (2009). We relied on these standard handbooks because the information was collated in a similar and stringent way, allowing comparison among species. We combined this with personal experience with all the species considered, including information from colleagues interested in urbanization of birds. A species being classified as urbanized had to fulfil the following two criteria: (1) Breeding populations occur inside towns and cities. (2) Population densities in towns and cities are larger than in nearby rural habitats. Based on these criteria, we made a list of 65 urbanized species of breeding birds out of the 526 species

recorded in the Western Palearctic, fulfilling the two criteria listed above. We note explicitly that many species are urbanized in some areas but not in others because urbanization is an ongoing process that initially starts in one area, followed by colonization of urban areas in other parts of the continent or independent urbanization in such cities. Still different indicators of urbanization are positively correlated and significantly repeatable among cities (review in Møller 2012). Regardless, we categorized these species as being urbanized.

Risk taking can also affect encounter rate, because shy species may be more difficult to observe than bold species, as they are more likely to avoid the traps or flee from the observer. Data on flight initiation distances were collected to reflect risk taking, as risk taking birds allow an observer to approach them at a closer distance without fleeing than risk aversive birds (Burger and Gochfeld 1991a, b; Fernández-Juricic et al. 2001, 2002; Blumstein 2003). One of us (APM) used a modified technique developed by Fernández-Juricic, Blumstein and their co-workers to score risk taking behaviour in birds observed in France and Denmark during February–September 2006–2009. A full description of the procedures and three different cross-validations of the data are reported in Møller (2008a, b, c). In brief, when an individual bird had been located with a pair of binoculars, APM moved at a normal walking speed towards the individual, while recording the number of steps (which approximately equals the number of meters). The distance at which the individual took flight was recorded as the flight initiation distance. If the individual was positioned in the vegetation, the height above ground was recorded to the nearest meter. Flight initiation distance was estimated as the Euclidian distance that equals the square root of the sum of the squared horizontal distance and the squared height above ground level (Blumstein 2003).

Data availability differs across predictor variable, as there was missing information for an unequal number of species. Therefore, sample size varies across analyses. The full dataset can be found in the electronic appendix.

Statistical analyses

Continuous variables were $\log_{10}$-transformed for the interspecific analyses, except combined estimates of sampling effort based on within-species sample size and missing data. Pearson's correlations between different estimates of research effort were performed to test for their consistent variation across sources. For this comparison, we present analyses based on the raw-species data without correcting for phylogenetic effects, as phylogenetic generalized least squares (PGLS) modelling revealed that none of the variables that reflect data availability and research effort are confounded by phylogeny (see "Results" section). In such

case, simple correlations are equivalent with the PGLS model that assumes zero covariance due to common descent ($\lambda 00$, see below). We also repeated these tests based on non-parametric statistics, as the assumption about normal distribution was violated in some cases. In a first set of analyses, we searched for correlates of within-species sample size across the five different datasets. Analogically, we also compared missing data between sources, for which we used contingency tables. Then, we used the combined estimates across sources and tested for a relationship between standardised within-species sample size and missing data by using correlation approaches (overall missing data is considered an ordinal variable). We also assessed the relationship of these variables with the general estimate of research efforts based on the number of Web of Science hits.

Nested analyses of variance (ANOVAs) on the mean of the standardized within-species sample size and on the summed estimate of missing data were performed to investigate if there was any taxa-specific variation at the above species level in data availability (order, family, and genus). To evaluate the possibility that closely related species are more likely to share sample sizes than more distantly related species, we used phylogenetic modelling. This exercise was carried out by using the PGLS approach, which incorporates a matrix of the expected covariances among species due to shared phylogenetic structure (Martins and Hansen 1997; Pagel 1999). This method enabled us to estimate the degree of phylogenetic signal within the data in the form of lambda (λ) statistics (Freckleton et al. 2002). We evaluated the likelihood surface depicted by a range of evolutionary models that assume different degrees of phylogenetic dependence. Models relying on $\lambda 00$ are equivalent with phylogenetic independence, while models constrained with increasing (up to 1) λ imply stronger phylogenetic structure in the data at hand.

We also used the PGLS method to test for correlations between the biological predictors and combined estimates of data availability. We conducted these analyses by setting the phylogenetic scaling factor (λ) to its maximum likelihood estimate that had been evaluated for the model investigated. We did this although we found that sampling is independent of phylogeny, because there might be phylogenetic structure in the predictor variables (see Abouheif 1999). We fitted all phylogenetic models in the R statistical computing environment, with additional unpublished functions by R. Freckleton (University of Sheffield, available upon request) for the PGLS procedure developed for continuous and discrete variables and multivariate models.

To determine the strength and direction of the predicted relationships, we estimated effect sizes (such as correlation effect size "r" sensu Cohen 1988) and the associated 95 % confidence intervals (95 % CI) for each particular relationship. We preferred focusing on effect sizes rather than significance levels because models with a large number of species would likely support weak, but statistically significant effects that only explain a small proportion of the variance (Nakagawa and Cuthill 2007; Freckleton 2009; Garamszegi et al. 2009a). Therefore, we focused on the strength of the relationships (effect size) by considering the precision (95 % CI) by which it can be measured given the available data. However, for illustrative purposes, we also present P values, but avoid discussing statistical significance. We combined effect sizes across the five sources of sample size data by adopting a meta-analytic approach, in which effect sizes were averaged by using sample size as weights (Garamszegi 2006).

Results

Consistent variation in within-species sample size and missing data

Overall, within-species sample sizes were positively correlated with each other across the five independent datasets (i.e., effect sizes and confidence intervals fall in the positive range, Table 1). This indicates that consistently larger

Table 1 The correlations of within-species sample size across five large-scale sources of interspecific data

Cells above the diagonal show effect sizes based on Pearson's correlation coefficient and the associated 95 % confidence interval (lower/upper). Below the diagonal, Kendall rank order non-parametric correlations are given with sample sizes. *$P<$ 0.05, **$P<0.01$, ***$P<0.001$

Dataset	Senescence	Taxidermist	Malaria	Brain size	Behavior
Senescence		0.480***	0.579***	0.048	0.424***
		0.344/0.596	0.450/0.685	−0.132/0.225	0.275/0.553
Taxidermist	0.487***		0.471***	0.092	0.355***
	(145)		0.287/0.622	−0.124/0.299	0.170/0.515
Malaria	0.603***	0.462***		0.203	0.250*
	(126)	(85)		−0.041/0.424	0.043/0.437
Brain size	0.026	0.099	0.202		0.220
	(121)	(85)	(66)		−0.023/0.439
Behaviour	0.430***	0.362***	0.253*	0.171	
	(135)	(100)	(88)	(66)	

sample size is available for some species while others are systematically represented with smaller sample size. The only exception was sample size from the "brain size" data, which showed considerably weaker effects. This might be due to the fact that the variation in sample size was smaller in this study than in others. The combined effect size across all correlations was r 0.333, implying that within-sample size is generally repeatable at this magnitude.

The picture was very similar for missing data (Table 2), in which "brain size" data also followed the general pattern. Therefore, it is very likely that if there is missing information for a species in one study, there will be no data for the same species in another study. On average, the magnitude of this effect was r 0.360.

The above analyses revealed that within-species sample sizes and missing data are species-specific traits, thus combining them across sources makes biological sense. These combined measures also showed correlations with each other (r 0.415, N 339, 95 % CI 0.323/0.499, $P < 0.001$, Fig. 1), and with the estimate of general research effort (within-species sample size: r 0.504, N 339, 95 % CI 0.420/0.579, $P < 0.001$; missing data: r 0.659, N 526, 95 % CI 0.608/0.705, $P < 0.001$; Fig. 2). These results further suggested that data availability varies systematically among species and reflects sampling effort in a broad sense.

Data availability and taxonomic and phylogenetic relationships

In our further search for taxonomic structure, we found that higher (i.e. above-species) taxonomic levels are unlikely to explain considerable variation (i.e., >10 %) in within-species sample size (nested ANOVA: order, $F_{18,165}$ 0.714, r 0.014, 95 % CI $-0.093/0.120$, P 0.793; family, $F_{42,165}$ 0.886, r 0.023, 95 % CI $-0.083/0.130$, P 0.670; genus, $F_{113,165}$ 0.657, $r < 0.001$, 95 % CI $-0.106/0.107$, P 0.991). That was also the case for overall missing data (nested ANOVA: order, $F_{18,303}$ 1.242, r 0.053, 95 % CI $-0.033/0.138$, P 0.226; family, $F_{50,303}$ 1.286, r 0
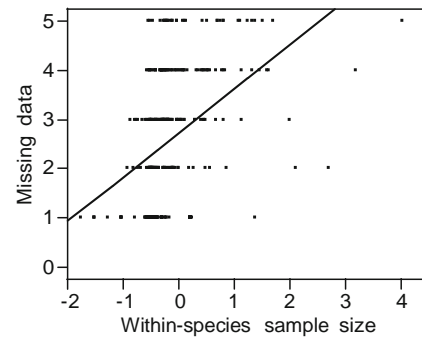

Fig. 1 The relationship between standardized within-species sample size and missing data. Data are combined estimates from five independent comparative projects. The line is the regression line

0.071, 95 % CI $-0.015/0.155$, P 0.106; genus, $F_{154,303}$ 0.907, r 0.014, 95 % CI $-0.072/0.099$, P 0.752).

We traced data availability along the phylogenetic tree of species and fitted different evolutionary models that consider different scaling of the lambda parameter. When examining the potential for phylogenetic signal in the within-species sample size data, we found that the maximum likelihood estimate of λ was 0. For missing data, the maximum likelihood was also achieved with a model that relied on λ 0. We tested whether models relying on the maximum likelihood estimate of lambda were considerably better than models completely incorporating phylogenetic effects by using λ 1. In both cases, there was strong evidence against the model that applied full phylogenetic correction (within-species sample size: likelihood ratio 284.086, r 0.915, 95 % CI 0.896/0.931, N 339 species, $P < 0.001$; missing data: likelihood ratio 333.294, r 0.796, 95 % CI 0.762/0.825, N 526 species, $P < 0.001$). Therefore, we detected no evidence for data availability being structured by the phylogenetic relatedness of species.

Biological predictors

The pair-wise relationships between combined within-species sample size and the seven predictors are given in

Table 2 The associations of missing data across comparative databases

Pearson correlation effect sizes and the associated 95 % confidence interval were derived from the χ^2 statistics (with d.f. 1, 524) of contingency tables. Sample size is 526, and $P < 0.001$ in all cases

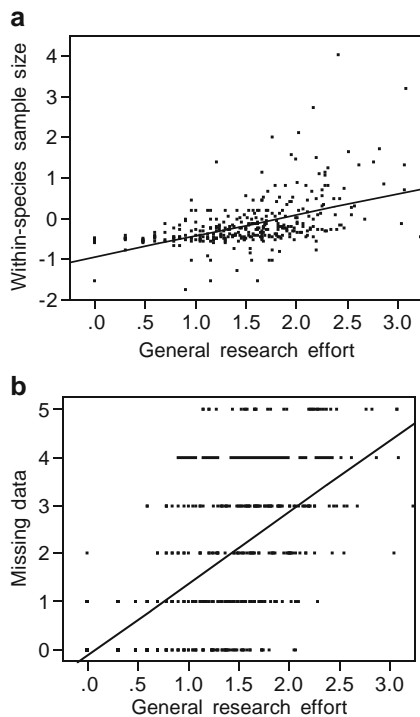| Dataset	Senescence	Taxidermist	Malaria	Brain size	Behaviour
Senescence					
Taxidermist	0.365 0.289/0.437				
Malaria	0.459 0.389/0.524	0.340 0.262/0.413			
Brain size	0.365 0.289/0.437	0.311 0.232/0.386	0.186 0.102/0.267		
Behaviour	0.528 0.463/0.587	0.464 0.394/0.529	0.401 0.327/0.470	0.177 0.093/0.259	

Fig. 2 The relationship between sampling effort and general research effort, when sampling effort is estimated as a within-species sample size, and b missing data. General research effort is measured as the number of papers published (the scale is $\log_{10}$-transformed). The line is the regression line

Table 3. These correlations yielded effect sizes of considerable importance for several variables. In the majority of cases, we can rely with 95 % confidence that variation in within-species sample size does not occur randomly with respect to the predictor variables (i.e., effect size is different from zero). However, the explanatory role of different predictors varies: some are represented by intermediate effects (e.g., population size, urban life), while others have smaller effects (e.g., distribution range, habitat generalism, and flight distance). Lambda estimates indicated that these effect sizes were independent of the phylogenetic relationships of species.

We detected similar tendencies when analyzing the determinants of variation in missing data. In fact, the predictor variables with considerable role (population size, distribution range, habitat generalism) explained more variance in data availability, as indicated by the corresponding effect sizes that were generally larger in magnitude than in the case of within-species sample size (Table 3). The exception was the relationship with flight distance, which covered an effect size with a 95 % CI that represented a weaker role. Moreover, as lambda parameters indicated, the relationships should be adjusted for a slight phylogenetic signal in the data (potentially caused by the predictors that vary nonrandomly with regard to the phylogeny).

Predictor variables can be correlated among themselves, and such correlation structure has to be taken into consideration. For example, population sizes and distribution ranges depend on body size, or urban life partially results from reduced habitat generalism. Moreover, making conclusions for population density only makes sense, if population size is estimated when distribution range is held constant. Therefore, we also performed a multi-predictor analysis to remove the confounding role caused by the association between the predictors to test for their independent effects. The multi-predictor models provided a slightly different picture compared to the pair-wise relationships (Table 4). In the within-species sample size model, the effect size for urban life decreased, while it increased for body size. In the missing data model, several variables lost importance, but still confidence range for population size and distribution range fell within the range of small-to-large effects (i.e., between 0.1 and 0.5). Importantly, the multi-predictor models were able to explain ∼50 % of the variance in sampling effort. Therefore, some of the correlations in the pair-wise comparisons (Table 3) might have spuriously emerged because of the intercorrelation between the predictors. However, interpretations for correlating predictors in a multiple regression should be made with caution if they are measured with different errors, as variables with less error can overwhelm variables with more error (Freckleton 2011). Unfortunately, with the available correlative data, we cannot trace these effects further.

Discussion

In this analysis of data availability in comparative studies, we have demonstrated that within-species sample size and missing data vary in a consistent manner across studies. This repeatable variation implies that there are species for which data are generally more accessible than for others. The within-species sample size and the occurrence of missing data are likely to reflect the same phenomenon. Missing data may represent the case when within-species sample size is zero, while unequal sample sizes can also be considered as a consequence of missing data, because missing information for unequal numbers of individuals serve as a basis for variation in within-species sample size. As a result of this phenomenon, we found that data availability constrains the intensity of research effort (Fig. 2). Species for which sample size is generally limited or data are often missing are more difficult to study, which has a strong impact on the amount of scientific information accumulated on these species. Therefore, future comparative studies should consider whether availability of data is subject to random noise or can cause systematic bias in interspecific samples.

Interspecific variation in sample size and missing data can be described by true biological variables, as we demonstrated

Table 3 The relationship between study effort (within-species sample size and missing data) combined across five sources and eight potential biological predictors that were hypothesized to mediate the probability of being observed

Pearson correlation effect sizes and the associated 95 % confidence intervals are based on phylogenetically adjusted correlations. Lambda values, which were first estimated based on maximum likelihood and then used in the model, are given. **$P<0.01$, ***$P<0.001$

Predictor	Within-species sample size	Missing data
Population size	$\lambda0$0.066, N0336	$\lambda0$0.274, N0487
	r00.351*** (0.254/0.442)	r00.665*** (0.612/0.712)
Distribution	$\lambda0$0.000, N0337	$\lambda0$0.134, N0524
	r00.192*** (0.087/0.293)	r00.391*** (0.316/0.461)
Body mass	$\lambda0$0.000, N0333	$\lambda0$0.155, N0496
	r00.000 (−0.108/0.107)	r0−0.014 (−0.101/0.075)
Habitat generalism	$\lambda0$0.000, N0142	$\lambda0$0.332, N0249
	r00.137 (−0.028/0.295)	r00.368*** (0.256/0.471)
Urban life	$\lambda0$0.000, N0339	$\lambda0$0.261, N0526
	r00.367*** (0.272/0.456)	r00.432*** (0.360/0.499)
Flight distance	$\lambda0$0.000, N0153	$\lambda0$0.000, N0153
	r0−0.215** (−0.362/−0.059)	r0−0.098 (−0.253/0.062)

strong evidence that data availability is determined by several species-specific characteristics. For example, species with larger population size and distribution range have higher probability to be represented with a larger sample size in comparative data than species with smaller population size and distribution area. This pattern is in accordance with the hypothesis that less abundant and rare species are generally difficult to capture or observe. Moreover, body size also seems to explain some additional variation in within-species sample size. Therefore, we obtained support for the hypothesis that species with larger body size are more likely to be studied than species with smaller size. This is an important finding given that body size correlates with an entire suit of life-history traits (Bennett and Owens 2002; Jeschke and Kokko 2009), implying that data availability can depend on how species organize their life along the "fast" and "slow" life-history continuum.

This study provides further evidence that sampling effort is shaped by risk aversion. It has been suspected for a while that trappability may be affected by fear response of animals, as shy individuals are more likely to show consistent trap-averse behavior and may be more difficult to capture than bold individuals (Wilson et al. 1993). At the intraspecific level, the relationship between trappability and boldness has been shown

to exist in fishes (Wilson et al. 1993), mammals (Réale et al. 2000; Malmkvist and Hansen 2001), and birds (Mills and Faure 2000; Garamszegi et al. 2009b). Such a relationship can have consequences for non-random sampling, as the pool of successfully caught animals is more likely to include bold individuals than the pool of animals that avoid the trap. This can cause systematic bias, if the investigated traits are also related to risk taking (Garamszegi et al. 2009b). Here, we provided the first evidence that such sampling bias caused by shy-bold behaviors could occur in interspecific datasets. Flight distance is an inverse estimate of risk taking (Fernández--Juricic et al. 2001, 2002; Blumstein 2003). The same species consistently show the same escape behavior in the presence of a potential predator, and thus it makes biological sense to discriminate between risk prone and risk aversive species (Møller 2008a, b, c). Such variation was found to relate to sampling effort, as risk prone bird species are usually represented by larger within-species sample size than risk aversive species. Interspecific variation in flight distance reflects variations in several other species-specific phenotypic traits (Møller 2008a, b, c). Therefore, if within-species sample size is biased by risk taking, it will also be structured by the correlates of flight distance.

Table 4 Regression models of the linear relationship between sampling effort (within-species sample size and missing data) and seven potential biological predictors

Pearson correlation effect sizes and the associated 95 % confidence intervals are taken from the PGLS model with the most appropriate lambda settings. *$P<0.05$, **$P<0.01$, ***$P<0.001$

	Within species sample size	Missing data
Full model	$\lambda0$0.000, $F_{6,95}$010.97***	$\lambda0$0.000, $F_{6,95}$019.98***
	$R^2$00.425	$R^2$00.574
Predictors		
Population size	r00.430*** (0.246/0.585)	r00.624*** (0.480/0.735)
Distribution	r00.219* (0.014/0.407)	r00.350*** (0.155/0.519)
Body mass	r00.225* (0.020/0.412)	r00.074 (−0.134/0.276)
Habitat generalism	r00.094 (−0.114/0.294)	r00.020 (−0.187/0.225)
Urban life	r00.088 (−0.120/0.289)	r00.096 (−0.112/0.296)
Flight distance	r0−0.243* (−0.427/−0.039)	r00.071 (−0.137/0.273)

We were unable to provide evidence for data availability being structured by the phylogenetic relationships of species, thus missing data and within species variation can occur randomly with respect to phylogeny. Therefore, closely related species do not necessarily impose similar challenges for data collection. This is an important finding, because it implies that the statistical control for phylogeny does not automatically remove errors due to unbalanced sampling. In other words, if data availability is not related to phylogeny, it does not override biases due to other factors (e.g., morphology, life-history, and behavior), which can still confound estimations about correlations or ancestral states (Garamszegi and Møller 2011). For example, comparative results are likely to change when the sample is tilted toward better-studied species, because these species have wider distributions, larger bodies, and take higher risk in the presence of predators. Justification for how non-random missing data can affect parameter estimation regardless of phylogeny can be found in Nakagawa and Freckleton (2008, 2011), while a case providing evidence for captured and non-captured fractions of birds showing different correlations is given in Garamszegi et al. (2009b). The lack of phylogenetic signal in the sample size data does not undermine the importance of phylogenetic corrections in the focal analyses. Closely related species can share several phenotypes, and thus it can happen that inadequately sampled species are associated with well-studied species with similar phenotypes when phylogenetic corrections are warranted (Felsenstein 1985). We note that the inability to demonstrate strong phylogenetic effects in this study may be specific to the data at hand, and for other animal taxa, it remains an empirical question if available within-species sample sizes are structured by common ancestry. In general, the causes of the variance in within-species sample size and missing data should be identified for each interspecific data set.

Our entire analysis was based on birds, thus making generalizations beyond this particular taxonomic group is unwarranted. Other vertebrate or invertebrate clades may have different characteristics that affect sampling. Therefore, future studies should further examine the factors responsible for shaping intra-specific sample sizes and the occurrence of missing data in comparative studies of mammals, fishes, or insects. However, on theoretical grounds, we could expect that sampling bias due to the unequal probability of individuals of different species being captured is likely to exist in non-avian taxa as well, because species can differ in several aspects of morphology, ecology, and behavior, which can all have consequences for trappability (Garamszegi and Møller 2011). For example, if a comparative study of primates primarily targets terrestrial species due to the difficulty of capturing arboreal species, the interspecific data will be loaded with biases due to body mass, social groups size, and locomotion as well as phylogeny

(Arnold and Nunn 2010). In fact, issues about non-random sampling should be considered when working on plants (Westoby 1999, 2002).

In conclusion, we derived empirical evidence in birds that available within-species sample sizes can depend on certain ecological, life history, and behavioral factors, suggesting that it is invalid to assume in comparative studies that variation in data availability occurs randomly. The non-random appearance of missing information poses several problems in interspecific comparative contexts. These problems include issues about statistical power and measurement error and also about biased parameter estimation and model selection results as well as erroneous ancestral state reconstructions (see details in Garamszegi and Møller 2011). We infer that these problems are currently being severely neglected in the comparative literature. Although data limitations arise from technical constrains and the shortcomings of non-random sampling are often hard to avoid during data collection, some practical advice on appropriate data collection can be followed in order to achieve more balanced and phylogenetically targeted samples (Mitani et al. 1996; Ackerly 2000; Maddison 2000; Harmon and Losos 2005; Arnold and Nunn 2010). By doing so, the observer could potentially determine the optimal within- and between-species sample sizes and the list of species to study for the biological question and animal taxa at hand, which would in turn minimize biases due to uneven sampling. However, if targeted sampling is not fully accomplishable, future interspecific studies can also benefit if issues about non-random sampling are considered at the level of analysis. For example, one can estimate diagnostic statistics to investigate if the available data represent a random sample of the natural diversity and more specifically determine whether variation in within-species sample size causes bias. In cases when non-random sampling is detected or suspected, data imputation methods can be applied to replace missing values by considering the phylogenetic relationships of species (Fisher et al. 2003). Moreover, phylogenetic methods are also available for incorporating within-species variances into the analyses (e.g., Ives et al. 2007; Felsenstein 2008; Paradis 2011) or that can balance for unequal sample sizes (Garamszegi and Møller 2010). Finally, issues about non-random study effort should be considered at the level of interpretation, as care should be made when generalizing beyond the data (Freckleton 2009). These practical suggestions are discussed in greater detail in a companion theoretical paper (Garamszegi and Møller 2011).

References

Abouheif E (1999) A method for testing the assumption of phylogenetic independence in comparative data. Evol Ecol Res 1:895–909

Ackerly DD (2000) Taxon sampling, correlated evolution, and independent contrasts. Evolution 54:1480–1492

Arnold C, Nunn CL (2010) Phylogenetic targeting of research effort in evolutionary biology. Am Nat 176:601–612

Badyaev AV (1997) Altitudinal variation in sexual dimorphism: a new patern and alternative hypotheses. Behav Ecol 8:675–690

Belliure J, Sorci G, Møller AP, Clobert J (2000) Dispersal distances predict subspecies richness in birds. J Evol Biol 13:480–487

Bennett PM, Owens IPF (2002) Evolutionary ecology of birds. Oxford University Press, Oxford

Bennett GF, Thommes F, Blancou J, Artois M (1982) Blood parasites of some birds from the Lorraine region, France. J Wildl Dis 18:81–88

Blondel J, Catzeflis F, Perret P (1996) Molecular phylogeny and the historical biogeography of the warblers of the genus Sylvia (Aves). J Evol Biol 9:871–891

Blumstein DT (2003) Flight-initiation distance in birds is dependent on intruder starting distance. J Wildl Manag 67:852–857

Burfield I, van Bommel F (2004) Birds in Europe. BirdLife International, Cambridge

Burger J, Gochfeld M (1991a) Human activity influence and diurnal and nocturnal foraging of sanderlings (Calidris alba). Condor 93:259–265

Burger J, Gochfeld M (1991b) Human distance and birds: tolerance and response distances of resident and migrant species in India. Environ Conserv 18:158–165

Cardillo M (2002) The life-history basis of latitudinal diversity gradients: how do species traits vary from the poles to the equator. J Anim Ecol 71:79–87

Cibois A, Pasquet E (1999) Molecular analysis of the phylogeny of 11 genera of the Corvidea. Ibis 141:297–306

Cohen J (1988) Statistical power analysis for the behavioural sciences, 2nd edn. Lawrence Erlbaum Associates, Hillsdale

Cramp S, Perrins CM (eds) (1977–1994) The birds of the Western Palearctic, vol 4-9. Oxford University Press, Oxford

Dickinson EC (ed) (2003) The Howard and Moore complete checklist of the birds of the world, 3rd edn. Princeton University Press, Princeton

Felsenstein J (1985) Phylogenies and the comparative method. Am Nat 125:1–15

Felsenstein J (2008) Comparative methods with sampling error and within-species variation: contrasts revisited and revised. Am Nat 171:713–725

Fernández-Juricic E, Jimenez MD, Lucas E (2001) Alert distance as an alternative measure of bird tolerance to human disturbance: implications for park design. Environ Conserv 28:263–269

Fernández-Juricic E, Jimenez MD, Lucas E (2002) Factors affecting intra- and inter-specific variations in the difference between alert and flight distances in forested habitats. Can J Zool 80:1212–1220

Fisher DO, Blomberg SP, Owens IPF (2003) Extrinsic versus intrinsic factors in the decline and extinction of Australian marsupials. Proc R Soc Lond B 270:1801–1808

Freckleton RP (2009) The seven deadly sins of comparative analysis. J Evol Biol 22:1367–1375

Freckleton RP (2011) Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error. Behav Ecol Sociobiol 65:91–101

Freckleton RP, Harvey PH, Pagel M (2002) Phylogenetic analysis and comparative data: a test and review of evidence. Am Nat 160:712–726

Garamszegi LZ (2006) Comparing effect sizes across variables: generalization without the need for Bonferroni correction. Behav Ecol 17:682–687

Garamszegi LZ, Møller AP (2010) Effects of sample size and intraspecific variation in phylogenetic comparative studies: a meta-analytic review. Biol Rev 85:797–805

Garamszegi LZ, Møller AP (2011) Nonrandom variation in within-species sample size and missing data in phylogenetic comparative studies. Syst Biol 60:876–880

Garamszegi LZ, Møller AP, Erritzøe J (2002) Coevolving avian eye size and brain size in relation to prey capture and nocturnality. Proc R Soc Lond B 269:961–967

Garamszegi LZ, Eens M, Erritzøe J, Møller AP (2005) Sperm competition and sexually size dimorphic brains in birds. Proc R Soc Lond B 272:159–166

Garamszegi LZ, Calhim S, Dochtermann N, Hegyi G, Hurd PL, Jørgensen C, Kutsukake N, Lajeunesse MJ, Pollard KA, Schielzeth H, Symonds MRE, Nakagawa S (2009a) Changing philosophies and tools for statistical inferences in behavioral ecology. Behav Ecol 20:1363–1375

Garamszegi LZ, Eens M, Török J (2009b) Behavioural syndromes and trappability in free-living collared flycatchers, Ficedula albicollis. Anim Behav 77:803–812

Glutz von Blotzheim UN, Bauer KM (eds) (1966–1997) Handbuch der Vögel Mitteleuropas, vol 1-15. Aula-Verlag, Wiebelsheim

Haberkorn A (1984) Observations on malaria in European perching birds (Passeriformes). Zbl Bakt Mik Hyg M 256:288–295

Harmon LJ, Losos JB (2005) The effect of intraspecific sample size on type I and type II error rates in comparative studies. Evolution 59:2705–2710

Ishak HD, Dumbacher JP, Anderson NL, Keane JJ, Valkiūnas G, Haig SM, Tell LA, Sehgal RNM (2008) Blood parasites in owls with conservation implications for the spotted owl (Strix occidentalis). PLoS One 3:e2304

Ives AR, Midford PE, Garland T (2007) Within-species variation and measurement error in phylogenetic comparative methods. Syst Biol 56:252–270

Iwaniuk AN, Nelson JE (2002) Can endocranial volume be used as an estimate of brain size in birds? Can J Zool 80:16–23

Iwaniuk AN, Nelson JE (2003) Developmental differences are correlated with relative brain size in birds: a comparative analysis. Can J Zool 81:1913–1928

Jeschke JM, Kokko H (2009) The roles of body size and phylogeny in fast and slow life histories. Evol Ecol 23:867–878

Kamilar JM, Bribiescas RG, Bradley BJ (2010) Is group size related to longevity in mammals? Biol Lett 6:736–739

Krone O, Priemer J, Streich J, Sommer P, Langgemach T, Lessow O (2001) Haemosporida of birds of prey and owls from Germany. Acta Protozool 40:281–289

Krone O, Waldenström J, Valkiunas G, Lessow O, Müller K, Iezhova TA, Fickel J, Bensch S (2008) Haemosporidian blood parasites in European birds of prey and owls. J Parasitol 94:709–715

Leisler B, Heidrich P, Schulze-Hagen K, Wink M (1997) Taxonomy and phylogeny of reed warblers (genus Acrocephalus) based on mtDNA sequences and morphology. J Ornithol 138:469–496

Maddison WP (2000) Testing character correlation using pairwise comparisons on a phylogeny. J Theor Biol 202:195–204

Malmkvist J, Hansen SW (2001) The welfare of farmed mink (Mustela vison) in relation to behavioural selection: a review. Anim Welf 10:41–52

Martins EP, Hansen TF (1997) Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am Nat 149:646–667

Mendes L, Piersma T, Lecoq M, Spaans B, Ricklefs RE (2005) Disease-limited distributions? Contrasts in the prevalence of avian malaria in shorebird species using marine and freshwater habitats. Oikos 109:396–404

Merino S, Potti J, Fargallo JA (1997) Blood parasites of passerine birds from central Spain. J Wildl Dis 33:638–641

Mills AD, Faure JM (2000) Ease of capture in lines of Japanese quail (Coturnix japonica) subjected to contrasting selection for fear or sociability. Appl Anim Behav Sci 69:125–134

Mitani JC, GrosLouis J, Manson JH (1996) Number of males in primate groups: comparative tests of competing hypotheses. Am J Primatol 38:315–332

Møller AP (2006a) Senescence in relation to latitude and migration in birds. J Evol Biol 20:750–757

Møller AP (2006b) Sociality, age at first reproduction and senescence: comparative analyses of birds. J Evol Biol 19:682–689

Møller AP (2008a) Flight distance and blood parasites in birds. Behav Ecol 19:1305–1313

Møller AP (2008b) Flight distance and population trends in breeding birds. Behav Ecol 19:1095–1102

Møller AP (2008c) Flight distance of urban birds, predation and selection for urban life. Behav Ecol Sociobiol 63:63–75

Møller AP (2008d) Relative longevity and field metabolic rate in birds. J Evol Biol 21:1379–1386

Møller AP (2009) Successful city dwellers: a comparative study of the ecological characteristics of urban birds in the Western Palearctic. Oecologia 159:849–858

Møller AP (2012) Behavioral and ecological predictors of urbanization. In: Gil D, Brumm H (eds) Avian urban ecology. Oxford University Press, Oxford (in press)

Møller AP, Merino S, Brown CR, Robertson RJ (2001) Immune defense and host sociality: a comparative study of swallows and martins. Am Nat 158:136–145

Møller AP, Erritzøe J, Saino N (2003) Seasonal changes in immune response and parasite impact on hosts. Am Nat 161:657–671

Møller AP, Martín-Vivaldi M, Soler JJ (2004) Parasitism, host immune defence and dispersal. J Evol Biol 17:603–612

Møller AP, Erritzøe J, Garamszegi LZ (2005) Covariation between brain size and immunity in birds: implications for brain size evolution. J Evol Biol 18:223–237

Møller AP, Garamszegi LZ, Spottiswoode C (2008a) Genetic similarity, breeding distribution range and sexual selection. J Evol Biol 21:213–225

Møller AP, Nielsen JT, Garamszegi LZ (2008b) Risk taking by singing males. Behav Ecol 19:41–53

Møller AP, Garamszegi LZ, Peralta-Sanchez JM, Soler JJ (2011) Migratory divides and their consequences for dispersal, population size and parasite–host interactions. J Evol Biol 24:1744–1755. doi:10.1111/j.1420-9101.2011.02302.x

Nakagawa S, Cuthill IC (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev 82:591–605

Nakagawa S, Freckleton R (2008) Missing inaction: the dangers of ignoring missing data. Trends Ecol Evol 23:592–596. doi:10.1016/j.tree.2008.06.014

Nakagawa S, Freckleton RP (2011) Model averaging, missing data and multiple imputation: a case study for behavioural ecology. Behav Ecol Sociobiol 65:103–116

Pagel M (1999) Inferring the historical patterns of biological evolution. Nature 401:877–884

Palinauskas V, Markovets MY, Kosarev VV, Efremov VD, Sokolov LV, Valkiûnas G (2005) Occurrence of avian haematozoa in Ekaterinburg and Irkutsk districts of Russia. Ekologija 4:8–12

Paradis E (2011) Analysis of phylogenetics and evolution with R, 2nd edn. Springer, Berlin

Phillimore AB, Orme CDL, Davies RG, Hadfield JD, Reed WJ, Gaston KJ, Freckleton RP, Owens IPF (2007) Biogeographical basis of recent phenotypic divergence among birds: a global study of subspecies richness. Evolution 61:942–957

Réale D, Gallant BY, Leblanc M, Festa-Bianchet M (2000) Consistency of temperament in bighorn ewes and correlates with behaviour and life history. Anim Behav 60:589–597

Seibold I, Helbig AJ, Wink M (1993) Molecular systematics of falcons (family Falconidae). Naturwissenschaften 80:87–90

Sheldon FH, Slikas B, Kinnarney M, Gill FB, Zhao E, Silverin B (1992) DNA–DNA hybridization evidence of phylogenetic relationships among major lineages of Parus. Auk 109:173–185

Shurulinkov P, Golemansky V (2003) Plasmodium and Leucocytozoon (Sporozoa: Haemosporida) of wild birds in Bulgaria. Acta Protozool 42:205–214

Sibley CG, Ahlquist JE (1990) Phylogeny and classification of birds: a study in molecular evolution. Yale University Press, New Haven

Suhonen J, Alatalo RV, Gustafsson L (1994) Evolution of foraging ecology in Fennoscandian tits (Parus spp.). Proc R Soc Lond B 258:127–131

Valkiûnas G, Iezhova T, Golemansky V, Pilarska D, Zehtindjiev P (1999) Blood protozoan parasites (Protozoa: Kinetoplastida and Haemosporida) in wild birds from Bulgaria. Acta Zool Bulg 51:127–129

Valkiunas G, Iezhova TA, Krizanauskiene A, Palinauskas V, Sehgal RNM, Bensch S (2008) A comparative analysis of microscopy and PCR-based detection methods for blood parasites. J Parasitol 94:1395–1401

Vitone ND, Altizer S, Nunn CL (2004) Body size, diet and sociality influence the species richness of parasitic worms in anthropoid primates. Evol Ecol Res 6:183–199

Webster AJ, Gittleman JL, Purvis A (2004) The life history legacy of evolutionary body size change in carnivores. J Evol Biol 17:396–407

Westoby M (1999) Generalization in functional plant ecology: the species sampling problem, plant ecology strategy schemes, and phylogeny. In: Pugnaire FI, Valladares F (eds) Handbook of functional plant ecology. M. Dekker, New York, pp 847–872

Westoby M (2002) Choosing species to study. Trends Ecol Evol 17:587

Wiersch SC, Lubjuhn T, Maier WA, Kampen H (2007) Haemosporidian infection in passerine birds from Lower Saxony. J Ornithol 148:17–24

Wilson DS, Coleman K, Clark AB, Biederman L (1993) Shy bold continuum in pumpkinseed sunfish (Lepomis gibbosus)—an ecological study of a psychological trait. J Comp Psychol 107:250–260