



Title	Rethinking birdsong evolution : meta-analysis of the relationship between song complexity and reproductive success
Author(s)	Soma, Masayo; Zsolt Garamszegi, László
Citation	Behavioral Ecology, 22(2), 363-371 https://doi.org/10.1093/beheco/arq219
Issue Date	2011-03
Doc URL	http://hdl.handle.net/2115/48179
Rights	This is a pre-copy-editing, author-produced PDF of an article accepted for publication in Behavioral Ecology following peer review. The definitive publisher-authenticated version Behavioral Ecology (2011) 22 (2): 363-371 is available online at: http://beheco.oxfordjournals.org/content/22/2/363
Type	article (author version)
File Information	BE22-2_363-371.pdf



[Instructions for use](#)

Rethinking birdsong evolution: Meta-analysis of the relationship between song complexity and reproductive success

Masayo Soma 1 & László Zsolt Garamszegi 2*

1. Laboratory for Animal Behavior and Intelligence, Department of Biology, Faculty of Science, Hokkaido University, N10 W8, Kita-ku, Sapporo, Hokkaido, 060-0810, Japan.

E-mail: masayo.soma@sci.hokudai.ac.jp

2. Department of Evolutionary Ecology, Estacion Biologica de Donana-CSIC, c/ Americo Vespucio, s/n 41092, Seville, Spain

E-mail: laszlo.garamszegi@ebd.csic.es, Fax: (+34) 954 621 125, Tel: (+34) 954 232 340

* Author for correspondence

Abstract

The theory of sexual selection predicts a relationship between male sexual traits and reproductive success. This prediction has been tested extensively using the complexity of birdsong as a model for trait elaboration. However, contradictory results have emerged. Some studies have demonstrated that males with large repertoires enjoy a reproductive advantage, while other studies have failed to support this prediction. To make general inferences from this mixed evidence, we quantitatively reviewed the relevant literature using a meta-analytic approach. The mean effect size for the song/mating success association was significant, but the effects were generally weak, affected by publication bias, confounded by uncontrolled variables and differing across the traits examined. Effect sizes were heterogeneous across studies due to species-specific effects, differences in mating systems and song phenotypes. The degree of association between song complexity and reproductive success was independent of the strength of sexual selection, as assessed by the degree of polygyny and extra-pair paternity. Our results highlight the importance of considering various biological factors to understand the role of repertoires in mediating mating success in different species.

Key words: extra-pair paternity, female choice, sexual selection, trait expression, repertoire size

When Darwin developed the theory of sexual selection in *The Descent of Man, and Selection in Relation to Sex* (Darwin 1871), he considered birdsong a spectacular example of sexual selection. He stated that, “The diversity of the sounds, both vocal and instrumental, made by the males of many species during the breeding-season, and the diversity of the means for producing such sounds, are highly remarkable. We thus gain a high idea of their importance for sexual purposes...” (Darwin 1871). Darwin (1871) suggested two important roles for birdsong, during both competition between males, and in female attraction, and associated these processes with song evolution.

Long after Darwin’s work, this issue remains highly topical. Birdsong has become one of the most studied models of sexual selection, giving birth to a plethora of studies examining the relationship between signal elaboration and reproductive success. Most of these studies have focused on song complexity as a potentially important target of female preference, predicting that large repertoire size or song versatility is associated with increased mating success. Accordingly, a number of field and laboratory studies have reported that complex songs may serve as indicators of male quality and thus should translate into mating and ultimately breeding success (Searcy and Andersson 1986; Searcy 1992; MacDougall-Shackleton 1997; Gil and Gahr 2002; Catchpole and Slater 2008). However, our understanding of the role of song complexity in sexual selection remains incomplete due to a high degree of inconsistency in the research results. While some studies have provided evidence for reproductive advantages in males with large repertoires, others fail to support this underlying prediction, and have reported non-significant or even negative associations between song and mating success (Byers and Kroodsma 2009). These conflicting patterns cause difficulty in making generalizations about the function and evolution of song complexity in relation to female preference.

There are at least 5 explanations for the diversity of previous findings, of which 2 are biological and 3 are technical. First, although song complexity is one of the likely candidate targets of sexual selection, this does not necessarily mean that this trait will relate to reproductive success in all songbirds. In some species, song traits other than repertoire size may evolve as indicators of male quality, leaving little, if any, scope for a correlation between song complexity and reproductive success. For example, song dialects, song matching, song performance, loudness, the presence of particular syllables or song consistency may also function on the grounds of sexual selection (Vallet and Kreutzer 1995, Podos et al. 2004,

Byers 2007, Botero et al. 2009), and in such cases, song variety may not be selected for or may even be detrimental for efficient communication (Beecher and Brenowitz 2005). Furthermore, not only song traits but other behavioral traits or plumage characteristics may also replace the signal function of repertoire size. Consequently, several alternative mechanisms may exist by which males can demonstrate their quality, which could explain why associations between song complexity (taken alone) and sexual selection are not universal.

Second, it is possible that song complexity does relate to mating success to some degree, but that certain biological factors mediate differences in the strength of sexual selection pressure on songs, leading to a wide range in variation in song complexity for mating success between species. For example, in species where there is an extensive range of male quality, female preference for signal elaboration and song complexity may be enhanced relative to species where the distribution of males is more homogenous with respect to their viability and song. Comparative studies of morphological or vocal ornamental traits have repeatedly shown that the reproductive biology of species can serve as an important factor accounting for trait elaboration and evolution (Hamilton and Zuk 1982; Read and Weary 1992; Møller and Birkhead 1994; Owens and Hartley 1998; Dunn et al. 2001; Rubenstein and Lovette 2009). If sexual selection differentially favors birdsong in different species, we can predict a greater effect of songs on mating in species with stronger sexual selection pressure caused by polygynous mating systems or frequent extra-pair fertilization.

Third, it may also be that the relationship between song and reproductive success is not causal, and is thus blurred by uncontrolled confounding factors. For example, most correlative studies ignore the possibility that repertoire size and reproductive output are both dependent on male age, territorial quality or another third factor, which could drive an indirect correlation between the focal traits (Searcy and Andersson 1986). Accordingly, it could be predicted that the strength of the song/mating relationship would be weaker when the effect of the confounding factors are taken into account.

Fourth, heterogeneity among previous studies might arise from differences in the experimental setups or the biological meaning of the focal variables used in each study. Some studies relied on experimental approaches while others used correlational methods. Such differences in study design may relate to different biological situations, and provide results corresponding to different levels of statistical power and the degree of control for unwanted

effects. These differences make generalizations between experimental and observational studies difficult. Moreover, the various methods for estimating song complexity (i.e. repertoire size or versatility) and reproductive success (e.g. timing of mating, number of offspring, or harem size) have been used, which may have led to differing conclusions between studies (MacDougall-Shackleton 1997; Byers and Kroodsma 2009). Also, as songs may be directly relevant to social and genetic mating success (Marshall et al. 2007), it might be confusing to make comparisons across these contexts, which occurs if the component of sexual selection the findings refer to is not taken into account. It is possible, even in the same species, that song complexity plays an important role in social mate choice but is not involved in shaping copulatory decisions, or vice-versa.

Finally, publication bias can cause an overestimation of effects, since non-published results are more likely to involve non-significant, weaker effects (Jennions and Møller 2002). Hence, it is possible that the true relationship between song and mating success is weaker than currently believed, if publication bias is accounted for. Byers and Kroodsma (2009) have pointed out the risk of the ‘file-drawer problem’ especially for laboratory studies, but the extent to which publication bias influences the detected relationship between these variables remains unknown.

To address the above-mentioned issues, we sought to review and quantitatively reevaluate previous studies. Meta-analytic approaches typically aim to examine the discrepancy among studies in a comprehensive manner. Meta-analysis could provide a useful technique for estimating the strength of the association between song complexity and reproductive success while correcting for publication bias, and for identifying factors that account for the heterogeneity of current findings (Arnqvist and Wooster 1995). Such a systematic appraisal has been made for visual signaling, revealing that approximately 10% of the variance in male mating success can be attributed to ornamentation (Gontard-Danek and Møller 1999). Only one previous study has assessed the link between birdsong and reproductive success using a meta-analytic approach. However, this study only focused on genetic mating success gained through extra-pair paternity, and showed weak and non-significant overall effects (Garamszegi and Møller 2004).

The aim of the current study was to fill the gaps in our knowledge on the function of song complexity in sexual selection by applying a meta-analytic approach, in which we statistically combined information from past studies on the topic. We estimated the influence

of publication bias, and calculated the cumulative effect size after controlling for this bias. We also conducted heterogeneity tests and meta-regression analyses to examine what biological or methodological factors are responsible for the variance and magnitude of effects. We were particularly interested in identifying species-specific effects, as different species may exhibit different selection mechanisms. Moreover, we predicted that in species showing intense sexual selection (where females are more choosy, for example), the relationship between song and mating success would be stronger than in species with relaxed selection. We also assessed the potentially confounding effects of phylogenetic history in different species, because closely related species are likely to experience similar selection pressure and exhibit similar song phenotypes.

Materials and Methods

Data collection

Using scientific search engines, including Web of Science and Google Scholar, we attempted to locate all publications reporting on the relationship between male song complexity and reproductive success in passerine birds. We performed key word searches using the terms “song complexity”, “song repertoire” or “song versatility”, in combination with the terms “reproductive success” or “mating success”. We also conducted cross-reference searches by consulting previous reviews on birdsong and sexual selection (Searcy and Andersson 1986; Read and Weary 1992; Andersson 1994; Searcy and Yasukawa 1996; MacDougall-Shackleton 1997; Gil and Gahr 2002; Catchpole and Slater 2008; Byers and Kroodsma 2009). In addition, as part of our effort to assess the effect of publication bias (see further details below), we collected unpublished results by contacting birdsong researchers, who were identified by their publications on birdsong.

We used the following criteria for inclusion of studies: (1) results on the relationships between song complexity and reproductive success were based on observational data in wild birds without experimental manipulations; (2) song complexity was measured by either song or syllable repertoire size (i.e. the number of unique song or syllable types in the full repertoire or in a given amount of vocal sample), versatility, or principal component score of multiple song variables that reflect song complexity; and (3) the measure of reproductive success was based on either the number of females mated (including both social and extra-pair mating), the number of offspring sired (including the number of eggs, fledglings

and recruits), or the timing of the onset of reproduction. In this way, we defined song complexity and reproductive success relatively broadly, which was necessary since the available data were limited but showed great diversity in terms of explanatory and response variables. However, in our subsequent analyses, we carefully examined whether the choice of variable in each study affected the strength of the focal relationship. Experimental studies of mate choice in laboratory conditions were not involved in the present study, because they are likely to be more susceptible to publication bias and heterogeneous methodologies, making standardized comparisons difficult (Byers and Kroodsma 2009).

In total, 44 studies met the inclusion criteria, including 39 published and 5 unpublished studies (Supplementary Table S1). Although we cannot rule out the possibility that we missed some relevant studies, our literature sample can be considered representative. The number of collected papers exceeds that of the latest descriptive review by Byers and Kroodsma (2009), which involved 27 field studies. The majority of studies contained multiple results pertaining to the effect of interest. Since multiple effect sizes derived from a single study are likely to be dependent on each other if they relied on the same individuals, we built two datasets as follows: a multiple effect size dataset (A) and a single effect size dataset (B). We calculated effect sizes from each study in all available combinations of the types of song complexity measures (R: repertoire or V: versatility) and the types of reproductive success variables (M: the number of acquired females, N: the number of offspring, or T: the timing of the onset of reproduction). For dataset A, we included all reported effects that were relevant. Thus, this dataset potentially included multiple (maximum of 6, i.e. 2 song variable types \times 3 reproductive success variable types) and thus non-independent entries for single studies. When more than one set of results was reported on the relationship between one song variable type and one reproductive success variable type (e.g. testing male repertoire size in relation to number of eggs, fledglings, and recruits) in a single study, they were combined by calculating the average effect size weighted by the sample size. For dataset B, we took a single effect size from each study. To ensure that dataset B was as uniform as possible in terms of the types of song complexity and reproductive success variables, we selected one effect from each study following the priority criterion of variable types: song complexity variable: R>V; and reproductive success variable: M>N>T. These priority criteria resulted in the most homogeneous dataset with respect to the variable type (i.e. most studies used R and M as focal traits). We created these different datasets because they allowed us to balance between

different constraints. Dataset A could be used to powerfully differentiate between the effects of different variables at the cost of violating statistical independence. On the other hand, dataset B fulfilled assumptions about the independence of data points, but incorporated limited power for identifying variable-specific differences. Therefore, for tests in which we focused on differences between variables we primarily relied on dataset A, while for the rest of the models we used dataset B.

In addition to the variable types, the studies reviewed in our analyses differed in some other respects. Some studies determined reproductive success (e.g. number of females acquired and number of offspring sired) based on social mating, while others determined reproductive success based on genetic mating. Some studies reported simple relationships between song and mating using correlation analyses or t-tests, while others conducted multivariate analyses (e.g. multiple regression, partial correlation tests) and reported statistics after controlling for confounding factors (e.g. territory quality or subject age). These methodological terms were also drawn from each study and we considered them when classifying data. Accordingly, we categorized effect sizes according to whether they referred to social or genetic mating, and whether they originated from bivariate or multivariate analysis.

Variables reflecting the strength of sexual selection

We predicted that biological factors related to sexual selection would be involved in the evolution of songs, and that the intensity of sexual selection would determine how singing behavior correlates with measures of reproductive performance. Hence, we tested whether the strength of sexual selection affected heterogeneity in the observed effect sizes. For these heterogeneity tests, which compare the variance of effect sizes among pre-defined groups, we established the following categories. Species were classified as having either 0: no polygyny, 1: irregular polygyny (species with less than 5% of males attracting more than one female), or 2: regular polygyny (species with more than 5% of males attracting more than one female; Møller 1986; Møller and Birkhead 1994; Hasselquist and Sherman 2001). Estimates of the proportion of extra-pair paternity (EPP) were taken as the percentage of offspring sired by males other than the attending male, in accord with previous studies (Møller and Birkhead 1994; Hasselquist and Sherman 2001; Spottiswoode and Møller 2004). Song and syllable repertoire sizes for each species were also compiled from published studies when available

(Read and Weary 1992; MacDougall-Shackleton 1997; Garamszegi et al. 2007). These continuous estimates of extra-pair paternity and song complexity were categorized into 3-5 levels based on their distribution to obtain similar sample sizes for each category (for details, see footnotes of Table 2).

We also conducted meta-regression analyses to assess the degree to which different selection forces related to sexual selection shaped the association between song complexity and mating success. For these regressions, the predictor traits were used as continuous variables, and we entered the numerical estimates of polygyny, extra-pair paternity and repertoire size into the models. In some species with extremely complex songs, the exact song type repertoire size is impossible to determine and thus is usually described as “large” or “infinite” in review articles (Read and Weary 1992; MacDougall-Shackleton 1997). In these species, non-enumerable song type repertoire size was assumed to be 1,000 in our analyses.

The full dataset is supplied as an online electronic appendix (Supplementary Table S1).

Phylogenetic information

The composite phylogeny for birds used in the phylogenetic meta-analysis was mainly based on Sibley and Ahlquist (1990), combined with information from other sources (Blondel et al. 1996; Leisler et al. 1997; Groth 1998; Grapputo et al. 2001; Yuri and Mindell 2002). We applied branch lengths from Sibley and Ahlquist (1990) for higher taxonomic levels. Within families the distance between different genera was set to $3.4\Delta T_{50H}$ units, and between species within genera to $1\Delta T_{50H}$ units. These branch length conventions are considered standard, and are widely used in the field (Bennett and Owens 2002). We did not use Barker et al.’s (2002) phylogeny because it did not allow the resolution of branch lengths within families.

Meta analysis

Effect sizes of the associations between song complexity and reproductive success were calculated as Pearson product-moment correlation coefficients (r), by converting published statistics if necessary based on standard methods (Cohen 1992; Walker 2003; Nakagawa and Cuthill 2007), and the normalized score of r , Fisher’s z , was used for performing meta-analyses, as described hereinafter. Positive relationships between song and

reproduction were assigned a positive sign (accordingly, the negative relations with T were also interpreted as supporting evidence for the preference for song trait, thus the corresponding effect size was treated as a positive effect). We used random effects models to calculate the mean effect size for each dataset (A or B), which are appropriate for a heterogeneous dataset after checking that heterogeneity is relatively high in the data (DerSimonian and Laird 1986). We applied Begg's method to identify publication bias (Begg and Mazumdar 1994), in which funnel-plot asymmetry was tested using Kendall's rank correlation between effect size and sample size. This method relies on the underlying assumption that studies with small sample size would be more prone to publication bias, while large studies would be likely to be published regardless of significance of the results. Thus, correlation between the effect estimates and their variances reflects a trend towards bigger effect sizes in studies with smaller samples, and is regarded as indicative of publication bias. Begg and Mazumdar (1994) stated that their test was powerful for large meta-analysis with 75 component studies or more, a sample size that we also reached in this study in dataset A. To calculate mean effect sizes while controlling for publication bias, we applied a Trimfill algorithm, where symmetry in the funnel-plot can be adjusted by filling theoretical missing data points arising from publication bias (Duval and Tweedie 2000).

To examine what factors were responsible for the heterogeneity among studies, we conducted heterogeneity tests with the inclusion of covariates as grouping variables, and tested whether effect sizes were heterogeneous among and within particular groups, in a similar way to one-way ANOVA. The following covariates were considered as grouping variables: species, types of studied variables (i.e. song variable type, reproductive success variable type, genetic/social mating), categories of reproductive ecology based on social mating system and EPP rate, and categories based on song type and syllable repertoire sizes.

We also conducted meta-regression analyses to test for the effect of species, social mating system, and EPP rate using dataset B.

To deal with phylogenetic dependence of species, we also conducted phylogenetic meta-analyses (Adams 2008), in which the phylogenetic covariance among taxa was taken fully into account and the phylogenetic cumulative effect size was determined. While conventional meta-analysis cannot account for the non-independence of data due to the phylogenetic relationships of species, it is of great importance to consider such dependence in a meta-analysis of different species, because these units are more likely to share similar

reproductive ecology and show similar song behaviors. Because Adams' model (Adams 2008) accepts single effect size per species, we adopted a combined effect size for each species by calculating average effect size weighed by sample size (i.e. the number of data points in each component study) when multiple studies for a single species were involved in dataset B. Adams' method allow assessing the importance of phylogenetic correction based on an information theoretic approach, in which model fit statistics in the form of Akaike's Information Criterion (AIC) can be used to determine if conventional meta-analysis or the phylogenetic meta-analysis performs better for the data at hand. Although, removing phylogenetic signal from the data may be warranted in theory, such correction is an empirical issue and is necessary only if the available data are phylogenetically structured (see Freckelton et al. 2002). Therefore, we preferred interpreting the model with the lowest AIC value.

All meta-analyses were conducted using R 2.9.0 (R Development Core Team 2009) and its 'meta' and 'metafor' packages.

Results

Main effect sizes

We collected 75 estimates of the associations between song complexity and reproductive success for 27 species from 44 studies, including five that were unpublished (Supplementary Table S1). These 75 estimates were treated independently in dataset A. Mean effect size (r) for this dataset was 0.245 (95% CI range = 0.179 - 0.312), which was statistically significant ($n = 75$, $z = 7.22$, $P < 0.0001$). However, a funnel plot analysis showed evidence of significant publication bias (Kendall's tau = 0.176, $P = 0.025$; Table 1, Fig. 1), suggesting that smaller effect sizes may have remained unreported for some effects, causing missing data to be distributed non-randomly. We also compared average effect sizes between published and unpublished studies, revealing that smaller effect sizes are more likely to remain unreported (unpublished: mean effect size $r = 0.136$, 95 % CI range = -0.071 - 0.343, $n = 8$, $z = 1.28$, $P = 0.199$; published: mean effect size $r = 0.256$, 95 % CI range = 0.186 - 0.326, $n = 67$, $z = 7.14$, $P < 0.0001$). When we controlled for publication bias, the mean effect size decreased to 0.175 (95 % CI range = 0.103 - 0.247), but was still positive and statistically significant ($z = 4.76$, $P < 0.0001$; Fig. 1). To deal with the non-independence from multiple estimates taken from one study, we also analyzed dataset B, which involved a single effect

size per study (Supplementary Table S1). We found similar mean effect size as above ($r = 0.262$, 95 % CI range = 0.161 - 0.364, $n = 44$, $z = 5.07$, $P < 0.0001$), which was again weakened when publication bias was accounted for ($r = 0.200$, 95 % CI range = 0.092 - 0.309, $n = 44$, $z = 3.309$, $P = 0.0003$).

Using dataset A, we compared studies based on the variables and methods used. The mean effect size was smaller when song complexity was measured by versatility ($r = 0.140$, 95 % CI range = -0.038 - 0.310, $n = 11$) relative to repertoire size ($r = 0.261$, 95 % CI range = 0.190 - 0.333, $n = 64$; Table 1, Fig 2). The mean effect size was smaller for genetic mating ($r = 0.044$, 95 % CI range = -0.136 - 0.225, $n = 14$) relative to social mating ($r = 0.274$, 95 % CI range = 0.206 - 0.339, $n = 61$; Table 1, Fig 2). Mean effect sizes also differed depending on how reproductive success was determined (timing: $r = 0.272$, 95 % CI range = 0.146 - 0.390, $n = 24$; number of mates: $r = 0.264$, 95 % CI range = 0.145 - 0.383, $n = 26$; number of offspring: $r = 0.194$, CI = 0.076 - 0.306, $n = 20$; Table 1, Fig. 2).

Mean effect size was also smaller when multivariate analyses were applied in the component studies ($r = 0.166$, 95 % CI range = 0.094 - 0.237, $n = 23$), indicating that associations between song complexity and reproductive success were weaker when confounding variables were taken into account.

In summary, overall mean effect sizes calculated for each dataset and subset were small (between 0.1-0.3) but were generally significantly larger than zero, except when we considered the subsets of studies that focused on genetic mating or song versatility as focal traits (Fig. 2). Effect sizes were systematically smaller when publication bias was controlled for. The strength of the song/reproductive success relationship was dependent on the traits considered, with stronger effects found for repertoire size, number of mates acquired and pairing date (Table 1, Fig. 2). Studies that controlled for potentially confounding variables reported smaller effect sizes than studies reporting pair-wise relationships that disregarded confounders.

Heterogeneity test

Because effect sizes were highly heterogeneous within datasets A and B, and each subset (Table 1), we examined possible biological factors that may have mediated this heterogeneity. Most of the heterogeneity arose between species in both datasets (Table 2). In dataset A, differences in the variable types measured (R vs. V or M vs. N vs. T) did not cause

heterogeneous effect sizes, while the definition of reproductive success (genetic or social) partially explained heterogeneity (Table 2). In dataset B, in addition to species-specific effects, effect sizes also varied differently between groups that were differentiated based on EPP rate and song repertoire size (Table 2). However, none of the covariates considered was fully responsible for the heterogeneity of effect sizes across studies, as in each case some within-group heterogeneity was left to be explained by unknown variables. This indicates that even when species-specific effects, reproductive ecology and song phenotype were held constant, differences between studies still caused considerable variation in the observed relationship between song complexity and reproductive success.

Meta-regression analyses

To investigate how different biological factors shaped the strength of the song/reproductive success association, we conducted meta-regression analyses in which the following factors were included in the models: species, social mating system and EPP rate. In the analysis of the entire dataset B, none of these variables had a significant effect (Table 3).

Phylogenetic meta-analysis

Given the phylogenetically structured data, it might be possible that the association between song complexity and reproductive success is shaped to some extent by the shared evolutionary history of species. The results of a phylogenetic meta-analysis (Adams 2008) showed a higher mean effect size ($r = 0.542$, $n = 27$, $P = 0.0001$) than that obtained from conventional meta-analyses using species as the unit of analysis ($r = 0.202$, $n = 27$, $P = 0.0038$). However, the conventional meta-analysis offered a considerably better model fit than the phylogenetic model (AIC = 25.8 vs. AIC = 71.67). This suggested that phylogenetic structure did not explain the variation in the data well, and that the results without phylogenetic correction were preferable.

Discussion

Following the proposal of Darwin's theory of sexual selection in 1871, it has become widely accepted that song complexity in at least some oscine birds evolved under the pressure of sexual selection, as a consequence of reproductive advantages conferred on males with elaborate songs. Therefore, it is not surprising that we detected a generally positive

association between song complexity and reproductive success when we integrated past field research using meta-analyses. However, the overall mean effect sizes (r) for each dataset and subset varied between 0.1 and 0.3, which can be considered a small effect (*sensu* Cohen 1992). This small effect size contrasts with the intermediate effect size that had been reported for the relationship between visual ornamentation and mating success (Gontard-Danek and Møller 1999). We demonstrated that the combined effect sizes were likely to have been overestimated because of publication bias by showing smaller mean effect size for unpublished studies. We were unlikely to have obtained all unpublished information relevant to our study, but we assume that our success in finding these non-published results was independent of the strength of the effects they covered, and propose that the gathered information thus provides a representative sample of unpublished results. In case this assumption is not supported, conclusions about publication bias should be based on the outcome of the companion analysis, in which we used meta-analytical approaches to assess the degree of the file-drawer problem by relying on a larger sample of published results (Fig. 1). However, it should be noted that the two approaches provided qualitatively similar results. In addition, the mean effect size was smaller in studies that conducted multivariate analyses than in studies that did not control for other potentially confounding factors. This suggests the possibility that the association between song complexity and reproductive success was to some extent mediated by other covariates, such as territorial quality or age, which remain uncontrolled in most previous studies relying on pair-wise correlations. This technical finding does not question the role of song complexity during reproduction, but highlights the importance of paying attention to which aspect of the reproductive process song influences, and which characteristics of individuals are more directly reflected by song quality.

The degree of association observed between song complexity and reproductive success was dependent on how each variable was defined (Table 1, Fig. 2). For example, when song complexity was measured by versatility (Fig. 2b), and when reproductive success was determined based on genetic mating (Fig. 2c), song and reproduction were less correlated. This indicates that repertoire size may be more often sexually selected than song versatility. The effect size for genetic mating success was not only small on average but exhibited a high degree of variance (Fig. 2c), in accord with the findings of a previous meta-analytic appraisal based on a smaller sample size ($r = 0.122$, Cohen's $d = 0.245$ [SE = 0.195]; Garamszegi and Møller 2004). Therefore, the current data suggest that the relationship between mating

outcome and song complexity might be stronger in the social than in the genetic context. The wide confidence intervals and the positive effect size estimate after controlling for publication bias for the relationship between song complexity and extra-pair paternity (Fig. 2c) imply that we still cannot reject the possibility that songs have an effect in the genetic context to some extent. We propose that more studies are needed to draw stronger conclusions about the general relationship between the degree of cuckoldry and repertoire size. Particularly, the effect sizes shown in Fig. 2d indicate that song complexity is likely to mediate mating success via quick mating and the number of mates acquired, while it does not necessarily translate into reproductive success in terms of the number of offspring. Note that the effect size for the number of offspring was a combined effect size and it involved weak effect sizes for the number of eggs and hatchlings and moderate effect sizes for the number of fledglings and recruits (data not shown). Based on the available data, we cannot fully exclude the possibility that there is an effect on the number of surviving offspring or that selection is acting on the quality and not the quantity of offspring.

Our meta-analytic findings suggest that the strength of the observed relationship between song complexity and reproductive success in a given study may depend on several factors. Therefore, it may be expected that there is a high degree of heterogeneity between the outcomes of studies testing the predictions of sexual selection using different focal variables in different contexts, and controlling for confounding variables in a non-standard manner. Supporting this expectation, we found a large amount of heterogeneity among effect sizes (Table 1), which could be partially attributed to the reproductive ecology and song phenotype of the focal species, and to the reproductive variable measured (Table 2). Moreover, effect sizes were highly heterogeneous between subject species (Table 2). This may indicate variation in the effect of sexual selection on song complexity across species. Species of songbirds greatly differ with regard to the function of their songs, and to the selection pressures that shape them during reproduction (Catchpole and Slater 2008). Such interspecific differences can be explained by various ecological and behavioral factors (Read and Weary 1992), and our results suggest that such selection forces can also mediate the strength of the relationship between song and reproduction that is observed at the within-species level. In addition, our meta-analysis indicated that some heterogeneity remained detectable even between different studies of the same species, thus selection factors and methodological differences can vary in a narrow context. In addition, it is also possible that the heterogeneity

is partially due to the existence of generally larger sampling errors in field studies relative to other paradigms. The detected patterns of heterogeneity help us understand why past studies on the topic produced inconsistent results.

In accord with the conclusions of earlier reviews in this field (Read and Weary 1992; MacDougall-Shackleton 1997, Searcy and Nowicki 2005), we propose that the current evidence indicates that a reproductive advantage of males with large repertoire sizes or complex songs is not universal for all oscine species, even if it is empirically well supported in some model species. In some species, song matching or local dialects are more important for increasing reproductive success (Beecher and Brenowitz 2005), while in other species sexual selection acts on consistency or performance (e.g. amplitude) of songs (Podos et al. 2004, Byers 2007, Botero et al. 2009), or on non-acoustic traits (see references in Gontard-Danek and Møller 1999), which can cause weak selection or even trade-offs with song complexity (Cardoso 2010). Heterogeneity in the current findings may also suggest the importance of considering methodological and biological factors that can mediate the predicted relationship and increase variability in the outcomes of different studies.

If sexual selection has favored larger repertoire sizes to enhance female choice, we should predict stronger associations between song complexity and reproductive success in species with high rates of polygyny or EPP. These predictions were not supported by our results, as we found no evidence for the focal association being stronger in species that are more strongly sexually selected as indicated by the higher degree of polygyny or EPP. However, this pattern might be explained if these variables are weak predictors of the strengths of sexual selection on song complexity.

To test this from interspecific comparative aspects, a number of studies have examined the relationship between song complexity and social mating systems (e.g. Catchpole 1980, Read and Weary 1992). However, a remarkable number of studies failed to find a strong link between song repertoire size and mating system at the interspecific level (New World blackbirds: Irwin 1990; wood warblers: Shutler and Weatherhead 1990; passerines: Read and Weary 1992; also see Price and Lanyon 2004). These interspecific comparisons suggest that social mating systems play a mixed role in mediating the evolution of song complexity, and that the effect of polygyny on selection for larger repertoires may vary across bird taxa.

Extra-pair paternity is another key component of sexual selection, and is often

considered to serve as one of the main contributing factors for the evolution of various sexual traits in males (Andersson 1994; Sheldon and Ellegren 1999). The link between extra-pair paternity and trait elaboration is repeatedly corroborated by the interspecific positive relationship between rate of extra-pair paternity and plumage brightness of males (Owens and Hartley 1998). On the contrary, features of song do not seem to strongly associate with the degree of extra-pair paternity either at the interspecific or at the intraspecific level, (Garamszegi and Møller 2004, this study). Only some exceptional evidence indicates that females prefer to copulate with males with more complex songs (e.g. Hasselquist et al. 1996), while one study showed that extra-pair males had smaller song repertoires than social mates (Marshall et al. 2007). Here, we demonstrated that the frequency of extra-pair paternity observed in a species cannot explain the degree of association between birdsong and reproductive success, suggesting that the evolution of the function of large repertoire size generally does not depend on sperm competition.

In conclusion, we demonstrated that the general relationship between song complexity and mating success is small and subject to large variation. Differences in study outcomes may arise because of several biological and methodological factors. Biological factors include the role of traits other than song complexity that also mediate mating success, and variation in the meaning and consequence of repertoire size between species. Technical explanations for the conflicting results include the effects of publication bias, uncontrolled confounders, differences in study design, and variables used to describe the focal traits.

Supplementary material

Supplementary material can be found at <http://www.behco.oxfordjournals.org/>.

Funding

JSPS Grant-in-Aid for Research Activity Start-up (22800002) to M.S.; “Ramon y Cajal” research grant from the Spanish National Research Council (Consejo Superior de Investigaciones Científicas, CSIC) to L.Z.G.

Acknowledgments

We are grateful to B. Ballentine and S. Hamao for providing us with unpublished information.

References

- Adams DC. 2008. Phylogenetic meta-analysis. *Evolution*. 62:567-572.
- Andersson M. 1994. *Sexual Selection*. Princeton: Princeton University Press.
- Arnqvist G, Wooster D. 1995. Meta-analysis: synthesizing research findings in ecology and evolution. *Trends Ecol Evol*. 10:236-240.
- Barker FK, Barrowclough GF, Groth JG. 2002. A phylogenetic hypothesis for passerine birds: taxonomic and biogeographic implications of an analysis of nuclear DNA sequence data. *Proc Roy Soc B*. 269:295-308.
- Beecher MD, Brenowitz EA. 2005. Functional aspects of song learning in songbirds. *Trends Ecol Evol*. 20:143-149.
- Begg CB, Mazumdar M. 1994. Operating characteristics of a rank correlation test for publication bias. *Biometrics*. 50:1088-1101.
- Bennett PM, Owens IPF. 2002. *Evolutionary ecology of birds: life history, mating systems and extinction*. Oxford: Oxford University Press.
- 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.
- Botero CA, Rossman RJ, Caro LM, Stenzler LM, Lovette IJ, de Kort SR, Vehrencamp SL. 2009. Syllable type consistency is related to age, social status and reproductive success in the tropical mockingbird. *Anim Behav*. 77:701-706.
- Byers BE. 2007. Extrapair paternity in chestnut-sided warblers is correlated with consistent vocal performance. *Behav Ecol*. 18:130-136.
- Byers BE, Kroodsma DE. 2009. Female mate choice and songbird song repertoires. *Anim Behav*. 77:13-22.
- Cardoso GC. 2010. Loudness of birdsong is related to the body size, syntax and phonology of passerine species. *J Evol Biol*. 23:212-219.
- Catchpole CK. 1980. Sexual selection and the evolution of complex songs among European warblers of the genus *Acrocephalus*. *Behaviour*. 74:149-165.
- Catchpole CK, Slater PJB. 2008. *Bird song: biological themes and variations* (2nd ed.). Cambridge: Cambridge University Press.
- Cohen J. 1992. A power primer. *Psychol Bull*. 112:155-159.
- Darwin C. 1871. *The descent of man, and selection in relation to sex*. London: John Murray.
- DerSimonian R, Laird N. 1986. Meta-analysis in clinical trials. *Control Clin Trials*.

- 7:177-188.
- Dunn PO, Whittingham LA, Pitcher TE. 2001. Mating systems, sperm competition, and the evolution of sexual dimorphism in birds. *Evolution*. 55:161-175.
- Duval S, Tweedie R. 2000. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. 56:455-463.
- Freckleton RP, Harvey PH, Pagel M. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *The American Naturalist* 160:712-726.
- Garamszegi LZ, Biard C, Eens M, Møller AP, Saino N. 2007. Interspecific variation in egg testosterone levels: implications for the evolution of bird song. *J Evol Biol*. 20:950-964.
- Garamszegi LZ, Møller AP. 2004. Extrapair paternity and the evolution of bird song. *Behav Ecol*. 15:508-519.
- Gil D, Gahr M. 2002. The honesty of bird song: multiple constraints for multiple traits. *Trends Ecol Evol*. 17:133-141.
- Gontard-Danek MC, Møller AP. 1999. The strength of sexual selection: a meta-analysis of bird studies. *Behav Ecol*. 10:476-486.
- Grapputo A, Pilastro A, Baker AJ, Marin G. 2001. Molecular evidence for phylogenetic relationships among buntings and American sparrows (Emberizidae). *J Avian Biol*. 32:95-101.
- Groth JG. 1998. Molecular phylogenetics of finches and sparrows: consequences of character state removal in cytochrome b sequences. *Mol Phylogenet Evol*. 10:377-390.
- Hamilton WD, Zuk M. 1982. Heritable true fitness and bright birds - a role for parasites. *Science*. 218:384-387.
- Hasselquist D, Bensch S, von Schantz T. 1996. Correlation between male song repertoire, extra-pair paternity and offspring survival in the great reed warbler. *Nature*. 381:229-232.
- Hasselquist D, Sherman PW. 2001. Social mating systems and extrapair fertilizations in passerine birds. *Behav Ecol*. 12:457-466.
- Irwin RE. 1990. Directional sexual selection cannot explain variation in song repertoire size in the New-World blackbirds (Icterinae). *Ethology*. 85:212-224.
- Jennions MD, Møller AP. 2002. Publication bias in ecology and evolution: an empirical assessment using the 'trim and fill' method. *Biol Rev*. 77:211-222.

- 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.
- MacDougall-Shackleton SA. 1997. Sexual selection and the evolution of song repertoires. *Curr Ornithol.* 14:81-124.
- Marshall RC, Buchanan KL, Catchpole CK. 2007. Song and female choice for extrapair copulations in the sedge warbler, *Acrocephalus schoenobaenus*. *Anim Behav.* 73:629-635.
- Møller AP. 1986. Mating systems among European passerines: a review. *Ibis.* 128:234-250.
- Møller AP, Birkhead TR. 1994. The evolution of plumage brightness in birds is related to extrapair paternity. *Evolution.* 48:1089-1100.
- Nakagawa S, Cuthill IC. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol Rev.* 82:591-605.
- Owens IPF, Hartley IR. 1998. Sexual dimorphism in birds: why are there so many different forms of dimorphism? *Proc Roy Soc B.* 265:397-407.
- Podos J, Huber SK, Taft B. 2004. Bird song: the interface of evolution and mechanism. *Annu Rev Ecol Evol Syst.* 35:55-87
- Price JJ, Lanyon SM. 2004. Patterns of song evolution and sexual selection in the oropendolas and caciques. *Behav Ecol.* 15:485-497.
- R Development Core Team. 2009. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: <http://www.R-project.org>.
- Read AF, Weary DM. 1992. The evolution of bird song: comparative analyses. *Philos Trans Roy Soc Lond B.* 338:165-187.
- Rubenstein DR, Lovette IJ. 2009. Reproductive skew and selection on female ornamentation in social species. *Nature.* 462:786-789.
- Searcy WA. 1992. Song repertoire and mate choice in birds. *Am Zool.* 32:71-80.
- Searcy WA, Andersson M. 1986. Sexual selection and the evolution of song. *Ann Rev Ecol Syst.* 17:507-533.
- Searcy WA, Yasukawa K. 1996. Song and female choice. In: Kroodsma DE, Miller EH, editors. *Ecology and evolution of acoustic communication in birds*. Ithaca, New York: Cornell University Press. p. 454-473.

- Searcy WA, Nowicki S. 2005. The evolution of animal communication. Princeton: Princeton University Press.
- Sheldon BC, Ellegren H. 1999. Sexual selection resulting from extrapair paternity in collared flycatchers. *Anim Behav.* 57:285-298.
- Shutler D, Weatherhead PJ. 1990. Targets of sexual selection - song and plumage of wood warblers. *Evolution.* 44:1967-1977.
- Sibley CG, Ahlquist JE. 1990. Phylogeny and classification of birds: a study in molecular evolution. New Haven: Yale University Press.
- Spottiswoode C, Møller AP. 2004. Extrapair paternity, migration, and breeding synchrony in birds. *Behav Ecol.* 15:41-57.
- Vallet E, Kreutzer M. 1995. Female canaries are sexually responsive to special song phrases. *Anim Behav.* 49:1603-1610.
- Walker DA. 2003. JMASM9: Converting Kendall's tau for correlational or meta-analytic analyses. *J Mod Appl Stat Meth.* 2:525-530.
- Yuri T, Mindell DP. 2002. Molecular phylogenetic analysis of Fringillidae, "New World nine-primaried oscines" (Aves : Passeriformes). *Mol Phylogenet Evol.* 23:229-243.

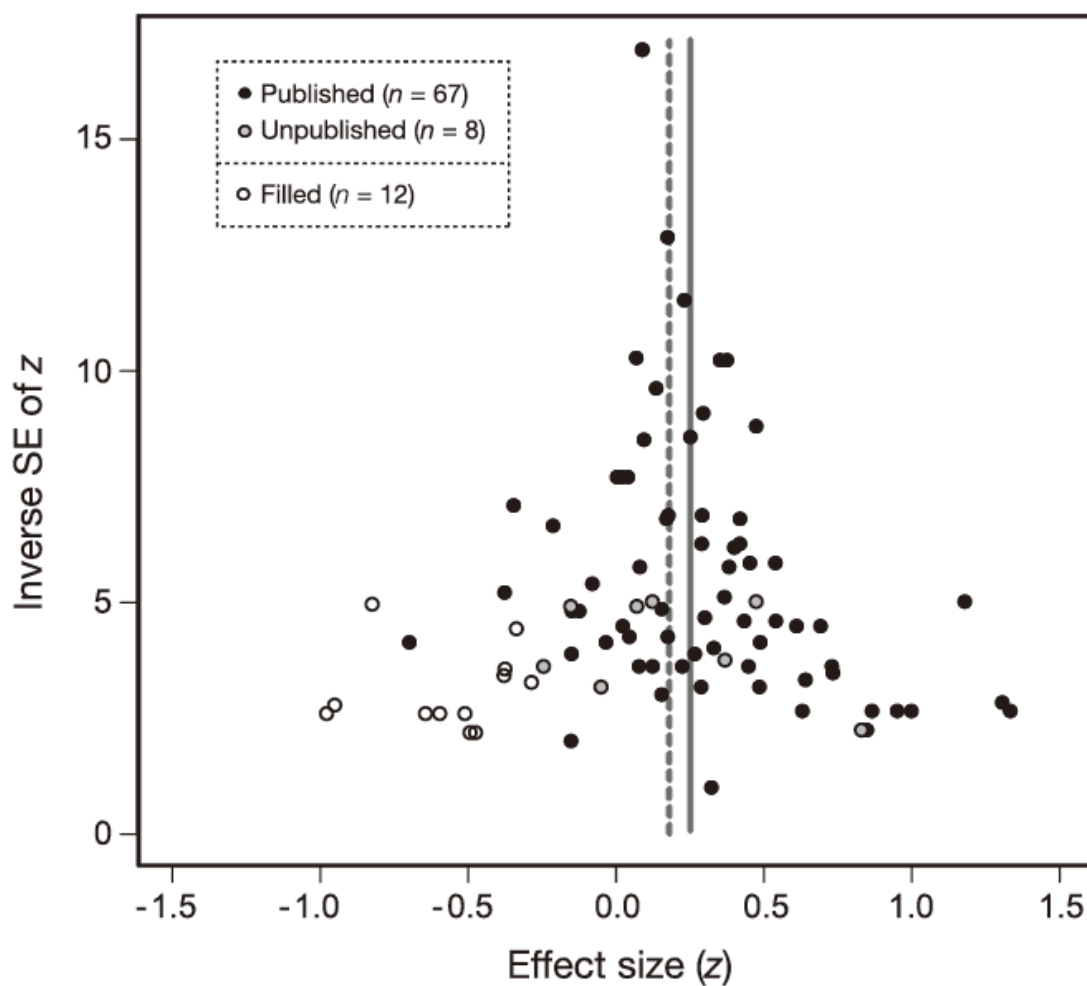


Figure 1. Funnel plot of effect sizes for the relationship between song complexity and reproductive success when using the multiple effect size dataset (dataset A). Black and gray circles show published and unpublished effect sizes, respectively. Solid and dotted lines show mean effect sizes before and after controlling for publication bias, in which theoretical missing data points (open circles, $n = 12$) were added to adjust funnel plot asymmetry.

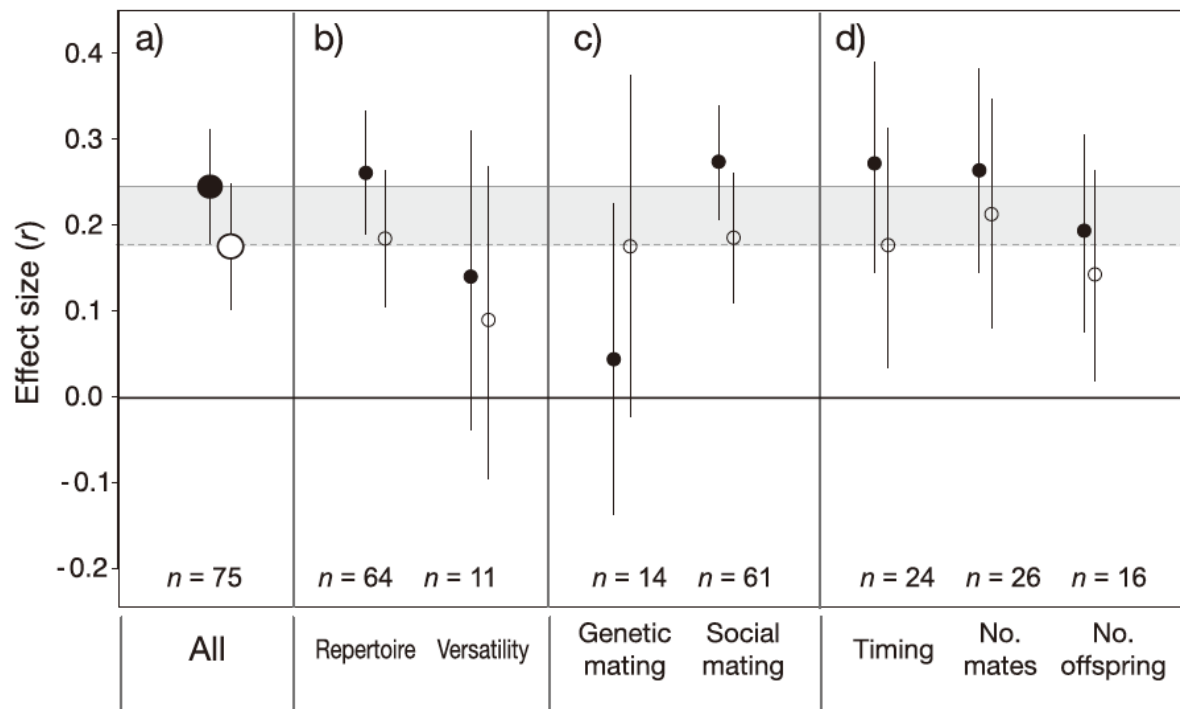


Figure 2. Mean (circles) and 95% CI (error bars) of effect sizes (r) for the song/reproductive success relationship are plotted for dataset A (a) and for subsets classified by the types of song measures (b), definitions of mating success (c), and reproductive success variables (d). Filled and open circles show mean effect sizes before and after controlling for publication bias, and the gray band shows the range of mean effect sizes in dataset A between those controlled (dotted line) and uncontrolled (solid line) for publication bias for reference.

Table 1. Summary of the meta-analysis results.

Data set category	n		Effect size	Mea effect size				Publication bias test Kendall's tau	Heterogeneity test		Mean effect size controlled for publication bias ^a			
	Species	Study		r	95%CI for r	z	P		Q	P	r	95%CI for r	z	P
Overall														
(A) Multiple/study	27	44	75	0.245	0.179–0.312	7.22	<0.0001	0.176*	189.7	<0.0001	0.175	0.105–0.247	4.76	<0.0001
(B) Single/study	27	44	44	0.262	0.161–0.364	5.07	<0.0001	0.118	130.2	<0.0001	0.200	0.092–0.309	3.62	0.0003
Subset														
Song complexity variable was measured by														
Repertoire	24	40	64	0.261	0.190–0.333	7.14	<0.0001	0.184*	168.0	<0.0001	0.184	0.106–0.262	4.64	<0.0001
Versatility	7	8	11	0.140	–0.038–0.310	1.54	0.123	0.220	19.5	0.035	0.091	–0.092–0.267	0.98	0.330
Reproductive success was determined based on														
Genetic mating	10	12	14	0.044	–0.136–0.225	0.48	0.630	<0.001	28.5	0.008	0.175	–0.02–0.371	1.76	0.079
Social mating	23	35	61	0.274	0.206–0.339	7.70	<0.0001	0.267*	156.9	<0.0001	0.185	0.110–0.259	4.74	<0.0001
Reproductive success was measured by														
Timing of reproduction ^b	17	24	24	0.272	0.146–0.390	4.13	<0.0001	0.350*	64.6	<0.0001	0.176	0.035–0.311	2.44	0.015
No. of obtained mates ^b	18	26	26	0.264	0.145–0.383	4.33	<0.0001	0.015	70.4	<0.0001	0.212	0.081–0.344	3.16	0.0016
No. of offspring ^b	16	20	20	0.194	0.076–0.306	3.21	0.0013	0.216	21.7	0.010	0.143	0.020–0.261	2.28	0.023

Mean effect sizes before and after controlling for publication bias and results of the publication bias test and heterogeneity test are shown for data set A and B and subsets.

^a Trimfill algorithm (Duval and Tweedie 2000) was applied for controlling for publication bias.

^b When one study reported multiple effect size for different song complexity measured, the effect size for repertoire size for repertoire size was chosen for the subset analyses. When there were multiple effect sizes based on repertoire size, they were combined by calculating average weighed by sample size.

* $P < 0.05$.

Table 2. Factors affecting heterogeneity among effects.

Factors	Data set	n			Group level	Between-group heterogeneity			Within-group heterogeneity		
		Species	Study	Effect size		Q	df	P	Q	df	P
Species	A	27	44	75		110.87	26	<0.0001	78.83	48	0.0033
	B	27	44	44		90.19	26	<0.0001	39.97	17	0.0013
Measured variable											
Song variable	A	27	44	75	R, V	2.23	1	0.135	187.47	73	<0.0001
Reproductive success definition	A	27	44	75	Genetic, social	4.23	1	0.040	185.47	73	<0.0001
Reproductive success variable	A	27	44	75	M, N, T	3.48	2	0.176	186.23	72	<0.0001
Species' reproductive ecology											
Social mating system	B	26	43	43	0, 1, 2	1.00	2	0.608	106.81	40	<0.0001
EPP rate	B	24	40	40	a	13.66	2	0.001	88.48	37	<0.0001
Species' song phenotype											
Song repertoire size	B	24	41	41	b	13.53	4	0.009	93.42	36	<0.0001
Syllable repertoire size	B	27	44	44	c	1.01	2	0.603	129.15	41	<0.0001

The number of examined studies are not the same across all sets of heterogeneity tests because available information on reproductive. df, degrees of freedom.

^a Different extrapair paternity rate were classified into 3 groups: <10% ($n = 11$), 10–30% ($n = 11$), and >30% ($n = 4$).

^b Different species' song repertoire size were divided into 5 groups: <10 ($n = 12$), 10–19 ($n = 7$), 20–50 ($n = 5$), >50 ($n = 17$).

^c Different species' syllable repertoire size were divided into 3 groups: <5 ($n = 14$), <5–9 ($n = 20$), ≥ 10 ($n = 10$).

Table 3. The results of meta-regression analyses for the dataset B.

Study		n = 40		
Effect size		n = 40		
Mean effect size (<i>r</i>)		0.219		
Factors affecting the magnitude of effects				
	<i>df</i>	Slope	<i>z</i>	<i>P</i>
Intercept		0.256		
Species	22	0.007	0.867	0.386
Soial mating system	2	-0.001	-0.001	0.999
EPP	1	-0.646	-1.699	0.089