Evol Ecol (2012) 26:1253–1274
DOI 10.1007/s10682-012-9555-5

ORIGINAL PAPER

Methodological issues and advances in biological meta-analysis

Shinichi Nakagawa · Eduardo S. A. Santos

Received: 29 November 2011 / Accepted: 25 January 2012 / Published online: 17 February 2012

Abstract Meta-analysis has changed the way researchers conduct literature reviews not only in medical and social sciences but also in biological sciences. Meta-analysis in biological sciences, especially in ecology and evolution (which we refer to as 'biological' meta-analysis) faces somewhat different methodological problems from its counterparts in medical and social sciences, where meta-analytic techniques were originally developed. The main reason for such differences is that biological meta-analysis often integrates complex data composed of multiple strata with, for example, different measurements and a variety of species. Here, we review methodological issues and advancements in biological meta-analysis, focusing on three topics: (1) non-independence arising from multiple effect sizes obtained in single studies and from phylogenetic relatedness, (2) detecting and accounting for heterogeneity, and (3) identifying publication bias and measuring its impact. We show how the marriage between mixed-effects (hierarchical/multilevel) models and phylogenetic comparative methods has resolved most of the issues under discussion. Furthermore, we introduce the concept of across-study and within-study meta-analysis, and propose how the use of within-study meta-analysis can improve many empirical studies typical of ecology and evolution.

Keywords Fixed-effect meta-analysis · Random-effects meta-analysis · Meta-regression · Egger's regression · I^2 · Heterogeneity · Multivariate meta-analysis · Trim and fill method

Introduction

In a similar manner that theoretical developments stir research into new directions, methodological developments open up new avenues of research, bringing fresh insights and findings. There is no doubt that meta-analysis has been one of the most important

S. Nakagawa (✉) · E. S. A. Santos
Department of Zoology, University of Otago, 340 Great King Street, Dunedin 9054, New Zealand
e-mail: shinichi.nakagawa@otago.ac.nz

S. Nakagawa
Max Planck Institute for Ornithology, Seewiesen, Germany

methodological developments in many fields of research including medical, social and biological sciences (Egger et al. 2001; Cooper et al. 2009; Koricheva et al. 2012). Meta-analysis has fundamentally changed how scientists conduct a review of related studies or "*How science takes stock*" (Hunt 1997). Meta-analysis is a quantitative method that combines results from different studies on the same topic in order to draw a general conclusion and to evaluate the consistency among study findings (Hedges and Olkin 1985). In a more statistical sense, meta-analysis combines common effect-size statistics (e.g., Hedges' d or the correlation coefficient, r) extracted from relevant studies by accounting for the sample sizes of the studies (i.e. sampling errors). For readers who are unfamiliar with specific equations (for relevant formulas, see below), it may be conceptually helpful to think of meta-analysis as a special case of weighed regression analysis (see Lipsey and Wilson 2001).

A broader use of the term 'meta-analysis' encompasses any procedure which draws together different sources of data regardless of modelling sampling errors. Therefore, phylogenetic comparative analysis, which synthesizes traits across many species and taxa to examine the evolutionary patterns of traits (Harvey and Pagel 1991; Pagel 1999), is sometimes referred to as meta-analysis. Indeed, as we will see below, phylogenetic comparative analysis incorporating measurement errors on traits (Lynch 1991; Ives et al. 2007; Felsenstein 2008) is statistically identical to meta-analysis that incorporates phylogeny (Adams 2008; Lajeunesse 2009; Hadfield and Nakagawa 2010). In this paper, we use the term 'meta-analysis' in a more specific (traditional) sense, focusing on empirical and methodological studies, which employ (normally distributed) effect-size statistics and their sampling error variance (for exceptions, see Cornwallis et al. 2010; Wehi et al. 2011).

Since meta-analysis has been introduced to the field of ecology and evolution, researchers have identified many methodological problems that are specific to meta-analyses in our field (Arnqvist and Wooster 1995; Gurevitch and Hedges 1999). Such problems seem to arise from the fact that meta-analytic methods were first, and still primarily are, developed in the medical and social sciences (Sutton and Higgins 2008). In these fields, the main research focus is obviously on one species, humans, and methodological homogeneity among studies is often achievable. Conversely, when synthesizing studies addressing the same or similar questions, ecologists and evolutionary biologists commonly deal with multiple species and systems, which require different methodologies.

We review the statistical developments in meta-analysis in the biological sciences, especially in ecology and evolution (referred to as 'biological' meta-analysis, hereafter). We survey the most common effect-size statistics and introduce standard meta-analytical models, highlighting their major shortcomings. Then, we concentrate on three methodological issues: (1) non-independence, (2) heterogeneity, and (3) publication bias. Although these problems are not unique to our field, the multi-species nature of biological meta-analysis adds complexity and a multitude of levels to each of them. Finally, we expand on the currently prevailing across-study meta-analysis by introducing the concept of within-study meta-analysis as a useful tool to evaluate results within a single study.

Effect-size statistics and standard meta-analytic models

Effect-size statistics in ecology and evolution

Three categories (comprising six types in total) of effect-size statistics are commonly used in meta-analysis (Borenstein et al. 2009); these three types are based on (1) two different

means, (2) relationships/correlations (r) between two variables, (3) 2-by-2 contingency tables (i.e. binary data). The first category incorporates two types: (a) standardized mean differences (e.g., d or g) and (b) response ratios (Hedges et al. 1999; for a recent development for the response ratio, see Lajeunesse 2011a). The third category includes three types: (a) odds ratio, (b) relative risk, and (c) risk difference. Mathematical definitions of each type of effect-size statistics are found elsewhere (e.g., Nakagawa and Cuthill 2007; Borenstein et al. 2009; Cooper et al. 2009). These six effect-size statistics have two properties in common, which are required for 'traditional' meta-analysis: (1) they, or their transformations, are normally distributed (e.g., Fisher's z transformation of r, Zr, or the natural logarithm of odds ratio, $\ln(OR)$), and (2) their, or their transformations', sampling variances are estimable from formulas.

In a survey of 100 recent biological meta-analyses we assessed which types of effect size statistics are most widely used in the field of ecology and evolution and ascertained the uses of meta-analytic and related methods (for details of methodological procedures and more results, see "Appendix" and also below). The results on the use of effect-size statistics in the 100 studies are summarised in Fig. 1a, b. Standardized mean difference was the most popular category of effect-size statistics reported (61.5%; e.g., Hedges' d and the response ratio). In fact, the response ratio was the most commonly used effect-size statistic overall (32.7%), although the use of this effect size statistic is almost exclusive to the field of ecology and evolution (Borenstein et al. 2009). The nature of meta-analysis usually dictates which effect-size statistic is used (Nakagawa and Cuthill 2007; Borenstein et al. 2009; Cooper et al. 2009), but this is not always the case nor straightforward (e.g., Osenberg et al. 1999; Lajeunesse and Forbes 2003).

Fixed-effect and random-effects models

There are two types of standard meta-analytic models: (1) fixed-effect models, and (2) random-effects models (Hedges and Olkin 1985; Hedges and Vevea 1998; see below for mixed-effects meta-analysis). A fixed-effect meta-analysis can be written as:

$$z_j = \mu + m_j, \tag{1}$$

$$\mathbf{m} \sim N(\mathbf{0}, \mathbf{M}), \tag{2}$$

where z_j is an effect size for the jth study ($j = 1,\ldots, N_{study}$; N_{study} is the number of studies), μ is the meta-analytic mean (or intercept), m_j is a sampling (measurement) error effect for the jth study, $\mathbf{m}$ is a 1 by N_{study} vector of m_j, which is normally distributed around 0 with the corresponding sampling error variance, σ_j^2 (easily estimable for the common effect-size statistics; Nakagawa and Cuthill 2007; Borenstein et al. 2009; Cooper et al. 2009), $\mathbf{0}$ is a 1 by N_{study} vector of 0, and $\mathbf{M}$ is a N_{study} by N_{study} matrix with its diagonal elements being σ_j^2 (which is assumed to be known). Fixed-effect models assume that studies included in a meta-analysis share a common effect size, or meta-analytic mean (irrespective of possible differences among studies). However, such an assumption is hardly ever met, especially in biological meta-analysis where studies include different designs, populations and species (Gurevitch and Hedges 1999; see also Higgins et al. 2009). A random-effects meta-analysis uses the more reasonable assumption that each study has a 'true' effect size different from each other. It can be written as:

$$z_j = \mu + u_j + m_j, \tag{3}$$

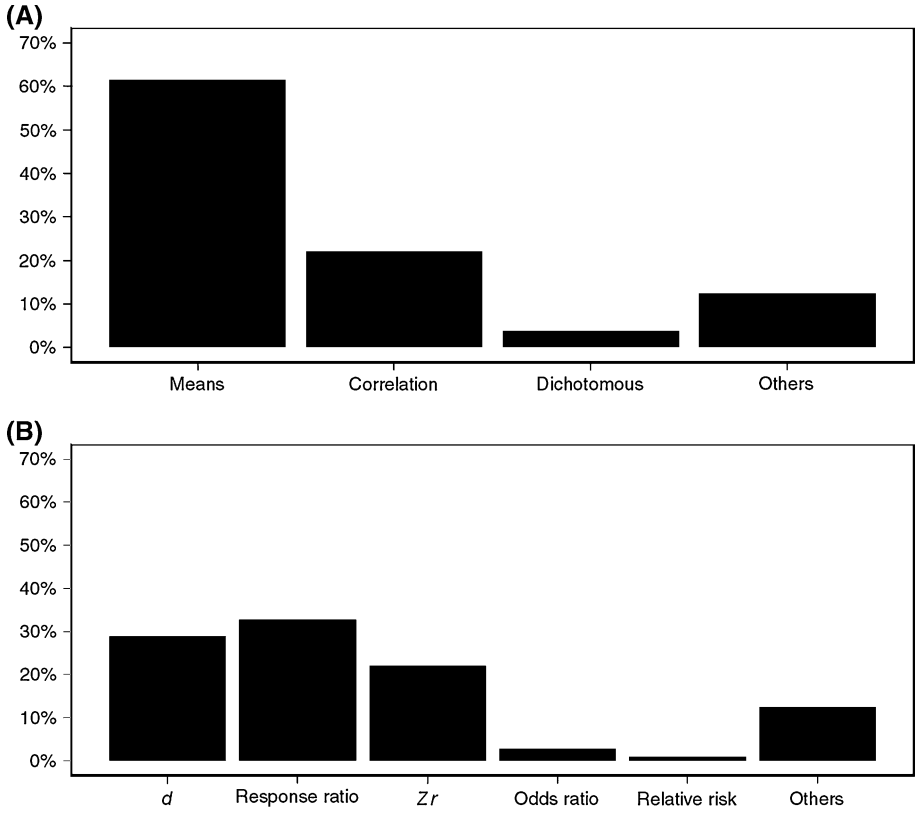

Fig. 1 *Bar*-plots: **a** the frequency with which the three categories of effect sizes (i.e., means, correlation coefficient, and dichotomous) occur in ecology and evolution meta-analyses published between 2009 and 2011 (a total of 104 records from 100 studies, as some studies used multiple effect-size statistics). We pooled all other effect sizes under the category 'Others', and **b** the frequency of each type of effect size from the three main categories (see the text for details)

$$\mathbf{u} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}), \tag{4}$$

where u_j is the study specific effect of the jth study, $\mathbf{u}$ is a 1 by N_{study} vector of u_j, which is normally distributed around 0 with the between-study variance of σ_u^2 (estimated from the data) and $\mathbf{I}$ is a N_{study} by N_{study} identity matrix; the other symbols are as in the fixed-effect model (Eqs. 1, 2).

Although the random-effects meta-analytic model described above is useful and widely used, there is one serious limitation. This model is designed for meta-analyses where the number of effect sizes equals the number of studies ($N_{effect\text{-}size} = N_{study}$). This is often the case for medical meta-analyses as each study provides one effect-size estimate (Egger et al. 2001). However, biological meta-analyses will frequently include studies that often provide more than one effect-size estimate. Unfortunately, many currently available easy-to-use meta-analytic programmes such as *MetaWin* (Rosenberg et al. 2000; for a list of software, see Borenstein et al. 2009; see also "Appendix") cannot appropriately deal with multiple effect size estimates from the same study. We will now focus on recent methodological developments that have overcome this and other non-independence problems, characteristic of biological meta-analyses.

Multi-level non-independence

Related effect sizes

We could categorise the non-independence problem in biological meta-analysis into two types: (1) non-independence of observations from related sources, and (2) non-independence from phylogenetic relatedness of species when multi-species are used. Although the distinction between these two categories is sometimes ambiguous in practice (e.g., effect sizes resulting from the same species can belong to both types), we use this categorisation for explanatory purposes. The first type of non-independence appears in many forms (see Gurevitch and Hedges 1999). For example, effect-size estimates originate from the same study groups, populations or species. Also, one study may measure multiple traits as proxies for, say, fitness (e.g., mating success, breeding success and survival). In the past, researchers often acknowledged this type of non-independence, but they, nonetheless, carried out meta-analysis with non-independent datasets using the standard fixed- or random-effects models (reviewed in Gurevitch et al. 2001; Gates 2002). They usually supplemented their study with alternative meta-analyses, which avoided the problem of non-independence. Some meta-analyses opted to include only one effect-size estimate per study, losing statistical power and potentially forgoing important information. Others calculated weighted study and species means, losing the information on within-study and within-species variance. Also, the latter approach of calculating study and species means could potentially inflate Type I error if these means were not calculated taking non-independence into account (for how to avoid this problem, see Davidson et al. 2011; Slatyer et al. 2011; also see below).

The first type of non-independence arises due to strata or multi-levels in data. Therefore, to address this issue most effectively, we should use multilevel/hierarchical models, which explicitly model correlations within the different levels (for an example of a Bayesian hierarchical meta-analysis, see Liermann and Hilborn 1997; for general references for multilevel/hierarchical models, see Gelman and Hill 2007; note that these models are usually called mixed-effects models in ecology and evolution, Pinheiro and Bates 2000; Bolker et al. 2009, but that we distinguish them from mixed-effects meta-analysis, see below). The simplest multilevel meta-analytic model can be written as (following notations by Gelman and Hill 2007):

$$z_i = \mu + u_{j[i]} + m_i, \tag{5}$$

where $u_{j[i]}$ denotes the study-specific effect for the jth study ($j = 1,\ldots, N_{study}$) applied to the ith effect size ($i = 1,\ldots, N_{effect\text{-}size}$; $N_{effect\text{-}size}$ is the number of effect-size estimates; note that $N_{effect\text{-}size} > N_{study}$); all symbols are comparable to the random-effects model above. This multilevel meta-analytic model can be implemented in, for example, the *metafor* package in *R* (Viechtbauer 2010) or using the *nlme4* package in *S-Plus* (Pinheiro and Bates 2000). Indeed, recent biological meta-analyses have taken advantage of versions of this multilevel meta-analysis (e.g., Evans et al. 2010; Weir et al. 2011; Hector and Nakagawa 2012). The shortcoming of this model is the assumption that all within-study variance comes from sampling variance (Eq. 2) when, in fact, within-study variance can be separated and estimated when studies have multiple effect sizes (Hadfield and Nakagawa 2010). Therefore, a more reasonable multilevel model is:

$$z_i = \mu + u_{j[i]} + e_i + m_i, \tag{6}$$

$$\mathbf{e} \sim N(\mathbf{0}, \sigma_e^2 \mathbf{I}), \tag{7}$$

where e_i is the effect-size-specific (within-study) effect (this is the residual term in a normal linear model) for the ith effect size, and $\mathbf{e}$ is a 1 by N_{study} vector of e_j, which is normally distributed around 0 with within-study variance, σ_e^2 (which is estimated from the data); the other symbols are as above. We can implement this meta-analytic model using *ASReml* (Gilmour et al. 2002), *BUGS* (Lunn et al. 2000), and the *R* package *MCMCglmm* (Hadfield 2010). Of course, this multilevel meta-analytic model (Eq. 6) can be extended to include more strata of data; for example, a more complex meta-analytic model including species-level effects can be written as:

$$z_i = \mu + v_{k[i]} + u_{j[i]} + e_i + m_i, \tag{8}$$

$$\mathbf{v} \sim N(\mathbf{0}, \sigma_v^2 \mathbf{I}), \tag{9}$$

where $v_{k[i]}$ is the species-specific effect for the kth species ($k = 1,\ldots, N_{species}$; $N_{species}$ is the number of species; note that usually $N_{effect\text{-}size} > N_{study} > N_{species}$) applied to the ith effect-size estimate, and $\mathbf{v}$ is a 1 by $N_{species}$ vector of v_j, which is normally distributed around 0 with species-specific variance σ_v^2 (which is estimated from the data); the other symbols are as above. Yet, even this model (Eq. 8) falls short of adequately treating the non-independence at the species level because it does not explicitly model the different degrees of relatedness among species.

Phylogenetic relatedness

The second type of non-independence, phylogenetic relatedness, can be dealt with by using the framework of phylogenetic comparative methods, which have been extensively developed in evolutionary biology (Harvey and Pagel 1991; Pagel 1999). A series of papers have demonstrated the importance of controlling phylogeny for comparing phenotypic traits across species (reviewed in Freckleton et al. 2002; Blomberg et al. 2003). The earlier attempts at controlling for phylogeny included running both standard meta-analysis and comparative analysis on the same datasets (see Dubois and Cezilly 2002; Verdu and Traveset 2004; see also Verdu and Traveset 2005), because software packages to conduct 'phylogenetic' meta-analysis (sensu Adams 2008) were not available. Adams (2008) first described how the fixed-effect meta-analytic model could incorporate phylogenetic relatedness. Lajeunesse (2009), then, extended phylogenetic meta-analysis to random-effects models. Their phylogenetic meta-analytic models can be written as:

$$z_k = \mu + a_k + m_k, \tag{10}$$

$$z_k = \mu + a_k + u_k + m_k, \tag{11}$$

$$\mathbf{a} \sim N(\mathbf{0}, \sigma_a^2 \mathbf{A}), \tag{12}$$

where a_k is the phylogenetic effect for the kth species (in these models, $N_{effect\text{-}size} = N_{study} = N_{species}$) and $\mathbf{a}$ is a 1 by $N_{species}$ vector of a_k, which is multivariate-normally distributed around 0, σ_a^2 is phylogenetic variance, $\mathbf{A}$ is a $N_{species}$ by $N_{species}$ correlation matrix of distances between species, extracted from a phylogenetic tree (see below), and the other notations are as above. Note that Eq. 10 is a fixed-effect meta-analysis (cf. Eq. 1) while Eq. 11 is a random-effects meta-analysis (cf. Eq. 3). For example, when $N_{species} = 3$, $\mathbf{A}$ can be written as:

$$\mathbf{A} = \begin{bmatrix} 1 & d_{12} & d_{13} \\ d_{21} & 1 & d_{23} \\ d_{31} & d_{32} & 1 \end{bmatrix}, \tag{13}$$

$$\mathbf{A} = \begin{bmatrix} 1 & \exp(-\alpha d_{12}) & \exp(-\alpha d_{13}) \\ \exp(-\alpha d_{21}) & 1 & \exp(-\alpha d_{23}) \\ \exp(-\alpha d_{31}) & \exp(-\alpha d_{32}) & 1 \end{bmatrix}, \tag{14}$$

where d is the distance from the root to the most recent common ancestor between two species (also note that the distance between every tip in the phylogenetic tree needs to have a unit length, i.e. an ultrametric tree) and α is a constant which represents the speed of divergence between species (set to a specific value or estimated from data). Equation 13 assumes a Brownian motion model of evolution (Felsenstein 1985) whereas Eq. 14 assumes the Ornstein–Uhlenbeck model of evolution (Martins and Hansen 1997; both types of models for meta-analysis were described by Lajeunesse 2009). Lajeunesse (2011b) implemented his methods in a stand-alone programme, named *phyloMeta*, which has been used in several meta-analytic studies (e.g., Carmona et al. 2011; Meunier et al. 2011).

An obvious limitation of these approaches (Eqs. 10, 11) is that effect-size estimates have to be collapsed to the species level because of the restriction $N_{effect\text{-}size} = N_{study} = N_{species}$ (Adams 2008; Lajeunesse 2009, 2011b). A solution to this limitation is to combine multilevel and phylogenetic approaches, as described by Hadfield and Nakagawa (2010). A phylogenetic multilevel meta-analysis can be written as:

$$z_i = \mu + a_{k[i]} + s_{k[i]} + u_{j[i]} + e_i + m_i \tag{15}$$

$$s \sim N(\mathbf{0}, \sigma_s^2 \mathbf{I}), \tag{16}$$

where s_k is the species-specific effect, which is not part of the phylogenetic effect (i.e. a_k), for the kth species, $\mathbf{s}$ is a 1 by $N_{species}$ vector of s_k, which is normally distributed around 0 with species-specific variance σ_s^2 and the other notations are as above. It is important to distinguish s_k from v_k in Eq. 8. Although both are species-specific effects, the latter includes both phylogenetic and non-phylogenetic parts of the effects. Versions of this model can be run in the aforementioned programmes such as *ASReml*, *BUGS* or *MCMCglmm* (relevant code provided in Hadfield and Nakagawa 2010). Several meta-analytic studies have already made use of this more flexible approach (e.g., Cornwallis et al. 2010; Horváthová et al. 2012). Incidentally, this phylogenetic multilevel meta-analysis can be seen as a phylogenetic comparative analysis with z being the trait of interest, u the between-population effect and m the measurement error variance obtained from multiple samples (see Lynch 1991; Ives et al. 2007; Felsenstein 2008).

Even with the development of phylogenetically controlled meta-analytic models, some researchers opted for not using these models (e.g., Kelly and Jennions 2011; Weir et al. 2011). One reason for avoiding the use of the phylogenetically controlled meta-analytic models is that often the collection of species in the data set spans several taxonomic classes so that it is difficult to construct or obtain reliable phylogenetic trees that extend to the species level. To address this issue, Hadfield and Nakagawa (2010) combined a traditional taxonomic model (Harvey and Pagel 1991) and phylogenetic models for meta-analysis. An example of this model when only a family-level phylogenetic tree is available can be written as:

$$z_i = \mu + a_{l[i]} + f_{l[i]} + s_{k[i]} + u_{j[i]} + e_i + m_i, \tag{17}$$

$$\mathbf{f} \sim N(\mathbf{0}, \sigma_f^2 \mathbf{I}), \tag{18}$$

where $f_{l[i]}$ is the family-specific effect, which is not part of phylogenetic effect (i.e. a_l), for the lth family ($l = 1,\ldots, N_{family}$; N_{family} is the number of taxonomic families; note usually that $N_{effect\text{-}size} > N_{study} > N_{species} > N_{family}$) applied to the ith effect-size estimate, and $\mathbf{f}$ is a 1 by N_{family} vector of f_l, which is normally distributed around 0 with family-specific variance σ_f^2 (estimated from data); the other symbols are as above.

It seems that we now have statistical models which can appropriately deal with non-independence faced by biological meta-analyses. However, to date phylogenetically controlled meta-analyses are only rarely performed in ecology and evolution (17%, see "Appendix"). Above, we showed appropriate but increasingly complex models for biological meta-analysis. An important thing to note here is that complex models require more data and meta-analysts need to attain balance between an ideal model and a statistically feasible model given the data at hand.

Heterogeneity

Quantifying heterogeneity

Next to finding a general trend or calculating meta-analytic means, the most important function of meta-analysis is to investigate inconsistency (or consistency) across studies, or quantifying heterogeneity in the data. The reliability of a general trend (meta-analytic mean) depends on the degree of consistency among studies (heterogeneity). Currently, the two test statistics most commonly used to quantify heterogeneity are Cochran' Q (Hedges and Olkin 1985) and I^2 (Higgins and Thompson 2002; Higgins et al. 2003; for an example in our field, see Smith and Blumstein 2008). All major meta-analytic programmes provide Q, which probably explains its widespread application in traditional meta-analyses. More recently, I^2 is increasingly being used and newer computer programmes such as *metafor* in *R* (Viechtbauer 2010) provide I^2 as well as Q. Let us now look at the formulations for both of these statistics. By using the notations of Eqs. 1–4, Cochran's Q can be written as:

$$Q = \sum w_j (z_j - \mu)^2, \tag{19}$$

$$w_j = 1/\sigma_j^2. \tag{20}$$

For statistical significance, Q is tested against a χ^2 distribution with $(k - 1)$ degrees of freedom ($k = N_{study}$). Similarly, I^2 can be expressed as:

$$I^2 = \sigma_u^2/(\sigma_u^2 + \sigma_m^2), \tag{21}$$

$$\sigma_m^2 = \sum w_j (k-1) / \left[\left(\sum w_j \right)^2 - \sum w_j^2 \right], \tag{22}$$

where σ_m^2 is the within-study variance which is estimated from the study-specific sampling/measurement-error variance σ_j^2 (cf. σ_e^2; 7); the other notations are as above (Eqs. 1–4, 19, 20). Higgins and Thompson (2002) refer to σ_m^2 as the 'typical' within-study variance; here we call it the typical sampling-error variance. The reliability and significance of Q is largely dependent on the number of studies (N_{study}) included in a meta-analysis. The statistic I^2 shifts the focus from statistical significance to actual variance and is therefore an improvement over Q as a meta-analytic heterogeneity measurement. Indeed, the statistic I^2

is somewhat similar to effect-size statistics such as d and r; $I^2 = 25$, 50 and 75% are considered as low, moderate and high heterogeneity, respectively (Higgins et al. 2003). As can be seen from their formulations, both Q and I^2 are designed only for quantifying the degree of between-study variance (σ_u^2). Therefore, they are unsuitable for multilevel meta-analysis (e.g., Eqs. 6, 8, 15 and 17). They cannot be used, because multilevel meta-analytic models will include, for example, species-specific and phylogenetic effects (i.e. Eq. 14). Here, we propose a simple solution to this problem by extending the idea on which the formulation of I^2 is based:

$$\sigma_t^2 = \sigma_a^2 + \sigma_s^2 + \sigma_u^2 + \sigma_e^2 + \sigma_m^2, \tag{23}$$

where all the symbols are as above. Then, for example, the amount of heterogeneity at the species level and study level is written as:

$$I_s^2 = \sigma_s^2/\sigma_t^2, \tag{24}$$

$$I_u^2 = \sigma_u^2/\sigma_t^2, \tag{25}$$

where I_s^2 and I_u^2 are the study-level and species-level heterogeneities, respectively. Incidentally, by using the notations above, H^2, a metric of phylogenetic signal named phylogenetic heritability by Lynch (1991) can be written as:

$$H^2 = \sigma_a^2/(\sigma_a^2 + \sigma_s^2 + \sigma_u^2 + \sigma_e^2). \tag{26}$$

When $H^2 = 0$, there is no phylogenetic relatedness among effect sizes (or traits of interest), while $H^2 = 1$ indicates that the effect-size/trait values among species are exactly proportional to their phylogenetic relatedness. It is worth mentioning that H^2 is equivalent to a more-widely used metric, Pagel's λ (1999). This new approach to quantify heterogeneity (or phylogenetic signal) at different strata has been applied in recent studies based on multilevel meta-analytic models (e.g. Sutton et al. 2011; Horváthová et al. 2012).

Explaining heterogeneity

Once heterogeneity is identified, the next step in meta-analysis is to incorporate moderators (referred to as explanatory or independent variables in the linear model framework), which may explain the observed heterogeneity. In other words, we move onto meta-regression (also called mixed-effects meta-analysis; see Hedges and Vevea 1998; Cheung 2008). A simple meta-regression building upon Eq. 3 (a random-effects model) can be written as:

$$z_j = \eta_j + u_j + m_j, \tag{27}$$

$$\eta_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \beta_3 x_{3j} \ldots, \tag{28}$$

where β_0 is the intercept, $\beta_{1\text{-}3}$ are slopes (regression coefficients), $x_{1\text{–}3}$ are moderators, and the other notations are as above (note $j = 1,\ldots, N_{study}$). Note that moderators ($x_{1\text{–}3}$) can be categorical variables with more than two levels; how categorical variables can be incorporated into the regression via dummy variables is well described elsewhere (we refer the reader to Schielzeth 2010). An example of a phylogenetic multilevel meta-regression building upon Eq. 15 can be written as:

$$z_i = \eta_i + a_{k[i]} + s_{k[i]} + u_{j[i]} + e_i + m_i, \tag{29}$$

$$\eta_i = \beta_0 + \beta_1 x_{1k[i]} + \beta_2 x_{2j[i]} + \beta_3 x_{3i} \ldots, \tag{30}$$

where x_1 is a species-level moderator, x_2 is a study-level moderator, x_3 is an effect-size-level moderator and the other notations are as above ($i = 1,\ldots, N_{effect\text{-}size}$; $j = 1,\ldots, N_{study}$; $k = 1,\ldots, N_{species}$). This ability to flexibly incorporate different levels of moderators highlights the advantages of multilevel meta-analytic (meta-regression) models. The collapsing of effect-size estimates to higher levels performed in previous studies causes not only the information regarding lower levels to be hidden in the response (z), but also potentially important moderators (x) to be lost. Because heterogeneity is almost always expected, we suggest that meta-regression should be the default meta-analytic model for biological meta-analysis. The aforementioned resemblance between meta-analytic models and normal linear regression models is especially apparent in meta-regression. Therefore, it is not surprising that one can use information criteria such as Akaike information criterion (AIC; e.g., Jones et al. 2009; Knowles et al. 2009; Weir et al. 2011) or deviance information criterion (DIC; e.g., Horváthová et al. 2012) to select 'better' fitting meta-regression models (for model selection, see Burnham and Anderson 2002; Grueber et al. 2011).

Before leaving this section, we introduce multivariate (i.e. multi-response) meta-analysis/meta-regression by building upon the meta-analytic models described above. Ecologists and evolutionary biologists have started to appreciate that relationships between two groups of effect-sizes (or more than two) can provide valuable biological insights. The meta-analytic work by Griffin and West (2003) who investigated the generality of Hamilton's rule (Hamilton 1964) across cooperatively breeding vertebrates is a case in point. They demonstrated a strong 'correlation' between two kinds of effect-size estimates (based on Zr, Fisher's transformation of r): 1) the correlation between the level of helping and relatedness (i.e. kin discrimination), and 2) the correlation between the level of helping and an increase in indirect fitness (i.e. the benefits of helping). Similarly, Griffin et al. (2005) showed a clear correlation between sex-ratio adjustments (r between the sex-ratios of offspring and the number of helpers) and the benefits of helping (as above) in cooperative breeding vertebrates. However, these studies did not incorporate sampling-error variance when they correlated two types of effect sizes (also see Cornwallis et al. 2010). An appropriate treatment of such data sets is to use bivariate meta-analytic models. An example of multivariate (phylogenetic) meta-analytic models building up Eqs. 10 and 11 can be written as:

$$z_{kh} = \mu_h + a_{kh} + u_{kh} + m_{kh}, \tag{31}$$

$$\mathbf{a}_h \sim N(\mathbf{0}, \sigma^2_{ah}\mathbf{A}), \tag{32}$$

$$\mathbf{u}_h \sim N(\mathbf{0}, \sigma^2_{uh}\mathbf{\Sigma}), \tag{33}$$

$$\mathbf{m}_h \sim N(\mathbf{0}, \mathbf{M}_h), \tag{34}$$

where z_{kh} is the kth effect-size estimate of hth type ($k = 1,\ldots, N_{species}$; $i = 1,\ldots, N_{type}$; N_{type} is the number of types of effect sizes) and therefore, $\mathbf{a}_1$ is the 1 by $N_{species}$ vector of a_{k1} multivariate-normally distributed around 0 with the variance of $\sigma^2_{a1}\mathbf{A}$ (a phylogenetic correlation matrix), and $\mathbf{\Sigma}$ is a $N_{species}$ by $N_{species}$ correlation matrix (estimated from the data); the other notations are comparable to corresponding notations above. When we have a bivariate meta-analytic model (i.e. $N_{type} = 2$), $\mathbf{\Sigma}$ can be written as:

$$\Sigma = \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix}. \tag{35}$$

This correlation ($r_{12} = r_{21}$) provides an appropriate measurement of the relationship between two types of effect sizes, controlling for sampling variance. Of course, the multivariate meta-analytic model (Eqs. 31–34) can be extended to more complex meta-analytic/regression, for example, by incorporating moderators and lower level data (e.g. the study or effect-size level). However, when a pair (or a set) of effect sizes is required for multivariate meta-analysis, data are probably only available at higher levels (e.g., the species level; see Griffin and West 2003; Griffin et al. 2005).

A major difficulty of the multivariate approach may be incompleteness of such data; this is probably the main reason why multivariate meta-analysis in the field is still very rare (but see Sutton et al. 2011). In a case where we are interested in two related effect-size groups, we probably find that some studies provide the first type, some provide the second type, and yet others provide both types. This type of missing data in multivariate meta-analysis is easily handled by data imputation or augmentations (see Nakagawa and Freckleton 2008; Hadfield and Nakagawa 2010) in programs such as *BUGS*, *ASReml* or *MCMCglmm*. For example, Cleasby and Nakagawa (2012) recently used *BUGS* to implement a multivariate response meta-analysis in which a correlation between two sets of effect sizes was quantified: (1) the relationship between male age and extra-pair paternity, and (2) the relationship between male age and within-pair paternity. In this study, Bayesian data augmentation was used to handle such missing data. More general issues regarding missing data in meta-analysis is addressed elsewhere (Pigott 2009). In addition to the aforementioned programs, the recent *R* package *mvmeta* can implement multivariate meta-analysis (Gasparrini 2011; for a recent discussion on multivariate meta-analysis in medical sciences, see Jackson et al. 2011).

In this section, we introduced more complex meta-analytic models, i.e., meta-regression to account for heterogeneity, along with multivariate-response meta-analysis. Again, we remind the reader of the fact that the more complex the model, the more data it requires. Therefore, it is important to use a meta-analytic model, which attains a reasonable compromise in terms of complexity given the number of data points and the nature of these data (e.g., meta-analytic data are often unbalanced and nested). We certainly hope that we did not give the reader an impression that more complex models are always better solutions. It is advisable to run several alternative models to confirm the robustness of one's meta-analytic results.

Publication bias

Publication bias affects any sort of literature synthesis, including meta-analysis because studies with statistically significant results are more likely to be published than otherwise (Rosenthal 1979). This very issue was memorably termed as the 'file drawer problem' by Rosenthal (1979); non-significant results are stored in file drawers without ever being published. Here, we use publication bias in this narrow sense, although publication bias can encompass more than bias in the literature, arising from preferential publication of statistically significant results (for broader definitions, see Møller and Jennions 2001; Lortie et al. 2007; Jennions et al. 2012; for a book dedicated to the relevant topic, see Rothstein et al. 2005). The ability to detect publication bias in a given field is a key strength of meta-analysis, because identification of publication bias will challenge the validity of common views in that area, and will spur further investigations. Other types of synthesis such as traditional qualitative reviews and vote counting clearly lack such ability. There are two types of statistical procedures for dealing with publication bias in meta-analysis: (1)

methods for identifying the existence of publication bias, and (2) methods for assessing the impact of publications bias (Sutton 2009). The former includes: (a) the funnel plot (and other visualisation methods such as the normal quantile plot), and (b) regression/correlation-based tests; while the latter includes: (i) the fail-safe (file-drawer) number (*N*), (ii) the trim and fill method, and (iii) selection model approaches (Sutton 2009).

Our survey of 100 biological meta-analytic studies (see above and also "Appendix") showed that only half of the studies (49%) employed some kind of procedure to deal with publication bias (Fig. 2a); this number is alarmingly low and methods for the identification and assessments of publication bias in biological meta-analyses should be mandatory. Of the studies that considered publication bias, only 45% used procedures to both identify and assess the magnitude of publication bias, while 41.1 and 13.9% of studies only used procedures either to identify or to assess the impact of publication bias, respectively

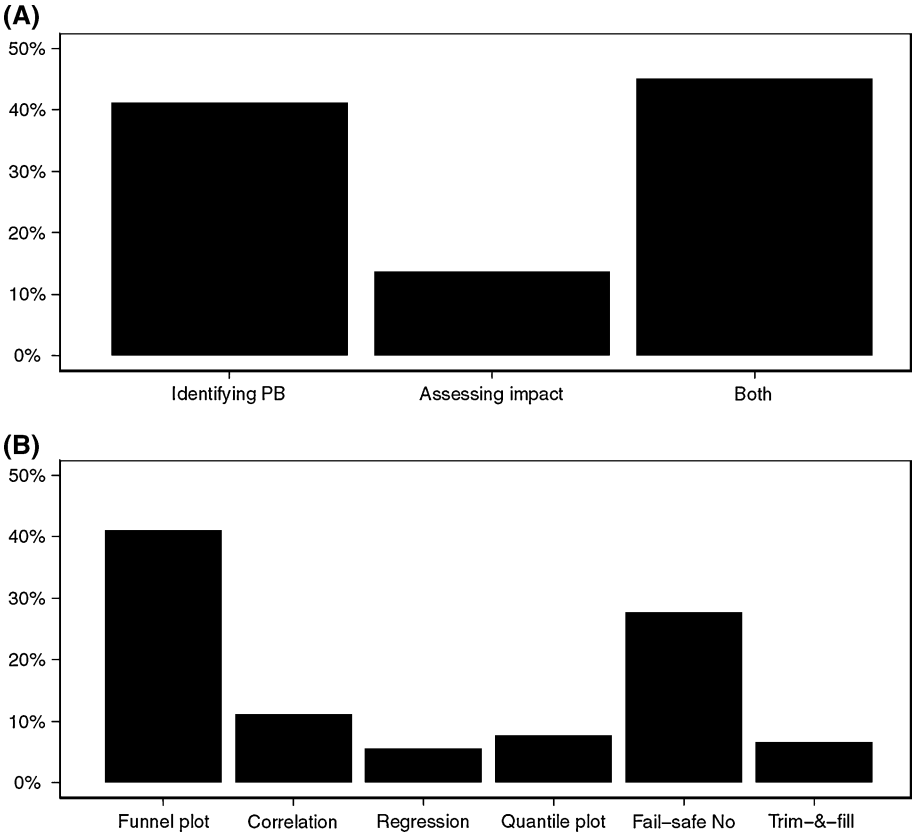

Fig. 2 *Bar*-plots: **a** the frequency of meta-analyses in ecology and evolution that used procedures to identify publication bias (labelled 'Identifying PB'), or to assess the impact of publication bias only ('Assessing impact'), respectively, or to do both ('Both'; this panel contains a total of 51 records), and **b** the frequency of different procedures to identify publication bias (funnel plot = funnel plots, correlation = correlation based tests, regression = Egger's regression, and quantile plot = normal quantile plots), and of different methods of assessing the impact of publication bias (fail-safe no = fail-safe number, and trim-&-fill = trim and fill methods); this panel contains a total of 90 records (some studies used more than one method)

(Fig. 2a, more details in Fig. 2b). We now examine both methods for publication bias, especially in the context of biological meta-analysis.

Identifying publication bias

A funnel plot usually displays effect-size estimates on one axis against their corresponding precision (i.e. the inverse of standard errors) or sometimes, corresponding weights (the inverse of variance or sample sizes; Egger et al. 1997; Sterne et al. 2005; Peters et al. 2008; for methods using the normal quantile plot, see Wang and Bushman 1998). The funnel plot is generally considered to be a very good visual tool to identify publication bias, which is manifested as funnel asymmetry (Sterne et al. 2005; Sutton 2009; see Fig. 3a, b; but see Terrin et al. 2005). However, publication bias is only one of the potential causes of funnel asymmetry. Other causes listed by Egger et al. (1997) include: (1) true heterogeneity (see above), (2) data irregularities (e.g., poor methodologies and qualities of small studies), and (3) chance. More recently, funnel plots have been displayed with areas of statistical significance, termed 'contour-enhanced' funnel plots, which better distinguish publication bias from other causes of asymmetry such as variable study quality (Peters et al. 2008; see Fig. 3c, d). We are not aware of any biological meta-analyses using contour-enhanced funnel plots, and suggest that these plots will be a very useful and practical tool to identify publication bias and other causes of funnel asymmetry. Contour-enhanced funnel plots can be easily implemented in programmes such as the *R* packages *meta* (Schwarzer 2010) and *metafor* (Viechtbauer 2010). Funnel plots are not statistical tests so that one cannot statistically judge whether observed funnel shapes are more asymmetric than symmetric. There are two commonly used statistical methods to quantify funnel asymmetry: (1) The rank correlation method proposed by Begg and Mazumdar (1994), and (2) the linear regression method proposed by Egger et al. (1997) (often referred to as Egger's regression or Egger test). The former is basically Kendall's rank correlation between effect sizes and corresponding sampling variance. The latter (Egger's regression) can be written as:

$$q_j = b_0 + b_1 p_j + \varepsilon_j, \tag{36}$$

$$q_j = z_j \sqrt{w_j}, \tag{37}$$

$$p_j = \sqrt{w_j}, \tag{38}$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_{\varepsilon}^2 \mathbf{I}), \tag{39}$$

where Eq. 36 is an ordinary linear regression with the intercept, b_0 and slope, b_1 with the residual, ε (ε_j is the jth residual; $j = 1,\ldots, N_{study}$), p is referred to as precision, w (weight; the inverse of sampling variance) is as in Eq. 19 and the other notations are as above. When the intercept (b_0) is significantly different from zero, one deems an observed shape in a funnel plot asymmetric, and concludes that there is evidence for publication bias. Unfortunately, the major shortcoming of this type of statistical method is that the detection of publication bias largely relies on sample size (i.e. $N_{effect\text{-}size}$); thus, caution is required when using the combination of funnel plots and these statistical tests, or just funnel plots (especially when sample size is small). Furthermore, both the rank correlation and Egger's regression cannot be directly applied to biological meta-analysis where effect-size estimates are not independent. Also, as mentioned above, statistically significant funnel asymmetry may only reflect heterogeneity. Here, we suggest a simple solution to these problems by modifying the original Egger's regression. Imagine that we have a dataset, which is suitable for a phylogenetic multilevel meta-regression (i.e. Eqs. 29, 30), and that we find the best model to

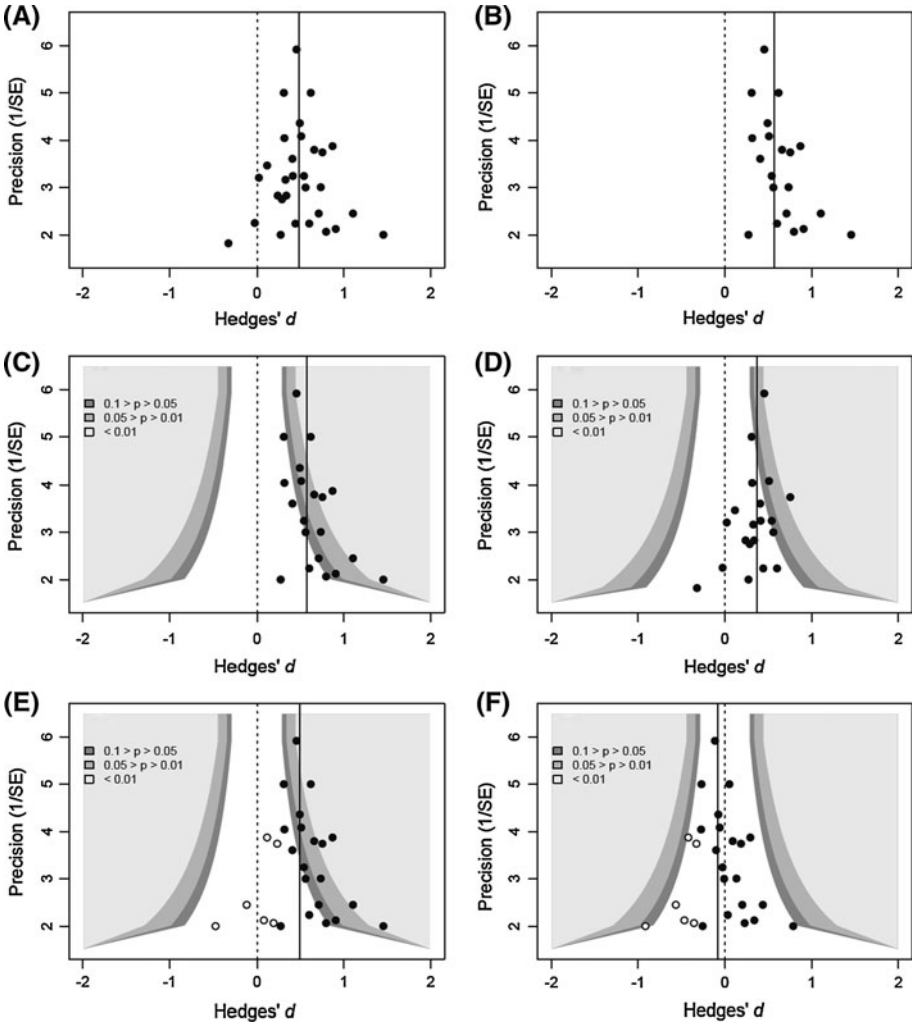

Fig. 3 Funnels plots: **a** an example of a symmetric funnel, **b** an example of an asymmetric funnel, **c** a counter funnel plot with statistically non-significant effect sizes potentially missing due to publication bias, **d** a counter funnel plot with statistically significant effect sizes potentially missing so that missing effect sizes are unlikely to be due to publication bias, **e** a funnel plot with trim and fill 'filled' data (*open circles*) and **f** a funnel plots with an example of meta-analytic residuals and the filled data (*open circles*); note that the black circles in **b**, **c** and **e** are the same data points. The *R* package *meta* (Schwarzer 2010) was used for creating the counter funnel plots

include several moderator variables. The funnel plot of the original data points looks very asymmetric, but this funnel asymmetry may be caused by heterogeneity, which can be accounted for by the moderators. Publication bias in this dataset can be assessed by use of a funnel plot, where the data points plotted against the corresponding (original) precision represent within-study effects, e_i plus their sampling-error effects, m_i (we refer to these as meta-analytic residuals), instead of effect sizes (see Roberts and Stanley 2005). In theory, meta-analytic residuals are independent from each other and free from effects of heterogeneity. An Egger's regression using meta-analytic residuals can be written as:

$$o_i = b_0 + b_1 p_i + \varepsilon_i, \tag{40}$$

$$o_i = (e_i + m_i)\sqrt{w_i}. \tag{41}$$

When b_0 (intercept) is significantly different from zero, one may conclude that there is evidence for publication bias after controlling for heterogeneity (note that the inclusion of moderators usually does not account for all heterogeneity). This procedure has been employed in recent biological meta-analytic studies (Sutton et al. 2011; Horváthová et al. 2012).

Another well-known type of publication bias is called time-lag bias. Time-lag bias arises when initially published studies have larger effects than those of later studies (Trikalinos and Ioannidis 2005). Interestingly, time-lag bias has been detected in the field of ecology and evolution on several occasions (Poulin 2000; Jennions and Møller 2002; Santos et al. 2011). Such temporal effects can easily be detected by including the publication year of the study as a moderator in the meta-analytic model. Alternatively, cumulative meta-analysis where meta-analytic means are cumulatively calculated over the years, in conjunction with forest plots (i.e. plotting such cumulative meta-analytic means by year), can visually detect temporal trends (see Leimu and Koricheva 2004). It is noteworthy that a recent 'meta-analysis' of meta-analyses by Barto and Rillig (2012) identified many types of so-called 'dissemination' biases. These occur when citation rates of papers are determined, regardless of sample sizes, by variables, such as journal impact factors and/or the magnitudes of effect sizes (see also Leimu and Koricheva 2005).

Assessing the impact of publication bias

In this section, we only focus on the trim and fill method among the three methods (selection model approaches, the trim and fill method and fail-safe *N*) for the following reasons. Selection model approaches explicitly model the processes of missing studies (e.g., the function of *p* values) and such approaches are possibly the best way to correct for the effect of missing studies (Hedges and Vevea 2005; Sutton 2009). However, the implementation is rather difficult and technical (for an implementation using *BUGS*, see Congdon 2003; also see the *R* package *copas* and its application by Schwarzer et al. 2010). The fail-safe *N* represents the number of studies required to refute significant meta-analytic means and several versions of fail-safe *N* have been widely used in earlier meta-analyses (Møller and Jennions 2001; Becker 2005; Jennions et al. 2012). However, these versions have been heavily criticised, mainly because such numbers are often misused and misinterpreted (reviewed in Becker 2005). The main reason for the criticism is that depending on which method is used to estimate the fail-safe *N*, the number of studies can greatly vary. Fail-safe *N*, although apparently intuitive, is in reality difficult to interpret not only because the number of data points (i.e. sample size) for each of *N* studies is not defined, but also because no benchmarks regarding fail-safe *N* exist, unlike Cohen's benchmarks for effect size statistics (Cohen 1988). We note that the method proposed by Rosenberg (2005) addresses some (but not all) concerns raised by Becker (2005).

The trim and fill method (Duval and Tweedie 2000a, b) has been implemented in many programmes (reviewed in Borenstein 2005), and used extensively in meta-analysis in general (Sutton 2009). The trim and fill method provides a conceptually easy and visually appealing way to adjust for the impact of missing studies in meta-analyses. It restores funnel symmetry by using existing data points to impute missing studies (for an accessible description of the procedure, see Duval 2005; see Fig. 3e). However, again this method has been designed for meta-analysis where independence of data points can be assumed.

Furthermore, the performance of the trim and fill method is limited when heterogeneity occurs (Peters et al. 2007). We propose that the same logic we applied to the Egger's regression above (39–40) can be used to resolve these issues. The trim and fill can be applied to meta-analytic residuals (which have a mean of 0). The difference between 0 and the new mean (intercept) calculated after the application of a trim and fill procedure can then be considered as the required adjustment (Fig. 3f). Sutton et al. (2011) used this approach and found for the trim and fill adjusted mean (intercept) a value of -0.167 (Hedges' *d*). They used this value to adjust the original meta-analytic mean along with its credible interval, and found that their original conclusions were robust.

It should be noted that the trim and fill method should be considered as sensitivity analysis rather than a true adjustment of meta-analytic means (Duval 2005; Peters et al. 2007). Unfortunately, none of the adjustment methods will recover 'true' meta-analytic means because the number of missing studies remains unknown.

Across-study and within-study meta-analysis

So far, the issues we discussed were related to meta-analyses across studies. An obvious merit of across-study meta-analysis is its ability to maximize statistical power (Arnqvist and Wooster 1995; Harrison 2011). The logic of meta-analysis can also easily be used to summarise results within studies, or within-study meta-analysis. It is easy to imagine situations in ecological and evolutionary empirical studies where within-study meta-analysis could be valuable. Empirical studies frequently use several approaches (e.g., observational and experimental) and measurements (e.g., molecular, behavioural and physiological) to test hypotheses. Thus, researchers will often obtain several related effect-size estimates, which can potentially be combined to draw a general conclusion for their empirical work. The idea of within-study meta-analysis is not new and it has been advocated before (Garamszegi 2006; Nakagawa and Hauber 2010) and a small number of empirical studies in ecology and evolution have indeed used this approach (e.g., Van den Bergh and Dewitte 2006; Milner et al. 2010). However, these previous within-study meta-analyses failed to adequately address the problem of non-independence among effect-size estimates. Such non-independence issues can be easily addressed by using the formula below:

$$\mathrm{var}\left(\frac{1}{n}\sum z_i\right) = \left(\frac{1}{n}\right)^2 \left(\sum \sigma_i^2 + \sum_{i \neq g} r_{ig}\sqrt{\sigma_i^2 \sigma_g^2}\right), \tag{42}$$

where the left-hand side of the equation represents the variance of the mean of n correlated effect-size estimates, r_{ig} is the correlation between the ith effect size (z_i; any normally distributed standardized effect size statistics) and the gth effect size (z_g) (r can be estimated or set by a researcher; $i = 1,\ldots, n$) and the other notations are as above (Borenstein et al. 2009). In recent across-study meta-analyses, this formula has been used to deal with non-independence incurred by multiple effect-sizes per study (i.e. collapsing effect-size estimates to a study level; e.g., Davidson et al. 2011; Slatyer et al. 2011). Santos and Macedo (2011) used this approach to carry out a within-study meta-analysis in an empirical study comparing maternal investments between cooperative breeding southern lapwings (*Vanellus chilensis*) with and without helpers. Maternal investment was measured via seven traits, such as clutch size, egg volume/content, and chick size. Only one trait (egg volume) was found to differ significantly between the two groups of females, most likely due to

limited statistical power. However, the within-study meta-analytic mean clearly showed that females with helpers invest less, as would be expected under the hypothesis of 'load lightening' (Hatchwell 1999).

We note that another way of doing within-study meta-analysis is to modify Eqs. 11, 12 by replacing **A** with a correlation matrix of effect-sizes (this matrix should be constructed prior to analysis). Such a within-study meta-analytic model can be easily implemented in *MCMCglmm*, *ASReml* or *BUGS*. We note that while there is no baseline number of effect-size estimates to warrant such an approach, it is clear that within-study meta-analysis will have to use simpler meta-analytic models due to limited degrees of freedom. Regardless of the approaches used, we encourage more empirical studies to conduct within-study meta-analysis to summarise their empirical work. With within-study meta-analysis, the methods of meta-analysis become relevant to many more researchers. This will promote 'effective thinking' among scientists; that is, the interpretation of results in terms of effect sizes rather than statistical significance *per se* (Nakagawa and Cuthill 2007; for the related concept, 'meta-analytic thinking', see Cumming and Finch 2001; Thompson 2002).

Final remarks

Gurevitch and Hedges (1999) wrote a review titled "*Statistical issues in ecological meta-analysis*". The last decade seems to have resolved many of the statistical issues they listed (but, for emerging practical problems in biological meta-analysis, see Lajeunesse 2010; Kueffer et al. 2011). These statistical developments, which we reviewed in this article, to some extent are not truly 'new' or 'recent'. In the field of quantitative genetics, most of the solutions and computer implementations presented here have existed at least since the mid 1990s, often labelled differently or hidden among technical literature (Hadfield and Nakagawa 2010; see also Ives and Zhu 2006). This fact highlights the importance of cross-disciplinary work, particularly for statistical methods. In fact, meta-analysis has been a great catalyst for inter-disciplinary exchange of ideas and methods (Gates 2002). Diverse statistical and meta-analytic techniques are actively being developed not only in medical and social sciences, but also in other disciplines; researchers in the fields of ecology and evolution should be open to such progress. Currently, methods dealing with phylogenetic relatedness and publication bias are rarely used in biological meta-analyses. We now have the appropriate meta-analytic tools and implementations to quantitatively synthesize most biological topics and their complexities.

Acknowledgments The authors thank J. Endler for conceiving this special issue and E. Schlicht, S. Chamberlain, M. Lagisz and two anonymous reviewers for the comments on an early version of this paper. S.N. acknowledges the Humboldt Fellowship and Marsden Fund for support and also is grateful to B. Kempenaers for hosting and providing an excellent working environment at the MPIO. E.S.A.S. acknowledges the support of a University of Otago PhD Scholarship.

Appendix: A survey of biological meta-analytic studies

We conducted a survey of the meta-analyses published in the last 3 years (2009–2011). We performed a search of the online database Scopus, by using the following keywords: "meta-analys*", and "meta-regression" (last updated in October 2011); we restricted our search to the following journal titles: *American Naturalist*, *Animal Behaviour*, *Behavioral Ecology*, *Behavioral Ecology and Sociobiology*, *Biological Reviews, Ecology*, *Ecological Application*, *Ecology Letters*, *Ecological Monographs*, *Evolution, Evolutionary Ecology*,

Functional Ecology, *Journal of Evolutionary Biology*, *Journal of Animal Ecology*, *Journal of Applied Ecology*, *Journal of Ecology*, *Molecular Ecology*, *New Phytologist*, *Oecologia*, *Oikos*, and *Quarterly Reviews of Biology*. Our search yielded a total of 390 original studies; we used the most recent 100 'biological' meta-analytic studies as defined in the Introduction. From each study, we retrieved the following information: (1) the type of effect size used, (2) whether the study accounted for phylogenetic relatedness or not, (3) whether the study used any procedure to deal with publication bias, and (4) the software used to conduct the meta-analysis. In addition, we classified the effect sizes into the three main categories (i.e., means, correlation coefficient, and dichotomous; see section 'Effect-size statistics in ecology and evolution'). All other types of effect sizes (e.g. repeatability and selection coefficients) were pooled in one category (named 'others'). If the studies used any procedure for dealing with publication bias, we classified them into one of three categories: (1) studies that only identified publication bias, (2) studies that only assessed the impact of publication bias; or (3) studies that conducted both types of procedures (see the section 'Publication bias'). In addition to the results reported in the main text, we found that most studies did not account (83%) for phylogenetic relatedness of the species included in the analyses. For this estimate, we excluded most of the ecological meta-analyses, because these studies mainly deal with effects on ecosystem or community scales, where phylogenetic non-independence does not apply. Most of the meta-analyses included in our survey used the software *MetaWin* to conduct their analyses (42.3% of studies). Eighteen studies did not report the software used; 19 different meta-analysis software packages were employed in the remaining studies.

References

Adams DC (2008) Phylogenetic meta-analysis. Evolution 62:567–572

Arnqvist G, Wooster D (1995) Meta-analysis: synthesizing research findings in ecology and evolution. Trends Ecol Evol 10:236–240

Barto EK, Rillig MC (2012) Dissemination biases in ecology: effect sizes matter more than quality. Oikos 121:228–235

Becker BJ (2005) Fail safe *N* or file-drawer number. In: Rothstein H, Sutton AJ, Borenstein M (eds) Publication bias in meta-analysis: prevention, assessment and adjustments. Wiley, Chichester, pp 111–125

Begg CB, Mazumdar M (1994) Operating characteristics of a rank correlation test for publication bias. Biometrics 50:1088–1101

Blomberg SP, Garland T, Ives AR (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:717–745

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White JSS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol 24:127–135

Borenstein M (2005) Software for publication bias. In: Rothstein H, Sutton AJ, Borenstein M (eds) Publication bias in meta-analysis: prevention, assessment and adjustments. Wiley, Chichester, pp 193–220

Borenstein M, Hedges LV, Higgins JPI, Rothstein HR (2009) Introduction to meta-analysis. Wiley, Chichester

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, Berlin

Carmona D, Lajeunesse MJ, Johnson MTJ (2011) Plant traits that predict resistance to herbivores. Funct Ecol 25:358–367

Cheung MWL (2008) A model for integrating fixed-, random-, and mixed-effects meta-analyses into structural equation modeling. Psychol Methods 13:182–202

Cleasby IR, Nakagawa S (2012) The influence of male age on within-pair and extra-pair paternity in passerines. Ibis. doi:10.1111/j.1474-919X.2011.01209.x

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

Congdon P (2003) Applied Bayesian modelling. Wiley, Chichester
Cooper H, Hedges LV, Valentine JC (2009) The handbook of research synthesis and meta-analysis, 2nd edn. Russell Sage Foundation, New York
Cornwallis CK, West SA, Davis KE, Griffin AS (2010) Promiscuity and the evolutionary transition to complex societies. Nature 466:969–972
Cumming G, Finch S (2001) A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. Educ Psychol Measurement 61:532–584
Davidson AM, Jennions M, Nicotra AB (2011) Do invasive species show higher phenotypic plasticity than native species and, if so, is it adaptive? A meta-analysis. Ecol Lett 14:419–431
Dubois F, Cezilly F (2002) Breeding success and mate retention in birds: a meta-analysis. Behav Ecol Sociobiol 52:357–364
Duval S (2005) The trim and fill method. In: Rothstein H, Sutton AJ, Borenstein M (eds) Publication bias in meta-analysis: prevention, assessment and adjustments. Wiley, Chichester, pp 127–144
Duval S, Tweedie R (2000a) A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. J Am Stat Assoc 95:89–98
Duval S, Tweedie R (2000b) Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 56:455–463
Egger M, Smith GD, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. Br Med J 315:629–634
Egger M, Smith GD, Altman DG (2001) Systematic reviews in health care: meta-analysis in context, 2nd edn. BMJ, London
Evans SR, Hinks AE, Wilkin TA, Sheldon BC (2010) Age, sex and beauty: methodological dependence of age- and sex-dichromatism in the great tit *Parus major*. Biol J Linn Soc 101:777–796
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
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
Gasparrini A (2011) Multivariate meta-analysis and meta-regression: package 'mvmeta'. In: R package version 0.2.3 edn. http://cran.r-project.org/web/packages/mvmeta/index.html
Gates S (2002) Review of methodology of quantitative reviews using meta-analysis in ecology. J Anim Ecol 71:547–557
Gelman A, Hill J (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge
Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R (2002) ASReml user guide release 1.0. VSN Internationl Ltd, Hemel Hempstead, UK
Griffin AS, West SA (2003) Kin discrimination and the benefit of helping in cooperatively breeding vertebrates. Science 302:634–636
Griffin AS, Sheldon BC, West SA (2005) Cooperative breeders adjust offspring sex ratios to produce helpful helpers. Am Nat 166:628–632
Grueber CE, Nakagawa S, Laws RJ, Jamieson IG (2011) Multimodel inference in ecology and evolution: challenges and solutions. J Evol Biol 24:699–711
Gurevitch J, Hedges LV (1999) Statistical issues in ecological meta-analyses. Ecology 80:1142–1149
Gurevitch J, Curtis PS, Jones MH (2001) Meta-analysis in ecology. Adv Ecol Res 32:199–247
Hadfield JD (2010) MCMC methods for multi-response generalised linear mixed models: the MCMCglmm R package. J Stat Softw 33:1–22
Hadfield JD, Nakagawa S (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. J Evol Biol 23:494–508
Hamilton WD (1964) The genetical evolution of social behaviour. I and II. J Theor Biol 7:1–52
Harrison F (2011) Getting started with meta-analysis. Methods Ecol Evol 2:1–10
Harvey PH, Pagel MD (1991) The comparative method in evolutionary biology. Oxford University Press, Oxford
Hatchwell BJ (1999) Investment strategies of breeders in avian cooperative breeding systems. Am Nat 154:205–219
Hector KL, Nakagawa S (2012) Quantitative analysis of compensatory and catch-up growth in diverse taxa. J Anim Ecol. doi:10.1111/j.1365-2656.2011.01942.x
Hedges L, Olkin I (1985) Statistical methods for meta-analysis. Academic Press, New York

Hedges LV, Vevea JL (1998) Fixed- and random-effects models in meta-analysis. Psychol Methods 3:486–504
Hedges LV, Vevea JL (2005) Selection method approaches. In: Rothstein H, Sutton AJ, Borenstein M (eds) Publication bias in meta-analysis: prevention, assessment and adjustments. Wiley, Chichester, pp 145–174
Hedges LV, Gurevitch J, Curtis PS (1999) The meta-analysis of response ratios in experimental ecology. Ecology 80:1150–1156
Higgins JPT, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. Stat Med 21:1539–1558
Higgins JPT, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. Br Med J 327:557–560
Higgins JPT, Thompson SG, Spiegelhalter DJ (2009) A re-evaluation of random-effects meta-analysis. J R Stat Soc A 172:137–159
Horváthová T, Nakagawa S, Uller T (2012) Strategic female reproductive investment in response to male attractiveness in birds. Proc R Soc B 279:163–170
Hunt M (1997) How science takes stock: the story of meta-analysis. Russell Sage, New York
Ives AR, Zhu J (2006) Statistics for correlated data: phylogenies, space, and time. Ecol Appl 16:20–32
Ives AR, Midford PE, Garland T (2007) Within-species variation and measurement error in phylogenetic comparative methods. Syst Biol 56:252–270
Jackson D, Riley R, White IR (2011) Multivariate meta-analysis: potential and promise. Stat Med 30:2481–2498
Jennions MD, Møller AP (2002) Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution. Proc R Soc B 269:43–48
Jennions MD, Lorite C, Rosenberg M, Rothstein H (2012) Publication and related biases. In: Koricheva J, Gurevitch J, Mengersen K (eds) The handbook of meta-analysis in ecology and evolution. Princeton University Press, Princeton (in press)
Jones KS, Nakagawa S, Sheldon BC (2009) Environmental sensitivity in relation to size and sex in birds: meta-regression analysis. Am Nat 174:122–133
Kelly CD, Jennions MD (2011) Sexual selection and sperm quantity: meta-analyses of strategic ejaculation. Biol Rev 86:863–884
Knowles SCL, Nakagawa S, Sheldon BC (2009) Elevated reproductive effort increases blood parasitaemia and decreases immune function in birds: a meta-regression approach. Funct Ecol 23:405–415
Koricheva J, Gurevitch J, Mengersen K (2012) The handbook of meta-analysis in ecology and evolution. Princeton University Press, Princeton
Kueffer C, Niinemets U, Drenovsky RE, Kattge J, Milberg P, Poorter H, Reich PB, Werner C, Westoby M, Wright IJ (2011) Fame, glory and neglect in meta-analyses. Trends Ecol Evol 26:493–494
Lajeunesse MJ (2009) Meta-analysis and the comparative phylogenetic method. Am Nat 174:369–381
Lajeunesse MJ (2010) Achieving synthesis with meta-analysis by combining and comparing all available studies. Ecology 91:2561–2564
Lajeunesse MJ (2011a) On the meta-analysis of response ratios for studies with correlated and multi-group designs. Ecology 92:2049–2055
Lajeunesse MJ (2011b) phyloMeta: a program for phylogenetic comparative analyses with meta-analysis. Bioinformatics 27:2603–2604
Lajeunesse MJ, Forbes MR (2003) Variable reporting and quantitative reviews: a comparison of three meta-analytical techniques. Ecol Lett 6:448–454
Leimu R, Koricheva J (2004) Cumulative meta-analysis: a new tool for detection of temporal trends and publication bias in ecology. Proc R Soc B 271:1961–1966
Leimu R, Koricheva J (2005) What determines the citation frequency of ecological papers? Trends Ecol Evol 20:28–32
Liermann M, Hilborn R (1997) Depensation in fish stocks: a hierarchic Bayesian meta-analysis. Can J Fish Aquat Sci 54:1976–1984
Lipsey MW, Wilson DB (2001) Practical meta-analysis. Sage, Beverly Hills
Lortie CJ, Aarssen LW, Budden AE, Koricheva JK, Leimu R, Tregenza T (2007) Publication bias and merit in ecology. Oikos 116:1247–1253
Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Stat Computing 10:325–337
Lynch M (1991) Methods for the analysis of comparative data in evolutionary biology. Evolution 45:1065–1080
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

Meunier J, Pinto SF, Burri R, Roulin A (2011) Eumelanin-based coloration and fitness parameters in birds: a meta-analysis. Behav Ecol Sociobiol 65:559–567
Milner RNC, Detto T, Jennions MD, Backwell PRY (2010) Experimental evidence for a seasonal shift in the strength of a female mating preference. Behav Ecol 21:311–316
Møller AP, Jennions MD (2001) Testing and adjusting for publication bias. Trends Ecol Evol 16:580–586
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 RP (2008) Missing inaction: the dangers of ignoring missing data. Trends Ecol Evol 23:592–596
Nakagawa S, Hauber ME (2010) Great challenges with few subjects: statistical strategies for neuroscientists. Neurosci Biobehav Rev 35:462–473
Osenberg CW, Sarnelle O, Cooper SD, Holt RD (1999) Resolving ecological questions through meta-analysis: goals, metrics, and models. Ecology 80:1105–1117
Pagel M (1999) Inferring the historical patterns of biological evolution. Nature 401:877–884
Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L (2007) Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. Stat Med 26:4544–4562
Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L (2008) Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J Clin Epidemiol 61:991–996
Pigott TD (2009) Handling missing data. In: Cooper H, Hedges LV, Valentine JC (eds) The handbook of research synthesis and meta-analysis. Russell Sage Foundation, New York, pp 399–416
Pinheiro JC, Bates DM (2000) Mixed-effects models in S and S-Plus. Springer, New York
Poulin R (2000) Manipulation of host behaviour by parasites: a weakening paradigm? Proc R Soc B 267:787–792
Roberts CJ, Stanley TD (2005) Meta-regression analysis: issues of publication bias in economics. Blackwell, Malden
Rosenberg MS (2005) The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. Evolution 59:464–468
Rosenberg MS, Adams DC, Gurevitch J (2000) MetaWin: statistical software for meta-analysis, 2nd edn. Sinauer, Sunderland
Rosenthal R (1979) The "file drawer problem" and tolerance for null results. Psychol Bull 86:638–641
Rothstein H, Sutton AJ, Borenstein M (2005) Publication bias in meta-analysis: prevention, assessment and adjustments. Wiley, Chichester
Santos ESA, Macedo RH (2011) Load lightening in southern lapwings: group-living mothers lay smaller eggs than pair-living mothers. Ethology 117:547–555
Santos ESA, Scheck D, Nakagawa S (2011) Dominance and plumage traits: meta-analysis and metaregression analysis. Anim Behav 82:3–19
Schielzeth H (2010) Simple means to improve the interpretability of regression coefficients. Methods Ecol and Evol 1:103–113
Schwarzer G (2010) Meta-analysis with R: package 'meta'. In: R package version 1.6-1 edn. http://cran.r-project.org/web/packages/meta/index.html
Schwarzer G, Carpenter J, Rucker G (2010) Empirical evaluation suggests Copas selection model preferable to trim and fill method for selection bias in meta-analysis. J Clin Epidemiol 63:282–288
Slatyer RA, Mautz BS, Backwell PR, Jennions MD (2011) Estimating genetic benefits of polyandry from experimental studies: a meta-analysis. Biol Rev 86:863–884
Smith BR, Blumstein DT (2008) Fitness consequences of personality: a meta-analysis. Behav Ecol 19:448–455
Sterne JAC, Becker BJ, Egger M (2005) The funnel plot. In: Rothstein H, Sutton AJ, Borenstein M (eds) Publication bias in meta-analysis: prevention, assessment and adjustments. Wiley, Chichester, pp 75–98
Sutton AJ (2009) Publication bias. In: Cooper H, Hedges LV, Valentine JC (eds) The handbook of research synthesis and meta-analysis. Russell Sage Foundation, New York, pp 435–452
Sutton AJ, Higgins JPI (2008) Recent developments in meta-analysis. Stat Med 27:625–650
Sutton JT, Nakagawa S, Robertson BC, Jamieson IG (2011) Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes. Mol Ecol 20:4408–4420
Terrin N, Schmid CH, Lau J (2005) In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. J Clin Epidemiol 58:894–901
Thompson B (2002) What future quantitative social science research could look like: confidence intervals for effect sizes. Educ Researcher 31:25–32

Trikalinos TA, Ioannidis JP (2005) Assessing the evolution of effect sizes over time. In: Rothstein H, Sutton AJ, Borenstein M (eds) Publication bias in meta-analysis: prevention, assessment and adjustments. Wiley, Chichester, pp 241–259

Van den Bergh B, Dewitte S (2006) Digit ratio (2D: 4D) moderates the impact of sexual cues on men's decisions in ultimatum games. Proc R Soc B 273:2091–2095

Verdu M, Traveset A (2004) Bridging meta-analysis and the comparative method: a test of seed size effect on germination after frugivores' gut passage. Oecologia 138:414–418

Verdu M, Traveset A (2005) Early emergence enhances plant fitness: a phylogenetically controlled meta-analysis. Ecology 86:1385–1394

Viechtbauer W (2010) Conducting meta-analyses in R with the meta for package. J Stat Softw 36:1–48

Wang MC, Bushman BJ (1998) Using the normal quantile plot to explore meta-analytic data sets. Psychol Methods 3:46–54

Wehi P, Nakagawa S, Trewick S, Morgan-Richards M (2011) Does predation result in adult sex ratio skew in a sexually dimorphic insect genus? J Evol Biol 24:2321–2328

Weir LK, Grant JWA, Hutchings JA (2011) The influence of operational sex ratio on the intensity of competition for mates. Am Nat 177:167–176