Evol Ecol (2012) 26:1187–1196
DOI 10.1007/s10682-012-9562-6

ORIGINAL PAPER

Meta-analysis of meta-analyses in plant evolutionary ecology

Maria Clara Castellanos · Miguel Verdú

Received: 13 November 2011 / Accepted: 13 February 2012 / Published online: 24 February 2012

Abstract After two decades of meta-analyses on plant traits, we can now look for global emergent patterns in plant evolutionary ecology. Hundreds of meta-analyses have focused on the effects of specific selection pressures on plant fitness, and the buildup of such results allows us to ask general questions regarding selection pressures and plant responses, a major focus of evolutionary ecology. Plant traits are affected by both abiotic and biotic factors. For example, biotic pressures like herbivory may affect physiological (i.e. secondary defences) and reproductive (i.e. seed predation) traits. Similarly, abiotic pressures such as increased CO_2 may affect both plant physiology and reproduction. We tested whether biotic or abiotic selective pressures are more important for plant traits, and if the strength of the response to those pressures depends on the plant trait studied by meta-analyzing published meta-analyses on plant responses. We classify meta-analyses according to the type of response variable studied (fitness and non-fitness traits) and the type of selective pressure examined (biotic or abiotic). Our database showed biases in the meta-analysis literature, for example that the majority of studies are focused on non-fitness traits, i.e. on traits that are not directly related to reproduction or survival, and furthermore, on non-fitness traits under abiotic selection pressures. The meta-meta-analysis showed that the strength of responses to selection depends on the nature of selection (stronger for biotic than for abiotic factors) but, unexpectedly, not on the type of trait under study as previously found. The stronger responses to biotic factors can be explained if biotic selection is more variable in space and time, driven by interactions with other organisms. The relative importance of biotic versus abiotic factors on plant traits has been little studied in the past, and would benefit from more studies and reviews that fill the under-represented combinations of selective pressures and plant traits (i.e. abiotic factors on fitness traits).

Electronic supplementary material The online version of this article (doi:10.1007/s10682-012-9562-6) contains supplementary material, which is available to authorized users.

M. C. Castellanos · M. Verdú (✉)
Consejo Superior de Investigaciones Científicas, Centro de Investigaciones sobre Desertificación
(CSIC-UV-GV), 46113 Moncada, Valencia, Spain
e-mail: miguel.verdu@uv.es

M. C. Castellanos
e-mail: maclacas@uv.es

Keywords Biotic and abiotic selective pressures · Fitness and non-fitness traits · Second-order meta-analysis

Introduction

After two decades of meta-analyses on plant traits, beginning with the introduction of the technique in the field of ecology in the early 1990s (see Gurevitch et al. 2001), it is timely to look for global emerging patterns in the literature. In the field of plant evolutionary ecology specifically, hundreds of meta-analytical studies have focused on the effects of specific selection pressures on a single or a few plant fitness components. As a result, generalizations can be made, for example on how increased N availability or increased herbivore activity can affect plant growth or reproductive output and how such effects hold across experiments and plant species. The buildup of meta-analytic results, however, allows us to go further and ask more general questions regarding selection pressures and plant responses, i.e. the types of questions on patterns of selection that are a major focus of modern evolutionary ecology. Key questions that can be explored are whether biotic or abiotic selective pressures are more important for plants, and if the strength of the response to those pressures depends on the plant trait studied.

Natural selection is an important force behind phenotypic differentiation across a wide range of plant traits (Kingsolver et al. 2001; Rieseberg et al. 2002). However, not all traits are expected to be targeted by selection in the same way. For example, traits closely related to fitness, such as life history traits, are expected to experience stronger selection than other types of traits (Merilä and Sheldon 1999). Tests of this idea have come to different conclusions, depending on the methodological approach. Kingsolver et al. (2001) compared selection gradients and differentials measured in wild populations across different types of traits, and found that morphological traits were subject to stronger selection than life history traits. Rieseberg et al. (2002) on the other hand, compared the signature of selection with a more “historical” approach, using the direction of effects of quantitative trait loci (the QTL sign test), and found evidence of stronger and more consistent selection on life history than on morphological characters.

Contrasting results are not necessarily surprising because selective pressures are expected to affect plant performance in complex ways (Bell 2010). For example, the strength, form and direction of selection can vary in time (Grant and Grant 2002; Siepielski et al. 2009; Kingsolver and Diamond 2011) and space (Linhart and Grant 1996; Schluter 2000; Herrera et al. 2006; but see also Morrisey and Hadfield 2012). In addition, the type of selection pressure, whether biotic or abiotic, could also exert different responses from plant traits. Biotic selective pressures depend on the interactions with other organisms, such as predators or mutualists, whose distributions and densities can vary rapidly and unpredictably and can therefore be expected to be less consistent in strength, space and time (Linhart and Grant 1996; Thompson 2005). Plant responses to biotic pressures could be then expected to be weaker and less consistent across species and populations than to abiotic pressures. However, a recent study suggests that biotically-selected traits are governed by fewer genes with a large effect, which could allow populations to move faster among variable peaks in an adaptive landscape (Louthan and Kay 2011). Although previous studies have tested for differences between measures of selection on fitness traits compared to other types of traits, to our knowledge no studies have specifically explored the potential differences in selection when the pressures are biotic or abiotic.

We assembled here a database of diverse meta-analyses that allows us to simultaneously test for the strength of the effects of biotic and abiotic selection pressures and the responses of different types of traits (fitness vs. non-fitness). We also tested for the interaction between them, which could reveal differential effects of biotic or abiotic factors on different types of traits. We address these questions quantitatively by performing a meta-analysis of published meta-analyses, or a second-order meta-analysis, an approach that has been little used in ecology so far, but is already common practice in the medical sciences (usually referred to as "umbrella reviews" when various reviews are compared in narrative form, or "multiple treatment meta-analysis" when multiple meta-analytic results are compared under specific models; Caldwell et al. 2010; Ioannidis 2009). Specifically, we compare, with meta-analytical techniques, (1) the global effect sizes of meta-analyses of biotic versus abiotic selection pressures, (2) the global effect sizes of fitness versus non-fitness response traits, and (3) the interaction between them. Note that we are not dealing with data on selection gradients or differentials (as defined by Lande and Arnold 1983), but with studies that control or measure the selective factors and record their effect on plant traits. In addition, we use our database to describe patterns in the published meta-analysis literature on plant evolutionary ecology and detect potential biases towards certain types of reviews.

Materials and methods

Clarification of the terminology used in the remaining of the article follows. The data base used in our qualitative and quantitative analyses is composed of *meta-analyses* mean effect sizes extracted from *publications* that may or may not include more than one meta-analysis. Each meta-analysis in turn included *original case studies*. Data points in our *second-order meta-analysis* are meta-analyses mean effect sizes and not the original case studies. Methods are detailed below.

We compiled the data set of published meta-analyses on plant traits by performing a literature search in the Web of Science with topic keywords "meta-analysis and plant" (as of September 2011). We purged down the initial list of around 440 publications to include meta-analyses that met the following requisites. (a) Studies had to perform a formal meta-analysis, that is, a comparison of weighted effect sizes across data sets. (b) Meta-analyses were revisions of the published literature designed to extract general patterns. This excludes studies that used formal meta-analytical techniques to compare various sets of original data. (c) We excluded meta-analyses performed exclusively on crop species under agricultural conditions, because a long history of artificial selection might affect current response to selective pressures. (d) We included only meta-analyses focused on plant traits that can be measured in individuals. Community level (e.g. species richness) or ecosystem level traits (e.g. litter decomposition) were not considered. (e) We also excluded allometric meta-analyses that were purely morphological (e.g. trunk diameter vs. leaf area), when they had no clear evolutionary implications.

We classified the remaining meta-analyses according to the type of response variable studied (growth, physiology, reproduction or survival) and the type of selective pressure examined (biotic or abiotic). Response variables were in turn grouped as fitness variables (reproduction and survival) or non-fitness (physiological and growth traits). This division might not seem straightforward, as it can be argued that growth or development are fitness components as well. Our rationale follows that of Merilä and Sheldon (1999), which assumes that reproductive traits and survival are more closely related to fitness itself than other traits.

From each meta-analysis we extracted global effect sizes and their associated sample sizes and sampling error variances to use them as weights. Sampling error variance is the square of the standard error, but these estimates are seldom reported in the literature. Instead, 95% confidence interval of the effect size is usually provided and half the width of the 95% CI divided by 1.96 is a good approximation to the standard error. We did not include partial effect sizes (predictor factors) that subdivide data sets already used to calculate a global effect (e.g. subdividing data sets to test the effect of ant mutualisms on herbivory in shrubs vs. herbs, Chamberlain and Holland 2009), to avoid pseudoreplication. When several global effect sizes were provided by the same publication to test separate response variable types (e.g. physiological, reproduction, etc.), we included all of them. For example, mutualism effects on growth and reproduction of target plants were studied independently by Trager et al. (2010) and therefore we included two global effects from this publication. Furthermore, if the original meta-analysis mixed the types of response variables we were interested in, we recalculated a global effect size for each variable type if the original data set was available. For example, Bailey et al. (2009) reported effects of introgression on a mixture of physiological, morphological, and reproductive response variables in *Populus*. We recalculated global effect sizes for growth and physiological response variables separately from their supplementary data set.

For our final second-order meta-analysis, we needed to transform individual meta-analyses' effect sizes to a common metric. However, the most common effect metric used in ecological studies, the log of the response ratio (lnRR) cannot be transformed into other metrics in a straightforward way (M. Lejaunesse, personal communication.). We therefore limited our quantitative analysis to meta-analyses reporting lnRR and closely related metrics (e.g. percentage of change) and excluded those reporting metrics based on standardized mean differences (i.e. Hedges *d*) or correlation coefficients. Because the $\mathrm{lnRR} = \ln(X^E) - \ln(X^C)$, i.e. the ratio of the outcome of an experimental group to that of a control group, our database for the quantitative analysis is composed mostly of meta-analyses of controlled experimental studies, but not exclusively, because some also include original case studies using natural variation (e.g. Chamberlain and Holland 2009; Trager et al. 2010).

There was no significant correlation between effect size and sample size ($r = 0.02$, $df = 137$, $P = 0.81$), suggesting against the biased publication of high effect sizes.

Statistical analyses

Because we were only interested in the strength of plant trait responses to selective pressures, the sign of the effect sizes was not informative in our analysis. We therefore used the absolute values of effect sizes (lnRR) to run Bayesian meta-analyses as explained below. Using the absolute values could introduce an upward bias when estimated effect sizes are non-significantly different from zero (Hereford et al. 2004). However, we do not expect this to affect our comparisons, because around 80% of the reported meta-analyses were significant. In addition, we are not testing for significance in effect sizes, but rather for differences in their strength.

We first calculated an index of heterogeneity among meta-analyses (I^2; Higgins and Thompson 2002) using the MCMCglmm R package as suggested by Nakagawa and Santos (2012). Values of I^2 around 25, 50 and 75% reflect small, medium and large heterogeneity (Higgins et al. 2003). For the second-order meta-analysis, we fitted generalized linear mixed models using Markov chain Monte Carlo techniques with the help of the MCMCglmm package for R (Hadfield 2010). The effect size was the dependent variable in the model, and two types of weights were used: (1) sample size and (2) inverse of the sampling error variance. Both weighting strategies have been used in social sciences (Hunter and Schmidt 2004 and references therein) as well as in ecology (e.g., van Groenigen et al. 2011). Comparisons of the performance of both methods can be found in Marín-Martínez and Sánchez-Meca (2010) and in Lajeunesse and Forbes (2003).Weights were passed to the *mev* argument of MCMCglmm (Hadfield and Nakagawa 2010). We ran 13,000 MCMC iterations with a burn-in period of 3,000 iterations and convergence of the chain was tested by means of an autocorrelation statistic. The priors used were $nu = 0$ and $V = I \times 1e + 10$, where I is an identity matrix of appropriate dimension. The type of selective variable (biotic and abiotic) and the type of response variable (fitness and non-fitness) were included as predictors in the MCMCglmm model, including an interaction. Although separate global effect sizes could come from the same publications, we decided against using the publication as a random grouping factor in the model. This is because (1) separate meta-analyses reported in the same publications are not necessarily non-independent, because they are derived from different sets of original study cases, and (2) publications deal with only one of the selective variable types (biotic or abiotic), so that including it as a random factor would remove important variance from the main predictors unintentionally. The effect of predictors was estimated by calculating the 95% credible interval of their posterior distribution (Nakagawa and Cuthill 2007).

Results

General patterns in the literature

Our final data set included 196 meta-analyses based on more than 17,800 original study cases, reported in 51 publications (Table 1 and Appendix in ESM). This sample reflects a bias in the literature towards meta-analyses of non-fitness traits (154 vs. 42 involving fitness responses), and particularly towards those of non-fitness traits under abiotic selection (102 studies). In contrast, only 9 meta-analyses in our data base dealt with biotic characters under abiotic selective pressures.

Most abiotic selective pressures were climatic variables (111 vs. only 3 related to disturbance). Among the climatic variables, there is a majority of meta-analyses dealing with responses to elevated CO_2 (50 meta-analyses) and exposure to UV-B radiation (25). Biotic pressures are all related to interactions, spanning from ant-plant mutualisms (10 meta-analyses), to herbivory (19), interactions with plant neighbors (11), and less often with plant-microbial interactions, pollinators, etc.

Finally, fitness responses are most often some measurement of reproductive output (37 of 42 studies), while survival is the response variable in only 5 studies. In contrast, within non-fitness variables there is a balance between growth and physiological responses (77 each).

Quantitative analysis

As explained above, we limit our quantitative analysis to the subset of meta-analyses in our database reporting lnRR as the effect size (N = 139 meta-analyses in 30 publications. Sampling error variance could only be obtained from 134 meta-analyses, see Appendix in

Table 1 Number of meta-analyses in each category of selective pressures and trait response types included in this revision

Selective pressure	Response trait type	Meta-analyses in this study
Biotic	Fitness	
	Reproduction	28
	Survival	5
	Non-fitness	
	Physiology	14
	Growth	38
Abiotic	Fitness	
	Reproduction	9
	Survival	0
	Non-fitness	
	Physiology	63
	Growth	39

Details and references are in Appendix 1 in ESM

ESM). We detected a large value of heterogeneity among meta-analyses ($I^2 = 99.6\%$; [99.5, 99.7]), which justified using predictors. Results were very similar for both weighting—sample size and variance-procedures. We found no significant interaction between the type of selective variable (biotic and abiotic) and the type of response variable (fitness and non-fitness) in their effect on effect sizes (posterior mean estimate = −0.010, 95% CI [−0.251 to 0.208] for sample size weighted and −0.043 [−0.197, 0.079] for variance weighted models). We therefore tested for the main effects of the two variables in a model without interaction. It showed no significant differences in effect sizes between fitness and non-fitness response variables (−0.059, [−0.172, 0.044] for sample size weighted and −0.014 [−0.086, 0.051] for variance weighted models). However, there was a significant effect of the type of selective variable analyzed, because biotic variables elicit higher responses than abiotic ones (0.188 [0.104, 0.273] for sample size weighted and 0.177 [0.120, 0.234] for variance weighted models). Raw mean effect sizes and their standard errors are shown in Fig. 1. These results are

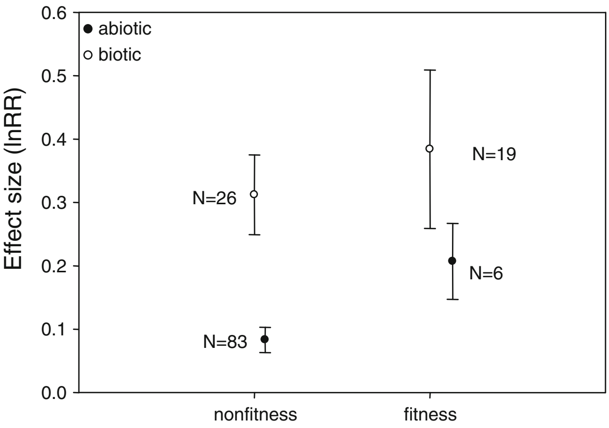

Fig. 1 Raw mean and 1 standard error of effect sizes (lnRR) for meta-analyses classified according to the type of selective variable (biotic and abiotic) and the type of response variable (fitness and non-fitness). Sample sizes for each group are included

unchanged if we include response variables as physiology, growth, or reproduction traits instead of grouping them as fitness or non-fitness.

Discussion

Our review of the meta-analytical literature of selection pressures on plants showed, on the one hand, that the majority of meta-analyses are studies of non-fitness traits and mostly on a few abiotic selection pressures such as increased CO_2 concentrations. On the other hand, these biases did not prevent a quantitative comparison of the effects of different selective pressures, which showed that the strength of responses to selection depends on the nature of selection (biotic versus abiotic factors) but, unexpectedly, not on the type of trait under selection. We discuss these results below.

Trends in the meta-analysis literature

Biases in our data base allowed us to detect biases in the meta-analysis literature. The majority of review studies are focused on non-fitness traits, i.e. on traits that are not directly related to reproduction or survival, and furthermore, on non-fitness traits under abiotic selection pressures. Certainly measuring a plant's reproductive output might be more difficult than measuring a morphological or physiological character and this can be one of the reasons for the unbalanced number of reviews. We suspect there is also a tradition of studying plant reproductive responses in a biotic context, and physiological and growth traits as influenced by abiotic environments (see Geber and Griffen 2003). These trends are reinforced by the recent boom of climate change studies, as reflected by the high number of CO_2 and UV radiation papers. The differential number of meta-analyses might then reflect a general bias in the plant literature. Louthan and Kay (2011), for example, also detected a bias towards abiotic-selected traits in a review of plant QTL mapping studies.

Strength of biotic and abiotic selection on fitness and non-fitness traits

Our approach to comparing the strength of selection on different types of traits differs from other review papers (Kingsolver et al. 2001; Rieseberg et al. 2002; Geber and Griffen 2003) in that we compare the results of multiple meta-analyses in a global, second order meta-analysis that includes thousands of results published in the literature. In addition, we do not focus on phenotypic selection as those articles, but on studies that control biotic or abiotic environmental variation and measure the resulting fitness and non-fitness responses. Because phenotypic selection studies do not formally measure environmental variation, such studies would not be appropriate to test our hypothesis. Still, we can compare our results on response variables to theirs. As opposed to those previous findings, we did not detect differences in the strength of responses to selection among different types of traits, either fitness or non-fitness. In contrast, when we looked for differences in the responses to selection elicited by biotic versus abiotic traits, we found a clear signal. Biotic-driven selection leads to stronger selection on traits in general when compared to abiotic selection pressures, at least for plants. It is possible that the biotic-abiotic comparison absorbs the differences between fitness and non-fitness traits detected in previous studies, as both variables are collinear in our database because of the biases described above.

The differential responses to biotic versus abiotic is a question that had been basically unexplored. The main exception is the recent study by Louthan and Kay (2011), who compiled mapping studies on plant traits and compared the direction and effect sizes of QTLs controlling biotic and abiotic-selected traits. Because they were not dealing with selection studies directly, but rather with the consequences of selection on the genetic architecture of traits, they classified traits a priori as putatively abiotic- or biotic-selected. Our study is the first that can confidently assign studies to the type of selection pressure. Our reviewed studies report more immediate responses and the results are therefore less historical than a QTL comparison, but in spite of the difference in approaches, the two studies found consistent results. Louthan and Kay (2011) found QTL's of larger effect associated with biotic-selected traits, and we found stronger observed responses of traits under biotic pressures. Both results are expected for traits that are under variable selective pressures, as can be the case for biotic selective agents. Biotic agents and interactions can vary strongly in space and time (Thompson 2005 and references therein), and consequently produce complex selective landscapes with multiple peaks or peaks that in turn shift in time and space. Such selective scenario can produce phenotypic responses that are stronger than under more subtle abiotic changes, and in turn select for QTLs of major effects.

To further explore the relative importance of biotic versus abiotic factors on plant character evolution it is clear that a higher diversity of studies would be very useful. In particular, case studies and meta-analyses in the under-represented categories (fitness traits under abiotic selection and non-fitness traits under biotic selection) would be very valuable. In addition, fully factorial case studies on the effects of biotic and abiotic pressures on both fitness and non-fitness traits in individual species are scant but potentially very informative.

Guide for future meta-analyses of meta-analysis

The broad use of formal meta-analytical techniques in plant ecology has undoubtedly contributed to our capacity for summarizing and extracting general results, based on the strength of combining many varied individual studies. We here take the next step of combining effect sizes of meta-analyses on diverse plant systems and traits in a second order meta-analysis. This approach is already frequently used in the health sciences, particularly to answer clinical questions, where for example different treatments for the same disease need to be compared but results are reported in independent reviews (Ioannidis 2009; Becker and Oxman 2011). Multiple-treatment meta-analysis is used to formally compare meta-analytic results in a network approach that incorporates direct and indirect comparisons of clinical treatments (Hasselblad 1998; Caldwell et al. 2010). Our analysis is a simplified version of such models.

One advantage of the approach of meta-analyzing meta-analyses is that it allows a high level of generalization using a very large number of individual case results already summarized in meta-analyses (in our case, more than 17,800) that would be very impractical to attempt with the original studies. Most meta-analyses, except perhaps the most recent ones, do not list each individual study case included and their associated effect size, sample size and variance, all required for a new meta-analysis based on the original studies. In a recent article that used published meta-analyses to find groups of papers on specific topics and extract individual study information (Barto and Rillig 2011), the authors report that they had to limit their analysis to a small fraction of the available publications, because few report the necessary data for each case study. In our case, using the original data would then imply going back to each case study and repeating the work done by meta-analytical studies. Another advantage of the second-order approach of using published meta-analyses compared to searching for original case studies is that meta-analyses are prepared by expert authors, who identify the relevant questions on each topic and the appropriate case studies to answer them. In a broad second-order meta-analysis like ours, such level of expertise is left to the original reviews.

Nonetheless, some aspects need to be considered carefully before combining review studies in second-order meta-analysis. First, it is possible that the same individual original studies are included in more than one of the meta-analytic publications available on a given topic. Our questions here were so broad and the number of individual studies on different topics so large, that it is unlikely that this form of pseudoreplication has affected our conclusions. Smaller and more focused meta-meta-analyses should probably be more concerned with excluding replicated results. Care should also be taken not to include meta-analyses that were not careful about another possible form of pseudoreplication, i.e., using the same case studies (and same experimental individuals) to conclude on different effects. Second, there are statistical problems with the conversion of effect sizes to a single common metric, as explained in the "Materials and methods" section. This can be a problem in ecological studies particularly, because a variety of effect sizes are commonly used and in particular response ratios, whose statistical properties have not been fully studied yet. Because of this problem, we had to limit our quantitative analysis to a single family of effect size metrics and exclude many potentially informative meta-analyses. Finally, future second-order meta-analyses addressing evolutionary issues should ideally include phylogenetic-informed effect sizes (Verdú and Traveset 2004), because of the ubiquity of phylogenetic signal in ecological traits (Blomberg et al. 2003).

Acknowledgments We thank J. Sánchez-Meca for comments and discussion on the use of meta-analysis of meta-analyses in ecology and other disciplines, and three reviewers for constructive comments on an earlier version of the manuscript. S. Nakagawa helped with the statistical analyses. This work was developed under the framework of projects VAMPIRO (CGL2008-05289-C02-01) and the European LinkTree project (BiodivERsA, EUI2008-03713). MCC was supported by a JAE-Doc CSIC scholarship.

References

Bailey JK, Schweitzer JA, Úbeda F, Koricheva J, LeRoy CJ, Madritch MD, Rehill BJ, Bangert RK, Fischer DG, Allan GJ, Whitham TG (2009) From genes to ecosystems: a synthesis of the effects of plant genetic factors across levels of organization. Philos T Roy Soc B 364(1523):1607–1616

Barto EK, Rillig MC (2011) Dissemination biases in ecology: effect sizes matter more than quality. Oikos. doi:10.1111/j.1600-0706.2011.19401.x

Becker LA, Oxman AD (2011) Chapter 22: overviews of reviews. In: Higgins JPT, Green S (eds) Cochrane handbook for systematic reviews of interventions version 5.1.0 (updated March 2011). The Cochrane collaboration. Available from http://www.cochrane-handbook.org

Bell G (2010) Fluctuating selection: the perpetual renewal of adaptation in variable environments. Phil Trans R Soc B 365(1537):87–97

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

Caldwell DM, Welton NJ, Ades AE (2010) Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. J Clin Epidemiol 63(8):875–882

Chamberlain SA, Holland JN (2009) Quantitative synthesis of context dependency in ant-plant protection mutualisms. Ecology 90(9):2384–2392

Geber MA, Griffen LR (2003) Inheritance and natural selection on functional traits. Int J Plant Sci 164(3 Suppl.):S21–S42

Grant PR, Grant BR (2002) Unpredictable evolution in a 30-year study of Darwin's finches. Science 296:707–711

Groenigen KJ, Osenberg CW, Hungate BA (2011) Increased soil emissions of potent greenhouse gases under increased atmospheric CO2. Nature 475:214–216

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 Sci 33(2):1

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(3):494–508

Hasselblad V (1998) Meta-analysis of multitreatment studies. Med Decis Making 18(1):37–43

Hereford J, Hansen TF, Houle D (2004) Comparing strengths of directional selection: how strong is strong? Evolution 58(10):2133–2143

Herrera CM, Castellanos MC, Medrano M (2006) Geographical context of floral evolution: towards an improved research programme in floral diversification. In: Harder LD, Barrett SCH (eds) Ecology and evolution of flowers. Oxford University Press, New York, pp 178–294

Higgins JPT, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. Stat Med 21(11):1539–1558

Higgins JPT, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. Br Med J 327:557–560

Hunter JE, Schmidt FL (2004) Methods of meta-analysis, 2nd edn. Sage, Thousand Oaks

Ioannidis JP (2009) Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. CMAJ 181(8):488–493

Kingsolver JG, Diamond SE (2011) Phenotypic selection in natural populations: what limits directional selection? Am Nat 177(3):346–357

Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hill CE, Hoang A, Gibert P, Beerli P (2001) The strength of phenotypic selection in natural populations. Am Nat 157(3):245–261

Lajeunesse MJ, Forbes MR (2003) Variable reporting and quantitative reviews: a comparison of three meta-analytical techniques. Ecol Lett 6:448–454

Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. Evolution 37:1210–1226

Linhart YB, Grant MC (1996) Evolutionary significance of local genetic differentiation in plants. Annu Rev Ecol Syst 27:237–277

Louthan AM, Kay KM (2011) Comparing the adaptive landscape across trait types: larger QTL effect size in traits under biotic selection. BMC Evol Biol 11:60

Marín-Martínez F, Sánchez-Meca J (2010) Weighting by inverse variance or by sample size in random-effects meta-analysis. Educ Psychol Measur 70:56–73

Merilä J, Sheldon BC (1999) Genetic architecture of fitness and nonfitness traits: empirical patterns and development of ideas. Heredity 83:103–109

Morrissey M, Hadfield J (2012) Directional selection in temporally replicated studies is remarkably consistent. Evolution. doi:10.1111/j.1558-5646.2011.01444.x

Nakagawa S, Cuthill IC (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev 82(4):591–605

Nakagawa S, Santos ESA (2012) Methodological issues and advances in biological meta-analysis. Evol Ecol. doi:10.1007/s10682-012-9555-5

Rieseberg LH, Widmer A, Arntz AM, Burke JM (2002) Directional selection is the primary cause of phenotypic diversification. Proc Natl Acad Sci USA 99(19):12242–12245

Schluter D (2000) The ecology of adaptive radiation. Oxford series in ecology and evolution. Oxford University Press, New York

Siepielski AM, DiBattista JD, Carlson SM (2009) It's about time: the temporal dynamics of phenotypic selection in the wild. Ecol Lett 12(11):1261–1276

Thompson JN (2005) The geographic mosaic of coevolution. University of Chicago Press, Chicago

Trager MD, Bhotika S, Hostetler JA, Andrade GV, Rodriguez-Cabal MA, McKeon CS, Osenberg CW, Bolker BM (2010) Benefits for plants in ant-plant protective mutualisms: a meta-analysis. PLoS One 5(12):e14308

Verdú 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