

## COMMENTARY

**On the utility of meta-analyses in the study of natural selection**

R. CALSBEEK &amp; M. A. MCPEEK

*Department of Biological Sciences, Dartmouth College, Hanover, NH, USA*

The history of attempting to measure natural selection on quantitative characters goes back nearly 120 years (Bumpus, 1899). On a wintery morning in February, 1898, D.H. Bumpus came upon 136 English sparrows knocked to the ground by a late season storm. Gathering them up and returning to his laboratory, he measured a handful of characters from these birds (64 of which subsequently perished) and attempted to differentiate the living and dead based on variation in phenotype. His conclusion? Selection tended to cull individuals most divergent from the average 'type'.

Seventy years passed before Lande (1979) and Lande & Arnold (1983) provided a statistically rigorous framework for estimating the strength and form of selection in the wild. The observation that selection gradients taken from traditional general linear models could be used to accurately estimate natural selection revolutionized the way evolutionary biologists thought about and measured selection. As a result, the number of studies reporting selection gradients has exploded in recent decades.

With the trove of data filling the scientific literature, researchers have begun asking various derived questions about the intensity and form of selection in the wild: How strong is selection in general? How does selection vary through time? How does selection resolve conflicts between the sexes? How does the environment influence evolution? Several recent papers have attempted to summarize the existing literature on natural and sexual selection to address these questions: typical strength of selection (Endler, 1986; Hoekstra *et al.*, 2001; Kingsolver *et al.*, 2001; Hereford *et al.*, 2004), temporal variability (Siepielski *et al.*, 2009) and sex specificity (Cox & Calsbeek, 2009). In Morrissey's (2016) study, these 'informal meta-analyses' are re-analysed using a novel mixed-model meta-analytical approach.

Morrissey explores three flavours of meta-analyses: informal, formal and a Bayesian mixed-model approach that is the major contribution of his paper. Informal meta-analyses typically do not account for the varied methods of collecting or variability associated with the observed data. Depending on the nature of the question

being asked, informal meta-analyses can range from perfectly reasonable to highly biased. In particular, the well-understood problem of upward bias that arises from summarizing mean absolute values makes informal meta-analyses problematic. Formal meta-analyses include some accounting for error in individual estimates of a quantity. Formal meta-analyses may therefore be preferable to informal meta-analyses but themselves suffer from other problems. In particular, formal meta-analyses tend to do well at minimizing error (i.e. improving precision) but do not necessarily get us any closer to knowing a true mean value. Morrissey offers a Bayesian mixed-model approach to measure some meta-quantity while accounting for sampling error. This new method he argues is free from the biases introduced by previous formal and informal meta-analyses. So if we wish to know the mean strength of selection in the wild, Morrissey suggests that this Bayesian mixed-model approach provides a less biased estimate than traditional methods.

When applied to the previously published studies cited above, Morrissey concludes that selection is not as strong on average as we thought, not as variable as we thought, not as sexually antagonistic as we thought, and that reaction norms evolve not by variation in shape but by differences in mean trait values across environments. Our initial reaction to these conclusions is that one should be unsurprised by the revision of results. The sample sizes used in most selection studies are small, at least relative to those needed for precise estimates of linear selection (not to mention quadratic selection estimation: (Brodie, 1992; Blows & Brooks, 2003)). Thus, error terms around the value of published values of selection tend to be large. Traditional meta-analyses may therefore provide a biased summary of how selection operates (on average) in the wild. Moreover, even the best of parameter estimates from meta-analyses are probably inflated, because studies in the literature are likely to be heavily influenced by publication bias (Hersch & Phillips, 2004). Researchers that attempt to measure selection only to find weak or undetectable relationships between fitness and phenotype are probably less likely to submit those results for publication or to have those results accepted for publication when submitted.

With that said, we pose the question: Does it matter? Previous reviews of selection (e.g. (Kingsolver *et al.*, 2001)) are interesting in our estimation, not because they provide an average strength of selection, but because they provide some insight into the range of values that selection gradients may take (see e.g. figure 3 in Kingsolver *et al.* (2001); figure 3 in Siepielski *et al.* (2009)). Although the mean value of selection in nature may be of passing interest, we should point out that any measure of selection is only relevant to that population in which it is measured and only for that time period during which it occurred. Calculating mean

*Correspondence:* Ryan Calsbeek, Department of Biological Sciences, Dartmouth College, Hanover, NH 03755, USA.  
Tel.: +603 646 9917; fax: +603 646 1345;  
e-mail: ryan.calsbeek@dartmouth.edu

values of selection over time, through space, or among species ignores this salient point. To see that this is true, one need only to consider the change to values of genetic variance and covariance, so central to the predicted evolutionary response to selection, that occurs each generation. The dynamic properties of quantitative genetic parameters make long-term future predictions of evolutionary change a fool's errand (Morrissey *et al.*, 2010).

A second problem related to the first is that differences in the shape or intensity of selection as it acts among populations are often the most interesting facets of these studies. Indeed, the most powerful insights regarding natural selection usually come from studies (whether experimental or observational) that compare differences in the agent of selection. However, these interesting ecological differences that underlie variation in selection are erased when selection coefficients are pooled into mean values. This is exemplified by Calsbeek & Cox (2010), who manipulated whole-island populations of lizards to study the importance of predation and competition as agents of natural selection. Calsbeek and Cox showed that the strength of selection was strongly tied to variation in population density (a proxy for competition intensity) but not with predation intensity. The ostensibly interesting result that natural selection on characters like body size and running stamina tended to be strong and positive at high population density but weakly negative at low population density would be completely missed by a study-wide averaging of selection gradients (which would indicate a net selection coefficient of approximately zero).

Finally, the diversity and heterogeneity of data represented among measures of natural selection are so great that the application of any meta-analytic methods, no matter how rigorous, may be problematic. Estimates of selection differentials and gradients are made on a huge diversity of phenotypic traits using a large set of proxies for fitness. Thus, the underlying statistical distributions from which these measures are drawn are surely not the same. Moreover, even for those that are drawn from the same distribution, the underlying parameters will not be the same. For example, we simulated various data sets of selection differentials and gradients by repeatedly applying directional selection to a population of 1000 individuals with a specified phenotypic variance and strength of selection. We found that the variance in the selection differentials and gradients changed as the initial phenotypic variance in the population changed. Thus, even differences in phenotypic variance among populations from which estimates are made cause violations of the assumptions of meta-analysis methods that Morrissey derives.

There is clearly value in a meta-analytical approach. As the scientific literature burgeons, the utility of summarizing major patterns in ecology and evolution will also grow. The mixed-model approach developed

by Morrissey offers a less biased means of estimating parameters using meta-analyses. However, distilling the multifarious nature of natural selection to a few summary statistic values seems uninformative. We echo Morrissey's (2016) final paragraph that real understanding of the operation of natural selection will come from synthesizing the similarities and the differences among studies on different traits, different measures of fitness and different taxa experiencing different ecologies. In the case of selection gradient studies, the meta-whole may be less than the sum of its parts.

## References

- Blows, M.W. & Brooks, R. 2003. Measuring nonlinear selection. *Am. Nat.* **162**: 815–820.
- Brodie, E.D. 1992. Correlational selection for color pattern and antipredator behavior in the garter snake *Thamnophis ordinoides*. *Evolution* **46**: 1284–1298.
- Bumpus, H.C. 1899. The elimination of the unfit as illustrated by the introduced sparrow, *Passer domesticus*. *Biol. Lectures, Woods Hole Marine Biol. Station* **6**: 209–226.
- Calsbeek, R. & Cox, R.M. 2010. Experimentally assessing the relative importance of predation and competition as agents of selection. *Nature* **465**: 613–616.
- Cox, R.M. & Calsbeek, R. 2009. Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *Am. Nat.* **173**: 176–187.
- Endler, J.A. 1986. *Natural Selection in the Wild*. Princeton University Press, Princeton, NJ.
- Hereford, J., Hansen, T.F. & Houle, D. 2004. Comparing strengths of directional selection: how strong is strong? *Evolution* **58**: 2133–2143.
- Hersch, E. & Phillips, P.C. 2004. Power and potential bias in field studies of natural selection. *Evolution* **58**: 479–485.
- Hoekstra, H.E., Hoekstra, J.M., Berrigan, D., Vignieri, S.N., Hoang, A., Hill, C.E. *et al.* 2001. Strength and tempo of directional selection in the wild. *Proc. Natl. Acad. Sci. USA* **98**: 9157–9160.
- Kingsolver, J.G.H., Hoekstra, H.E., Hoekstra, J.M., Berrigan, D., Vignieri, S.N., Hill, C.E. *et al.* 2001. The strength of phenotypic selection in natural populations. *Am. Nat.* **157**: 245–261.
- Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* **33**: 402–416.
- Lande, R. & Arnold, S.J. 1983. The measurement of selection on correlated characters. *Evolution* **37**: 1210–1226.
- Morrissey, M.B. 2016. Meta-analysis of magnitudes, differences, and variation in evolutionary parameters. *J. Evol. Biol.* **29**: 1882–1904.
- Morrissey, M.B., Kruuk, L.E.B. & Wilson, A.J. 2010. The danger of applying the breeder's equation in observational studies of natural populations. *J. Evol. Biol.* **23**: 2277–2288.
- Siepielski, A.M., DiBattista, J.D. & Carlson, S.M. 2009. It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecol. Lett.* **12**: 1261–1276.

Received 26 March 2016; accepted 31 March 2016