NEW PERSPECTIVES

Body size and fitness in Odonata, stabilising selection and a meta-analysis too far?

DAVID J. THOMPSON and OLA M. FINCKE. Population and Evolutionary Biology Research Group, School of Biological Sciences, University of Liverpool, U.K. and Department of Zoology, University of Oklahoma, U.S.A.

Key words. Damselflies, dragonflies, fitness components, meta-analysis, size effects, stabilising selection.

Introduction

Sokolovska et al. (2000) used 57 weighted correlation coefficients derived from 33 published studies of 20 species to produce the first meta-analysis of body size and fitness in odonates. They concluded that 'there is a general fitness benefit to large size in odonates'. Given the inherent biases of their analyses, such a conclusion is hardly surprising. In this paper, their study was used to illustrate major problems that arise from (1) ignoring the possibility of stabilising selection on body size, (2) using statistics from the literature indiscriminately, when the original studies are inappropriate for comparative analysis, and (3) relying heavily on transformed statistics derived from small samples. Suggestions are made as to how data can be presented and analysed in ways conducive to future comparative analyses on size relations to fitness in odonates and other insects.

Critique

Stabilising selection

There is a simplistic view, not uncommon in behavioural ecology, that bigger is always better. If this were the case, animals might become larger and larger over ecological time, subject to phylogenetic constraints. Given the lack of increase in size over time, it seems that sexual or natural selection favouring large size is nearly always balanced by some advantages to small size.

Stabilising selection can be inferred from the persistence of the status quo in body size, providing an a priori reason to suspect that stabilising selection is the most common selection acting on natural populations. To a naive reader, the behavioural ecology literature, on the other hand, suggests that directional selection for size is the norm, especially for species in which large size provides greater mating success to territorial males and greater fecundity to females. The focus of this paper is on factors contributing to a myopia best illustrated by reviews on body size and fitness. For example, in a selection of 186 species 'in which sexual selection has been studied quantitatively in relation to some character', Andersson (1994) found that large size was advantageous to male fitness in 51 cases and to female fitness in 27 cases. Notably absent, however, were quantitative studies indicating a small size advantage or those demonstrating stabilising selection with respect to size. Although Andersson conceded that in some cases sexual selection 'may be self-limiting', stabilising selection on size in relation to fitness has rarely been considered (e.g. Thornhill & Alcock, 1983; Arnqvist et al., 1996; Sokolovska et al., 2000).

Stabilising selection explicitly contradicts the assumption that fitness is a linear function of size, yet this type of selection is often the most difficult to document because very large sample sizes are needed to detect it. To date, for example, stabilising selection on body size with respect to measures of lifetime mating success has been demonstrated only for three species of insect, all odonates (Fincke, 1982, 1988; Banks & Thompson, 1985; Stoks, 2000). The sample size of these studies ranged from 186 to 489. Too often, authors publish only a correlation coefficient for a fitness variable and size, without even testing for a non-linear fitness function or even providing a scatter plot of the data. Examination of the data might suggest stabilising selection, even though a non-linear function is not statistically significant. For example, significant stabilising selection on body size with respect to the number of total mates was found for 489 male Enallagma hageni but not for 298 females, though the trend was in that direction (Fig. 1; for similar trends, see Cordero, 1995; Anholt, 1997). Yet another problem is the paucity of studies designed to identify selective constraints acting on body size. A major value of measuring fitness...
correlates over the lifetime of individuals is that such studies provide the data needed to identify potential selective trade-offs. The most commonly suggested constraints on body size are trade-offs between natural and sexual selection acting on adults (Banks & Thompson, 1985; Fincke, 1988; Thompson, 1989; Fig. 2a,b), though they may also occur during the adult maturation stage (Anholt et al., 1991). Larval ecology, though more difficult to study, is equally likely to constrain adult size (Fincke, 1992; Partridge & Fowler, 1993; Johansson & Rowe, 1999; Fig. 2c).

The first synthetic review of the effect of male body size on fitness estimates in Odonata relied conservatively on 11 species for which measures of lifetime reproductive success were available (Fincke et al., 1997). A simple comparative analysis revealed only a weak trend ($P = 0.12$) for a large male mating advantage in territorial species compared with non-territorial species. There was a stronger trend ($P = 0.01$) for large body size to provide a mating advantage to males in species in which males were larger than females, than in species in which females were larger (i.e. the most common case for Odonata). Mechanistic studies of odonates suggested that for territorial males, which compete during flight, manoeuvrability and endurance were more advantageous for mating efficiency rather than large size per se. The authors called for more long-term studies on under-represented taxa and greater attention to how females affect male mating success and how larval ecology acts as a selective pressure on male and female size and behaviour.

**Biases inherent in the analysis**

Sokolovska et al. (2000) began with the assumption that fitness is a linear function of body size by relying entirely on correlation coefficients between body size and fitness estimates (i.e. mating rates, longevity, fecundity). Hence, even though stabilising selection was demonstrated for some of the studies used, the advantage of intermediate-sized males translated simply to a low effect size; no methods were described that...
might detect a general trend for stabilising selection. Despite a priori reasons why the typical size advantages may be less pervasive for odonates engaging in aerial combat, one-tailed tests were used, making the size bias explicit.

Selection of studies

Inclusion of at least 11 studies is questionable because the data were not comparable or were used for the wrong purpose. Among these are six studies reported by Sokolovska et al. (2000) to be studies of lifetime mating success, which should be among the most compelling examples used. Gibbin and Thompson (1991) indicated explicitly that their data were not appropriate to use for comparisons of lifetime mating success because the ponds at which the study was undertaken were not sufficiently isolated to guarantee that all reproductive activity took place at those ponds. Van Buskirk (1987) also stated explicitly that he was unable to measure lifetime mating success because the males were long-lived and the study ended while many were still alive. Marden (1989) used 30 males to look at the effect of the ratio of flight-muscle mass to body mass. At the end of each observation period, the males were captured, dissected, and weighed. While this procedure ensured that they did not achieve any more matings, it is not what is meant conventionally by a lifetime mating success study. Nevertheless Sokolovska et al. entered all three sets of data under the lifetime mating success heading. Cordoba-Aguilar’s (1995) study of Hetaeria cruenta also is cited as an example of a lifetime mating success study, something that the author had, quite rightly, not claimed. The data were extracted from a sample of 13 males that Cordoba-Aguilar regarded as resident. The part of the study used by Sokolovska et al. (2000) from Michiels and Dhont (1991) was also not a lifetime mating success study. Sokolovska et al. extracted two estimates of lifetime mating success from Koenig (1991). This paper involved a re-analysis of field data collected in 1984 and 1986, some of which had been published by Koenig and Albano (1987) and from which another estimate of lifetime mating success had already been extracted.

In addition to the six studies cited above that were wrongly included under the heading of lifetime mating success studies, five other studies could not be viewed as examining fitness of mature odonates under natural conditions. In the work described by Forbes and Baker (1990) on Enallagma ebrium, predators were excluded and the 20 animals used were allowed to starve to death. They were newly emerged animals, not the mature adults described in the paper’s title, so this work should have been excluded on two counts. In one study of Pyrrhosoma nympha (Harvey & Corbet, 1985), the r statistic was derived from a comparison between sexually immature males and mated males. Another study on Pyrrhosoma that provided the comparable data showing that there was no difference in the sizes of the winners and losers of territorial disputes in mature Pyrrhosoma (Gibbin & Thompson, 1991) was used (incorrectly) for a different purpose. Most studies of territorial defence compared the size of territory owners with those that were not able to hold a territory, however two sources included in the territorial defence variable (i.e. Fincke, 1992; Cordoba-Aguilar, 1995) compared males using different, size-dependent alternative strategies, which is not comparable to successful and unsuccessful males using the same territorial strategy. In Conrad’s (1992) study of Argia viva, the data extracted by Sokolovska et al. appear to combine data from both of the sites used by Conrad. This is unfortunate because at one of these sites, Argia has a 3-year life cycle and at the other, a geothermal spring, a 1-year life cycle. Males were smaller in general at the hot springs site, so the result calculated by Sokolovska et al. depended simply on the sample sizes collected by Conrad at each site. The second mating rate study of Forbes (1991) is treated in a similar manner by combining data from 2 years.

Omissions

It would be churlish to criticise reviewers for omitting studies that they might have included, however there are several instances in which better data than those included in the review were ignored in studies that were cited. For example, there are only two studies from which clutch size data have been extracted (Koenig & Albano, 1987; Gibbin & Thompson, 1990). In the latter, clutch size is the number of mature eggs that would have been laid on a visit to the pond, whereas in the former it is measured in terms of time (number of seconds spent ovipositing). Curiously, the most detailed study on the relationship between skeletal size and clutch size (Banks & Thompson, 1987), in which a rich pattern of relationships that depended on interclutch interval was described, was ignored. In this study, there was a negative correlation between clutch size and body size for 1-day clutches ($r = -0.422$, $n = 25$, $P < 0.05$) and no significant correlation for 2-day clutches ($r = 0.167$, $n = 11$, $P = NS$). Interclutch interval depends on weather, so in good summers small females will lay larger clutches but this relationship breaks down in summers that include a higher proportion of days unsuitable for reproductive activity (Thompson, 1990).

Biases from errors in citing data and selective use of statistics

Of the 57 sets of statistics presented in table A1 by Sokolovska et al. (2000), there were 20 incorrect citings (four errors in r-value, eight errors in n, eight errors in assigning fitness correlates; corrected version of statistics in table A1 are available on request). Unfortunately, the errors of the greatest magnitude were those that biased the outcome towards finding a positive effect of body size. For example, in Orthetrum japonicum (Kasuya et al., 1997a), the authors extracted only the positive size effect, between body width and territorial defence, while they ignored the insignificant effect of wing length. Indeed, in another study of the same species, Kasuya et al. (1987)
reported a negative correlation (−0.205) between territory duration and wing length, and discussed why large size did not affect male mating success significantly (see also Kasuya et al., 1997b). The sample size for *Orthetrum chrysostigma* (Miller, 1983) was 11 rather than 77, and 79 for *Pyrrhosoma nymphula*, not 155 as cited. Both errors increase the effect size. The published mating rate statistics were omitted for *Enallagma hageni* (*r* = 0.02, *n* = 489) and for *Megaloprepus coerulatus* (*r* = 0.33, *n* = 33). The highest *r* for territorial defence of the analysis was 0.976 from *M. coerulatus*, however this was not a measure of territorial defence but rather the most exaggerated difference in size that could be chosen, between territorial and satellite males. Anholt (1991) stated that ‘Combining males from both years, a regression of mass against relative fitness was significant (*t* = 3.97, d.f. = 69, *P* = 0.0002) ... with light males getting the most matings’, but this was recorded as a sample size of 78 and a population correlation coefficient of 0.000. The correlation coefficient calculated by Sokolovska et al. from Conrad (1992) was 0.764 but in three of Conrad’s four samples, males that obtained matings were actually smaller than those that remained unmounted, and the overall mean size of mated males was smaller than that of unmounted males.

**Biases arising from small sample sizes and inflated *n* values**

Lower sample sizes are more likely to lead to a significant positive correlation with body size for a number of reasons. First, there was a strong, negative correlation between the value of *r* and the sample [data from table A1 in Sokolovska et al. (2000); Fig. 3]. The correlation becomes insignificant only for sample sizes ≥ 150 (*r* = −0.26, *P* = 0.21, *n* = 24) or ≤ 25 (*r* = −0.17, *P* = 0.75, *n* = 6). No negative correlations were found for any study with a sample size < 50. Few of the small samples exhibited a significant size effect. In the case of *M. coerulatus*, multiple statistics for the effect were provided, and the effect was magnified artificially in the smallest sample. For short-term studies, time constraints usually preclude equal sampling of sites and, given a choice, a researcher will naturally choose the most active sites (e.g. Tsubaki & Ono, 1987). In the case of *M. coerulatus*, more prolonged sampling indicated that smaller males can get matings before being displaced by larger males, whereas very small males can mate by playing an alternative strategy. In a later study of the same population, the *r* value between size and mating success (controlling for seasonal effect on size, partial *r* = 0.15, *P* < 0.05, *n* = 263; Fincke & Hadrys, 2001) was as low as the largest sample used in the meta-analysis (*r* = 0.152, *n* = 206).

Sokolovska et al. (2000) treated multiple populations of the same species as independent measures. Their graphs (Fig. 1a–d) included from two to six repeated measures; similarly, the sample sizes in table A1 are inflated by two to five repeated measures on the same species. When the sample sizes of those repeated measures are small, as in the meta-analysis, non-independent samples create a real bias. Even though population averages were weighted by sample size, the use of multiple statistics with small sample sizes biased the outcome towards a size effect. Taking *M. coerulatus* as an example, the two statistics used for mating rate were 0.134, *n* = 206 and 0.332, *n* = 33 (used incorrectly for mating rate). A weighted mean would be 0.18 whereas using only the statistic with the greatest sample size would have been a more conservative approach, given the extremely strong negative correlation between *r* and *n* (Fig. 3).

**Biases from insignificant *t* values coupled with the assumption of linear fitness functions**

Meta-analysis has been lauded as a way to maximise statistical power and uncover significant trends from multiple studies, the small sample sizes of which may preclude documentation of a significant effect individually

---


(Arnvist & Wooster, 1995). While true, given that the method reduces even significant, non-linear fitness functions to linear trends, the validity of the conclusions drawn is questionable. Of the 57 statistics used by Sokolovska et al. (2000), 35 were not significant. Of the 20 species on which their analyses were based, only half were represented by a study of more than 150 individuals. Hence, at best only half were suitable for detecting stabilising selection, and most researchers did not test for stabilising selection. Small samples may result more often in a positive rather than a negative correlation if small individuals have a greater chance of being sampled than large individuals. This might often result from only partial sampling of the flight season, for species in which body size declines seasonally, as is often the case (e.g. Banks & Thompson, 1985; Fincke, 1988; Michiels & Dhont, 1989; Fincke & Hadrys, 2001). Some authors corrected for decline in body size with time by calculating partial correlation coefficients (e.g. Banks & Thompson, 1985; Fincke, 1988; Fincke & Hadrys, 2001) while others made no such correction. Sokolovska et al. (2000) included the statistics provided by the original authors, whether they were strictly comparable or not.

**Phylogenetic and related problems**

Sokolovska et al. (2000) point out in their conclusions that a phylogenetically controlled analysis of effect sizes would allow determination of whether effect sizes depend, in part, on membership of a given clade. They also pointed out that the data set, and thus their analysis, is biased strongly towards one family, the Coenagrionidae. They are right but they have understated the extent of the bias. It is currently agreed that there are 10 families of Anisoptera (dragonflies) from which data were used from just one, the Libellulidae. Nineteen studies of libellulids are included in the analysis but nine of these (47%) come from one species, Plathemis lydia (five from one paper; Koenig & Albano, 1987). Thus nearly half of the analysis for the Sub-Order Anisoptera (approximately 3000 species) comes from studies of this one species. There are 22 families of Zygoptera (damselflies) from which data were used from four: Lestidae, Coenagrionidae, Pseudostigmatidae, and Calopterygidae. Coenagrionids are particularly well represented, with 27 out of 38 studies. As with the Libellulidae, some species are represented on several occasions (e.g. Coenagrion puella, Ischnura gralliisi). The authors make the wholly laudable call for more studies that relate fitness to body size in odonates, presumably from a wider range of taxa, however it seems unreasonable to blame the data for the narrowness of taxa available on which to perform this meta-analysis, and suggests that the data were simply not sufficient for the authors to assess size effects across the Odonata.

**Conclusions**

Of the eight species for which studies provide comparable data on lifetime mating success, only five reported statistics for both longevity and mating rates, the two major fitness components comprising lifetime mating success. Hence, only five of the species listed in table A1 (Sokolovska et al., 2000) could be used correctly to address the question of possible trade-offs between longevity and mating rate with respect to size (though the meta-analysis missed the published statistic for mating rate in *E. hageni*). Of these, only Moore (1990) reported a significant, unambiguous effect of body size. Fincke (1982) and Banks and Thompson (1985) reported stabilising selection for size, and Koenig (1991) found no significant effect of size on any fitness correlate (see also Fincke et al., 1997). Cordero (1995) reported a size effect in only one of two populations and, even then, the data appear to be bell-shaped, suggestive of stabilising selection. Given the multiple weaknesses of the analysis, the conclusion of Sokolovska et al. (2000) that 'large size is associated with increased lifetime reproductive success in odonates of both sexes' is at best overstated and at worst misleading. Indeed, as a taxon, the Odonata seem to be an exception to any trend for selection on large size; the advantage of small size for aerial manoeuvrability probably counters any large size advantage in many species (Fincke et al., 1997). Future meta-analyses may be used fruitfully to determine trends with respect to selection for size and phylogeny, mating systems, or the degree and direction of sexual size dimorphism, however the application of the technique demands a knowledgeable and critical reading of the studies in the literature. It remains unclear whether even a meta-analysis can handle the dual possibility of stabilising and directional selection acting within a taxon.

Finally, lifetime mating success should not be equated with lifetime reproductive success. The latter usually refers to the number of eggs fertilised by males or laid by females. The use of either measure as a fitness correlate remains untested for most species. For *Megaloprepus coerulatus*, a highly territorial species in which males are under strong sexual selection on body size, neither male fertilisation success nor female clutch size was a good predictor of realised fitness among mated individuals (Fincke & Hadrys, 2001). In both sexes, larval ecology constrains selection on adult size.

Future studies designed to identify contrasting selective pressures acting on body size in insects at both the adult and immature stages of their life history are critical to clarify the true picture of how body size affects fitness in organisms with a complex life cycle. As pointed out by Sokolovska et al. (2000), far more emphasis is needed on the larval stage of most odonates. After all, any selection on adult size of odonates begins when they are still in the water.

**Acknowledgements**

We thank M. Kaspari for discussion, and A. Cordero, I. Harvey and E. Kasuya for comments on the interpretation of their work.
References


Accepted 15 October 2001