# Pollinator size and its consequences: Predictive allometry for pollinating insects

Liam K. Kendall1 2, Romina Rader1, Vesna Gagic2, Daniel Cariveau3, Matthias Albrecht4, Katherine C.R. Baldock5, Breno Freitus, Mark Hall1, Andrea Holzschuh6, Janeally Pereira, Juanita Rodriguez7, Laura Russo8, Louis Suter4, Nicolas J. Vereecken9 and Ignasi Bartomeus10

1. School of Environmental and Rural Sciences, University of New England, Armidale, NSW 2351, Australia

2. CSIRO Agriculture, GPO Box 2583, Brisbane, QLD 4001, Australia

3. Department of Entomology, University of Minnesota, Minneapolis, MN, USA

4. Eidgenössisches Departement für Wirtschaft, Agroscope Agrarökologie und Umwelt, CH-8046 Zürich, Switzerland

5. School of Biological Sciences & Cabot Institute, University of Bristol, Bristol, BS8 1TQ, UK

6. Animal Ecology and Tropical Biology Biocenter, University of Würzburg, 97074 Würzburg, Germany

7. Australian National Insect Collection, CSIRO, Canberra, ACT 2601, Australia

8. Botany Department, Trinity College Dublin, Ireland

9. ULB, Brussels, Belgium

10. Dpto. Ecología Integrativa, Estación Biológica de Doñana (EBD-CSIC), 41092 Sevilla, Spain

Corresponding author: Liam Kendall, Ecosystem Management, Building W55, School of Environmental and Rural Sciences, University of New England, Armidale 2351, NSW, Australia. Email: [lkendal2@myune.edu.au](mailto:lkendal2@myune.edu.au)

**Abstract**

Pollinator insect body size can predict influential ecological traits such as metabolic rates and foraging distances, however measuring body size can be complicated. Allometric scaling laws have been used to overcome this problem yet available allometric models are outdated, rely upon geographically restricted sampling and have limited applicability for non-bee taxa. More accurate predictions of pollinator body size require dynamic models that consider biogeography, intraspecific variation and phylogenetic relatedness within an iterative, updatable framework. We improved upon pre-existing equations for estimating body size in two key pollinating taxa (bees and hoverflies). We measured dry weight and intertegular distance (ITD) of 298 bee species and 103 hoverfly species across three biogeographic regions: Australia, Europe and USA. We test the power of ITD alone and in interaction with region, sex and taxonomy or phylogeny to predict interspecific pollinator body size and assess intraspecific ITD - body size relationship for five species per taxa. Overall differences between these models were minimal with PGLS models performing similarly to LME models. Intraspecific models found ITD an unreliable predictor of body size for bees (*R2*: 0.02 – 0.46) and hoverflies (*R2*: -0.11 – 0.44). These highly applicable models form the basis of the dynamic R package, '*pollimetry*’, and provide an updated resource for allometric research concerning wild and managed pollinators globally.

**Keywords**: allometry, Apoidea, body size, pollination, Syrphidae

**Introduction**

Body size is an intrinsic trait of all organisms that influences key patterns across all levels of biological organisation. Adult body size variation (both intra- and interspecific) in insects is the outcome of natural selection affecting physiological and biochemical processes during ontogeny (see Chown & Gaston 2010’s review on body size variation). Therefore, body size is central to physiological (e.g. metabolic and growth rates (Angilletta et al. 2004; Ehnes et al. 2011; Harrison et al. 2014)), life history (e.g. life span, reproductive rate and type (i.e. capital or income breeders) (Speakman 2005; Teder et al. 2008)) and ecological attributes (e.g. species abundance and richness, trophic interactions, geographic range size and dispersal ability) (Brown et al. 2004; White et al. 2007; Chown & Gaston 2010, Rall et al. 2011; Stevens et al. 2012; Velghe & Gregory-Eaves 2013; DeLong et al. 2015). These effects lead to differing spatial and temporal size-frequency distributions within populations and communities as well as drive key ecosystem functions and services such as decomposition, carbon cycling, predation (REF!), primary productivity and pollination (Woodward & Hildrew 2002; Greenleaf et al. 2007; Rudolf & Rasmussen 2013; Schramski et al. 2015).

Studies of body size variation utilise allometric theory. Gould (1966) defined allometry as the ‘study of size and its consequences.’ Allometric scaling laws refer to how traits, which can be morphological, physiological or chemical, co-vary with an organism’s body size, often with important ecological and evolutionary implications (Gould 1966; Huxley 1993). However, direct measurements of body size, traits and inferred allometric relationships can be impractical for a number of reasons. First, direct measurements can be time consuming and require destructive methods, which are unfeasible for museum specimens and threatened species (Rogers et al. 1977; Henschel & Seely 1997). Secondly, in diet/food web studies, body size estimates come from digested prey items (e.g. Hodar 1997). Thirdly, a lack of life-history information, especially for ecologically cryptic and rare species, may not be known. As such, predictive allometry, which attempts to estimate body size or a hypothesised allometric characteristic using a co-varying trait, has emerged across many biological disciplines.

Most commonly, body length has been used extensively to predict body size, such as in fish (e.g. Karachle & Stergiou 2012), mammals (e.g. Trites & Pauly 1998) and both aquatic (e.g. Burgherr & Meyer 1997; Benke et al. 1999) and terrestrial invertebrates (e.g. Rogers et al. 1977; Sample et al. 1993; Sabo et al. 2002). These models often show considerable support (*R2* > 0.9), which has led to the proliferation of multiple models for a wide range of taxa worldwide, especially insects. However, when compared, these models show considerably different coefficients both within- and between insect orders (Schoener 1980; Sample et al. 1993; Ganihar 1997; Benke et al. 1999; Brady & Noske 2006), often due to biogeography (i.e. latitude, see Martin et al. 2014), and/or methodological influences such as sampling biases (e.g. range of body sizes sampled(Sage 1982)).

A number of key pollination traits exhibit allometric scaling. In bees, body size affects insect activity rates/periods (Streinzer et al. 2016), pollen load (e.g. Ramalho et al. 1998), foraging distance (e.g. Greenleaf et al. 2007; van Nieuwstadt & Iraheta 1996), proboscis length (Cariveau et al. 2016) and plant-pollinator trait matching (Stang et al 2009; Olesen et al., 2010). Despite the ubiquity of pollination-related allometric traits, few predictive models exist for pollinating insects, with one notable exception. Cane (1987) established a predictive allometric model for bee body size as a function of the intertegular distance (ITD) (the distance between the wing-attachment points on either side of the thorax). Cane (1987)’s pioneering model was developed with a sample of females from 20 North American solitary bee species that represented six major bee families. It is now the most commonly used metric for estimating bee body size and has used in ecological (e.g. Williams et al. 2010), sensory (e.g. Spaethe & Chittka 2003; Kapustjanskij et al. 2007) and behavioural studies (e.g. Oliveira & Schlindwein 2010). It also firmly developed ITD as an important body size proxy for establishing other ecologically important allometric relationships (e.g. foraging distances and bee proboscis length; Greenleaf et al. 2007; Cariveau et al. 2016).

. Iterative ecological forecasting, whereby equations are periodically updated as well as tested against novel data (e.g. Dietze et al. 2018; Harris et al. 2018) represent an as-yet untested avenue and overarching framework for greater accuracy and wider applicability of predictive allometry. Critical to this framework in the case of predictive allometric models are considerations of model choice, development and validation. Traditional predictive allometry has typically utilised ordinary least squares or major axis regression. Although both have appropriate statistical properties for allometric prediction (Warton et al. 2006; Legendre & Legendre 2012), phylogenetic model structures such as phylogenetic generalized least squares (PGLS) or phylogenetic generalised linear mixed models (Harvey & Pagel 1991; Hadfield and Nakagawa 2010) represent an attractive alternative, especially as body size has been repeatedly linked to phylogenetic signal, compelling allometric studies to incorporate species topology (Garland and Ives 2000; Blomberg et al. 2003). Model validation techniques, such as cross-validation, which are common-place in statistics (Stone 1974), have been overlooked in predictive ecology, except within species distribution and resource selection modelling (e.g. Boyce et al. 2002; Wenger & Olden 2012; Boria et al 2014).

The utility of Cane’s equation has not previously been tested beyond North American solitary bee species except in bumblebees (Hagen & Dupont 2013). Further, it hasn’t been tested in conjunction with biogeography, within more complex model structures or in other key pollinating taxa, such as hoverflies (Diptera: Syrphidae). Therefore, we aimed to develop predictive allometric equations within a forecasting framework ~~our prescribed iterative framework~~ that take into account these factors and place them alongside a catalogue of pre-existing equations for key pollinating insect taxa within the unified resource of an *R* package, entitled “*pollimetry*”.

**Methodology**

*Specimen collection and measurements*

We obtained specimens from recent field research concerning pollinators and insect biodiversity. In Australia, collections were made in the following states: New South Wales, Victoria, Queensland, South Australia and the North Territory. In Europe, we collected and measured specimens in the following countries: Belgium, England, Germany, Ireland, Spain and Switzerland. In the Americas, we included collections from USA and Brazil.

The majority of specimens were dehydrated and weighed within three-six months of collection, although some, in particular, those from Victoria, Australia, Belgium, Switzerland and Cane’s original samples were of variable ages: from three-five years since collection. Damaged specimens were excluded from measurement. For every included specimen, we obtained preservative time, sample location (latitude and longitude), collection method (pan trap, sweeping, malaise trap) and taxonomic designation.

*Body size and intertegular distance*

Dry weight (mg) was measured on an analytical balance to an accuracy to 0.001g. Both fresh and curated specimens were dehydrated at 70 °C for 24 - 48hrs prior to weighing to remove residual humidity. Specimen pins were not removed prior to weighing. Instead, we identified the pin type and weighed a sample of 10-50 pins per type. The mean weight was then subtracted off total weight. Pin weight variance was minimal (range of standard errors: 6.3-4 to 2mg). Intertegular distance was measured in millimetres using a stereo-microscope, either mounted with a calibrated scale or microscope camera.

*Data analysis: Model structures*

All analyses were undertaken in *R* v3.5.0 (R Core Team , 2018). For each taxon, we constructed species mean datasets. Every female and/or male of each species was averaged within biogeographical region of origin. The European honeybee, *Apis mellifera*, and the sweat bee, *Halictus rubicundus*, were present in multiple regions so specimens were removed from their introduced regions, Australia and North America respectively, prior to analysis. We used a power function in model formulation which is typical of predictive allometry:

Ln(Y) = ln(aj) + bj\*ln(IT)

where Y = body weight/size, aj = intercept and bj = allometric co-efficient and its interactions. To test whether allometric relationships vary with sex, biogeographic region and taxonomic family, we included an interaction between IT and these factors. We constructed linear mixed effect models using *MCMCglmm* (Bates et al. 2014) to predict body size as a function of these explanatory variables, with species included as ~~crossed~~ a random term.

We also assessed the correlation between ITD and body length and compared each in predicting body size using OLS regression.

*Data analysis: Incorporating phylogeny*

We explored the influence of phylogenetic relatedness in predicting pollinator body size in bees by first identifying phylogenetic signal in body size and then implementing a phylogenetic generalised linear mixed model to take into account phylogenetic dependence and intraspecific differences (sexual dimorphism). We first obtained a super matrix genera backbone tree (Hedtke et al. 2013) and removed non-represented genera using *ape* (Paradis et al., 2004). Species tips were added to genera nodes as pure-birth subtrees using *phytools* (Revell et al. 2012). As such, we made the explicit assumption that phylogenetic patterns in body size were assessed at and above the genera level. We estimated node ages using the mean path lengths method of Britton et al. (2002) as branch lengths in the original genera tree were mean number of substitutions per site.

We assessed the significance of phylogenetic signal of both body size and ITD using Pagel’s λ (Pagel 1999) with *phytools* (Revell et al. 2017). A PGLMM was formulated with a modified formula (i.e. without taxonomic family) to GLMMs with phylogeny incorporated to account for the co-variance between species, as a random term.

*Data analysis: Model selection and cross-validation*

We first fitted the full model with all predicted explanatory variables. We then performed model selection assessing all subset models using the ‘dredge’ function within *MuMIn* (Barton 2018). The best fitting models were then ranked by their Akaike Information Criterion (AIC) weighting. For wider applicability of evaluated models, we also extracted the highest-ranked models which excluded region, sex and family for both taxa from the model selection tables.. Lastly, we considered ITD in isolation.

*Data analysis: Cross-validation*

We implemented k-fold cross validation to test overall model performance and compare prediction error (Stone 1974; Kohavi 1995). Species mean datasets were divided into five equal sets containing a random subset of species. Each model was then evaluated iteratively upon each k-1 set (training set consisting of four folds), and then the actual and predicted values were compared within the one left out ‘test’ set. This was done repeatedly so each set was both the test set and contained within the training sets. New levels of random terms were allowed within each predicted set. We then assessed model performance on the basis of average root-mean square error (RMSE), ~~cross-validated~~ *~~R~~~~2~~* and AIC across the five sets. For PGLS models, λ for each model was fixed at the optimised value from the full dataset. Simplified cross validation was used to test the predictive accuracy of Cane’s (1987) original equation:

IT = 0.77(body size)0.405., We compared the actual and Cane (1987)predicted values for each of the k test sets and derived a mean RMSE.

*Data analysis: Intraspecific predictions*

We assessed the utility of ITD in predicting intraspecific body size variation. For the five most abundant species of both bees and hoverflies we tested the utility of ITD in predicting intraspecific female body size variation using species-level OLS regression. Furthermore, for each species, we plotted trait means independently against increasing sample size to estimate the adequate sample size whereby variance stabilised within confidence intervals of the actual sample size.

*Data availability*

All data including R code and R package are freely available here: <https://github.com/liamkendall/pollimetry>/

**Results**

*Species and specimen distribution*

In total, we measured 320 bee species from Australia, Europe, North America and South America, including Cane’s (1987) original 20 species and 103 hoverfly species from Australia and Europe (Table 1). Five out of six bee families and both hoverfly subfamilies were represented. Number of individuals per bee species ranged from one (109 ♀ only, 74 ♂ only) and 201 (♀ *Homalictus urbanus*) and 68 (♂ Bombus impatiens) with a mean of eight ♀ and four ♂ per species. In hoverflies, specimen number per species ranged from one (49 ♀ only hoverfly species, 34 ♂ only hoverfly species) to 32 (♀ *Austrosyrphus* spp.) and 18 (♂ *Austrosyrphus* spp.) with a mean of three per species for both sexes.

*Interspecific model selection and performance*

Body length and IT were highly co-linear (bees: F(1,1834): 1629, *R2*: 0.89, log(IT) = -1.573 (0.017) + 1.102 (0.008) \* log(BL) ; hoverflies: F(1,165): 429, *R2*: 0.72, log(IT) = -1.982(0.137) + 1.295 (0.063) \* log(BL). ITD better described body size in both taxa,

For bees, LME analyses found that models which considered region, family and sex in interaction with ITD best-predicted body size on the basis of AIC (Table 1A). In particular, family and region in interaction with ITD were most informative, however for hoverflies, region and sex were more important than subfamily across the best fitting models.

Phylogenetic signal was highly significant (*p* <0.001) for bee ln body size (λ: 0.71) and ln ITD (λ: 0.75) (Figure 1). Larger body size was most pronounced within Apidae, the largest bee in our dataset being *Xylocopa (Koptortosoma) lieftincki* (♀ mean weight: 260.25g), whereas Halictid (i.e. Halictus, Homalictus and *Lasioglossum* species)and Colletid bees (in particular, *Euhesma* sp. (♀ mean weight: 0.71g, ♂ mean weight: 0.66g) and *Hylaeus* (e.g. *Hylaeus communis* (♀ mean weight: 6.15g, ♂ mean weight: 2.76g ) were smaller than average (Figure X). PGLS models exhibited similar trends to LME models with the model that included an interaction between ITD and region having the lowest AIC weighting (Table 1). Further, PGLS models were significantly better fitting than non-phylogenetic GLS models (ITD \* Region: χ2: 27.47, p<0.001, λ: 0.54; ITD + Region: χ2: 27.88, p<0.001, λ: 0.54; ITD χ2: 45.28, p<0.001, λ: 0.6).

All new bee models out-performed Cane’s (1987) original equation on the basis of RMSE in cross-validation. Cross-validation indicated that high predictive accuracy persisted across all tested models for both bees and hoverflies. Overall the ranges of model precision was higher in bee models than hoverfly models (bees: RMSE: 0.381 – 0.432, *~~R~~*~~2~~~~: 0.858 – 0.89,~~ AIC 394.85-446.23; hoverflies: RMSE: 0.417 – 0.438, *~~R~~*~~2~~ ~~=0.815 – 0.832,~~ AIC: 167.31 – 173.64 (Table 2, Figure 2). In hoverflies, there was less overall dispersion in prediction error~~. Interestingly, the top-ranked LME models differed on the basis of RMSE,~~ *~~R~~~~2~~* ~~and AIC in both data sets.~~ Incorporating phylogeny into model-fitting resulted in a marginal increase (mean RMSE decrease ln(g): 0.014 to IT + Region model, ln(g) 0.002 from best-fitting LME model) in predictive precision for bees when considered in interaction with biogeographic region. Minimal differences in model precision were observed between phylogenetic- and non-phylogenetic models which considered only ITD or in addition to region.

*Intra-specific predictions*

Across the five most abundant species of bees and hoverflies (females only), the strength of intraspecific predictions of body size using ITD was varied (Table 3; Figure 3). All bee species exhibited a significant relationship between ITD and dry weight, however *R2*differed considerably from 0.02 in *Homalictus urbanus* to 0.46 for *Lasioglossum lanarium* (Table 3, Figure 3). Three of five hoverfly species, *Austrosyrphus* sp., *Helophilus parallelus and Melanostoma scalare* exhibited significant trends.

Sample size exhibited an interesting trend in relation to both ITD and dry weight. In bees, mean ITD and dry weight stabilised within the 95% confidence intervals of the total sample size with >20 specimens per species (Figure S1A). For hoverflies, the lower overall sample sizes of each species limited inference of sample-size / mean stabilisation (FIGURE S1B).

*Summary of R package functions*

The accompanying R package, ‘pollimetry’, includes a total of X functions for estimation of pollinator body size using pre-existing equations (see Table S1 for description of collation and co-efficients) or our new equations (Table 1). Also included are Greenleaf et al.’s (2007) and van Nieuwstadt and Iraheta’s (1996) allometric equations for estimating bee foraging distances using ITD or head width respectively, as well as Cariveau et al’s (2016) allometric equations for estimating bee tongue length. New equations will be periodically updated (~biannually) in the form of package updates as novel data becomes available. In the case of novel data from new biogeographical regions or new allometric trait evaluations, updates will be accompanied by a formal publication.

**Discussion**

Herein, we used a forecasting framework to develop and test a suite of dynamic predictive allometric models for two key pollinating taxa, bees and hoverflies. The hoverfly models represent the first predictive allometric models for this important insect group. Addditionally, we present the most comprehensive examination of both bee and hoverfly body variation to date that takes into account sex, biogeography and relatedness. As was first suggested by Cane (1987), we re-iterate the utility of the intertegular distance (ITD) in predicting ~~intra- and~~ interspecific body size variation, albeit highlight its sensitivity the beforementioned factors.. Overall, both LME and PGLS model structures exhibited high predictive precision, resulting in a suite of highly applicable models for researchers worldwide. By incorporating biogeography, gender and/or taxonomy/phylogeny we improved model performance and overcame the limitations of traditional predictive allometry. These three predictors represent fundamentally related causes of body size variation in pollinating insects.

Terrestrial invertebrates show considerable geographic variation in shape and biogeographical differences in predictive allometry are well-established. We found subtle yet significant body size variation in interaction with ITD between all three regions. The most pronounced differences in our study were between Australia and North America or Europe: both North American and European bees were heavier per unit ITD and exhibited reduced slopes. Martin et al (2014) assessed biogeographical patterns of order-level insect body-length allometric coefficients and contributed differences between geographic regions to a latitudinal gradient, suggesting comparable geographic regions should exhibit similar allometric coefficients. Our results conform to this prediction, although latitude wasn’t explicitly tested due to poor overall coverage within regions. However, previous studies have found both similarities (Gowing & Recher 1984 and Rogers et al. 1977) and differences in allometric coefficients between comparable regions (Schoener 1980 and Rogers et al. 1977) and only one prior predictive allometric study examined multiple biogeographic regions in concert, albeit as individual models (Schoener 1980).

The key difference between these studies and our own, is that this study demonstrated that the influence of different biogeographic regions appears in conjunction with species evolutionary histories and sexual dimorphism. Observed biogeographical differences can arise from differing diversification patterns and as well as from sampling biases, such as random variation among species and the ranges of body size measured. Extrapolating estimates for specimens beyond the range of sizes/lengths used for model creation can lead to imprecise estimation of larger insects’ body mass because of the non-linear nature of allometric models (Sage 1982). Yet in both taxa, we sampled systematically across a wide range of body sizes (see Figure 1 for bees) and incorporated far more species than the majority of prior allometric studies that formulated predictive models at either the superfamily or family level. Therefore, differing diversification patterns and random variation in species sampled appear central to biogeographic differences in allometric coefficient patterns in this study. As a result, predictive allometric model development requires large-scale efforts incorporating multiple regions, biomes and ecosystems for greater accuracy in order to develop biological hypotheses and reasoning driving observed trends.

Sex was retained as an integral predictor either in addition or in interaction with ITD for both taxa. This suggests that although ITD is a robust predictive trait, other morphological aspects must differ between the sexes resulting in sexual size dimorphism (SSD) in absence of ITD differences.SSD is common among insects. In both Diptera and Hymenoptera, SSD is predominantly (>80%) female-biased, including in Apoidea and Syrphidae (Shreeves and Field 2008; Francuski et al. 2011; Milankov et al. 2013). Female-biased SSD is hypothesised to be a result of the greater fitness and increased fecundity inferred by larger female body size (Teder and Tammaru 2005: Stillwell 2010). In bees, SSD is attributed to the physical requirements of nest provisioning and construction. As such, female-biased SSD is more pronounced in ground nesting and/or provisioning taxa than non-provisioning and/or cavity nesting taxa (Shreeves and Field 2008). SSD is also smaller in eusocial species (Medina et al. 2016). Furthermore, female bees exhibit specialised morphological structures for resource collection (i.e. scopal hairs and corbiculae) as well as self-preservation (i.e. a stinging ovipositor). In hoverflies, there exist few examples of specialised morphological sexual dimorphism, other than eye shape. However, our results demonstrate that ITD as a predictive trait is sensitive to sexual dimorphism making sex an important consideration in prediction, especially as ecologically relevant allometric traits are gender-related (e.g. foraging distance).

Few previous studies have assessed the utility of predictive models in describing intrageneric or intraspecific allometric traits (e.g. Hagen & Dupont 2013; Cariveau et al. 2016). Our results provide reasonable proof that intraspecific body variation can be predicted using ITD. However, the large variation in predictive power suggests that it is potentially sensitive to environmental conditions and small sample sizes. Adult body size variation in holometabolous insects is a direct result of diet and environment during ontogeny and larval development (Davidowitz et al. 2004). Sampled specimens in many cases, came from differing environments and likely developed in disjunct conditions, potentially leading to prediction error. Despite this, our results suggest there are reasonable constraints upon the overall influence of ontogeny and larval development on adult body size, resulting in adequate predictive power in intraspecific models. Sample size is also an important consideration for intraspecific predictions. The two hoverfly species which showed no-trend had small sample sizes. Although *Melanostoma scalare* exhibited a significant trend with the same sample size, we caution the development of intra-specific predictive allometric models with a sample size < 10. In any case, the interspecific model should provide adequate estimations of body size distributions within a given population.

~~Our prescribed iterative framework provides a new and much needed re-invigoration of the field of predictive allometry. A key principle of this is the incorporation of complex model structures. Traditional predictive allometry have typically utilised ordinary least squares or major axis regression. Although both have some attractive statistical properties for allometric prediction (Warton et al. 2006; Legendre & Legendre 2012), both fail to statistically incorporate the non-independence, pseudo-replication and random variation that are part and parcel of biological research. LME structures allowed us to accurately account for species overlap within regions and potential differences in equipment measurement error to assess macro-ecological trends.~~

Phylogenetic signal in body size variation has been inferred in a number of vertebrate and invertebrate groups (see Ashton 2004). Failing to account for dependent phylogenetic patterns heightens the risk of inaccurate predictions (Martins 1991; Martins et al. 2002; Garland et al. 2005). Incorporating phylogeny considerably decreased AIC in our predictive PGLS models relative to GLS models. This suggests that PGLS methods are more informative than non-phylogenetic allometric models. Where phylogenetic information is unavailable, incorporating taxonomy represents an adequate compromise. Including taxonomic family within our LME bee models improved predictive accuracy however in hoverflies, including subfamily was less informative in describing body size variation. Of interest from a predictive sense, incorporating taxonomy can aid in elucidating dependent patterns within a phylogeny, such as the disproportionately larger body sizes observed in Apid bees such as inthe genera *Bombus* and *Xylocopa*. Taxonomy has been used successfully before as a phylogenetic proxy (Cariveau et al. 2016). As such, our results suggest both phylogenetic and taxonomic models can improve model predictions.

An important yet underutilised aspect of predictive ecology is estimating model performance on untested data. As independent data is most-often unavailable to researchers, k-fold cross validation is an attractive alternative (Stone 1974; Kohavi 1995). By dividing our data into sequential validation and testing sets, we were able to assess how accuracy differs between test sets, each comprising a different random subset of our entire dataset. Such an approach has garnered interest in validating species distribution models, in particular, to address spatial auto-correlation (Hijmans 2012; Wenger & Olden 2012). However, despite its utility here, it remains largely unused in other fields of predictive ecology. The function of cross-validation is most apparent herein given the differential results, in terms of model performance metrics: AIC, CV *R2* and RMSE between k datasets and the full dataset used in model selection. Each in isolation lead to differing selections of the ‘best’ LME models for both bees and hoverflies. The PGLS ITD \* Region model was the only model congruent in both full-model selection and cross-validation. By considering multiple metrics (AIC, RMSE and CV *R2*) a better predictive picture is produced and disseminating the most accurate models becomes a hypothesis-driven formula. ~~It was also for this reason that we selectively removed predictors (i.e. region or sex).~~ In essence, predictive allometry requires acceptance of multiple models as equal-best or ‘most accurate’, with model choice and usage becoming the decision of the end-user and based on the aims of the proposed study.

The accompanying R package, “pollimetry”, provides a user-friendly interface to estimate pollinator body size and some allometric traits (bee foraging distance and proboscis length). Sampling regimes and research questions may not garner investigation of sex-related allometric differences and will occur outside the included biogeographic regions. ~~It was for these reasons we chose not to average models.~~ More useful predictive allometries require multiple reduced models. These models will continue to be updated as new body size data becomes available, including from other biogeographic regions (South America and Africa), and will enable investigation and re-evaluation of other allometric traits at both intra- or inter-specific levels. The consequence of size is ubiquitous within pollination research yet few have utilised allometric theory in studying pollinating taxa beyond bees. The prescribed framework heralds a dynamic new direction for predictive allometry and should provide greater prediction through hypothesis-led model choice, testing and investigation in allometric research.

Table 1. Distribution of measured specimens. Numbers in parenthesis denote total specimens and species per country, family and/or subfamily.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Taxa | Region | Country | Family | Subfamily |
| Bee | Australasia | Australia (899, 93) | Apidae (185, 20)  Colletidae (130, 15)  Halictidae (519, 38)  Megachilidae | Apinae (125, 13), Xylocopinae (60, 7)  Colletinae (76, 7), Euryglossinae (49, 3)  Hylaeinae (5, 5)  Halictinae (441, 25)  Nomiinae (78, 13)  Megachilinae (65, 20) |
|  | Europe | Belgium | TBC |  |
|  |  | England (46, 4) | Apidae | Apinae (46, 4) |
|  |  | Germany (765, 63) | Andrenidae  Apidae (189, 12)  Colletidae  Halictidae  Megachilidae | Andreninae (197, 14)  Apinae (188, 11), Nomadinae (1, 1)  Hylaeinae (13, 6)  Halictinae (337, 18)  Megachilinae (29, 13) |
|  |  | Ireland (52, 15) | Andrenidae  Apidae  Colletidae  Halictidae | Andreninae (4, 2)  Apinae (29, 6)  Hylaeinae (5, 2)  Halictinae (14, 5) |
|  |  | Spain (74, 46) | Andrenidae (18, 13)  Apidae (27, 17)  Colletidae  Halictidae (11, 6)  Megachilidae  Melittidae | Andreninae (10, 8), Panurginae (8, 5)  Apinae (16, 10), Nomadinae (6, 4)  Xylocopinae (5, 3)  Colletinae (4, 2)  Halictinae (8, 5), Rophitinae (3, 1)  Megachilinae (7, 6)  Dasypodainae (7, 2) |
|  |  | Switzerland (210, 63) | Andrenidae  Apidae (60, 20)  Colletidae  Halictidae  Megachilidae  Melittidae | Andreninae (54, 14)  Apinae (54, 16), Nomadinae (4, 3), Xylocopinae (2, 1)  Colletinae (3, 1)  Halictinae (76, 20)  Megachilinae (15, 6)  Melittinae (2, 2 |
|  | North America | USA (555, 73) | Andrenidae (24, 10)  Apidae (193, 22)  Colletidae (62, 7)  Halictidae (244, 26)  Megachilidae  Melittidae | Andreninae (19, 5), Oxaeinae (1, 1), Panurginae (4, 4)  Apinae (177, 17), Nomadinae (1, 1), Xylocopinae (15, 4)  Colletinae (2, 2), Hylaeinae (60, 5)  Halictinae (242,24), Nomiinae (1, 1), Rophitinae (1, 1)  Megachilinae (31, 7)  Dasypodainae (1, 1) |
|  | South America | Brazil (204, 19) | Apidae  Halictidae  Megachilidae | Apinae (149, 10), Xylocopinae (25, 5)  Halictinae (11, 2)  Megachilinae (11, 2) |
| Hoverfly | Australasia | Australia (120, 19) |  | Eristalinae (25, 7), Syrphinae (95, 12) |
|  | Europe | Ireland (39, 15) |  | Eristalinae (8, 5), Syrphinae (31, 10) |
|  |  | Spain (8, 8) |  | Eristalinae (6, 6), Syrphinae (2, 2) |
|  |  | Switzerland (232, 79) |  | Eristalinae (114, 37), Syrphinae (118, 42) |

Table 2. AIC and ΔAIC for interspecific models. Model types: i) LME: linear mixed effect models and ii) PGLS: phylogenetic generalised least squared models. As these two model types used different datasets, AIC values are not directly comparable.







Table 2. Full k-fold cross validation results. MSE: Mean standard error, RMSE: Root mean square error, *R2*: Cross-validated *R2* and AIC: Akaike Information Criterion. Values are the mean across five folds.







Table 3. Model parameters of intraspecific ln(body size)~ln(IT) size relationships. F: F-statistic and degrees of freedom for each model. A and B: intercept and IT co-efficients ± standard error, *R2*: Adjusted R-squared and P: p-value of full model. Only females were used in both analyses.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Taxa | Species | F(df) | A | B | *R2* | P |
| Bee | *Andrena flavipes* | 21.07 (1,57) | 1.308 ± 0.394 | 2.03 ± 0.442 | 0.257 | <0.001 |
|  | *Bombus lucorum* | 81.15 (1,101) | 1.413 ± 0.297 | 1.966 ± 0.218 | 0.44 | <0.001 |
|  | *Homalictus urbanus* | 6.055 (1,209) | -0.164 ± 0.033 | 1.166 ± 0.474 | 0.024 | 0.014 |
|  | *Lasioglossum lanarium* | 53.87 (1,61) | 0.702 ± 0.198 | 2.13 ± 0.29 | 0.46 | <0.001 |
|  | *Lasioglossum pauxillum* | 43.92 (1,110) | 0.402 ± 0.057 | 3.112 ± 0.469 | 0.279 | <0.001 |
| Hoverfly | *Austrosyrphus spp.* | 12.7 (1,30) | 0.087 ± 0.458 | 2.032 ± 0.57 | 0.274 | 0.001 |
|  | *Helophilus parallelus* | 14.84 (1,17) | 0.286 ± 0.857 | 2.485 ± 0.645 | 0.435 | 0.001 |
|  | *Sphaerophoria macrogaster* | 0.04 (1,8) | 0.361 ± 0.274 | 0.195 ± 0.907 | -0.11 | N.S. |
|  | *Episyrphus balteatus* | 0.08 (1,8) | 1.334 ± 1.885 | 0.885 ± 2.229 | -0.11 | N.S. |
|  | *Melanostoma scalare* | 6.38 (1,7) | -2.172 ± 1.324 | 7.619 ± 3.016 | 0.4 | 0.03 |

Figure 1. Bee genera tree from Hedtke et al. (2013) with species subtrees. Colour denotes mean ln(body size) of each bee species included in analyses.

Fig 2. Root mean square error (RMSE) across k-fold training and test sets for each model. Model numbers refer to these described in Table 2.

**Figure 3.** Intraspecific predictions of female body size with intertegular distance. Lines denote line of best fit from OLS regression.

SUPPLEMENTARY; *Description of* *pre-existing equations*

*Existing equations*

We selected three key pollinating insect orders: Diptera, Hymenoptera and Lepidoptera and collated all known predictive allometric models. From an initial literature search, we obtained the publications analysed by Martin et al. (2014). We then reviewed each publication individually, including their references and citations for additional equations.

Diptera: 26 allometric models for Diptera were collated (Table S1A). 11 models were reported for the entire order, including nine without any taxonomic breakdown of samples used. 12 for the three main suborders Nematocera (6), Brachycera (4) and Cycllorapha (2) and two for specific families; Asilidae and Bombyliidae.

Hymenoptera: 38 allometric models for Hymenoptera were collated (Table S1B). These included eight combined, seven excluding ants (Formicidae) as well as ten for Formicidae. There are three equations for Vespidae and two equations for Apidae (Cane 1987 & Sabo et al. (2002). Sample et al’s (1993) body length (BL) and body length\*width (BW) equations are provided for Braconidae, Ichneumonidae, Halictidae and Pompilidae.

Lepidoptera: 21 allometric models for Lepidoptera were collated (Table S1C). This includes 13 with varying taxa and without lower classifications. Hodar (1997) provides specific models for Heterocera (moths) and Ropalocera (butterflies). Sample et al. (1993) provide BL and BL\*BW models for Microlepidoptera and two moth families: Geometridae and Arctiidae.

Foraging distance and proboscis length: We collated six bee foraging distance models from van Nieuwstadt and Iraheta (1996) and Greenleaf et al. (2007) and three models for bee proboscis length from Cariveau et al. (2016)

Table S1. Allometric equations for pollinating taxa. BL = Body length, BW = Body width, IT = intertegular distance. OLS = Ordinary Least Squares regression. MA = Major axis regression, MU = Multivariate regression. Equations are present in the form of y = ln(B0) + ln(B1), which is equivalent to y = b0Xb1. \* = Included body width as well as length

1. **Diptera**.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | | **Taxa** | | **Families** | | **Sample size** | | **Biogeographical region** | | | **Measure** | **Range in body length** | **Reg. Type** | **Type** | **Equation** | | | |
|  |  | |  | |  | | **(Families: species)** | |  |  | | **(mm)** |  | | ***B*0 ± S.E.** | ***B*1 ± S.E.** | **Resi. SE** | ***R2*** |
| Rogers et al. (1977) | |  | |  | | (#:84) | | Washington, USA | | | BL | 0.9-34 | OLS | PF | -3.298 ± 0.115 | 2.366 ± 0.078 | 0.57 | **0.96** |
| Schoener (1980) | |  | |  | |  | | Dry forest, CR | | | BL | N.P | OLS | PF | A=-2.603 ± 0.0688 | B = 1.64 ± 0.1224 | NA | 0.795 |
| Schoener (1980) | |  | |  | |  | | Rain forest, CR | | | BL | N.P. | OLS | PF | A = -2.688 ± 0.051 | B = 1.59 ± 0.1173 | NA | 0.775 |
| Schoener (1980) | |  | |  | | (#:171) | | Massachusetts | | | BL | N.P. | OLS | PF | A=-3.816 ± 0.561 | B=2.42 ± 0.0969 | NA | 0.89 |
| Gowing and Recher (1984) | |  | |  | | (100) | | NSW, Australia | | | BL | 2-11 | OLS | PF | 3.653 ± 0.129 | 2.546 ± 0.071 | 0.37 | **0.93** |
| Sample et al. (1993) | |  | | Combined | | (15:257) | | West Virginia, USA | | | BL | 2.9-23.65 | OLS | PF | -3.184 ± 0.184 | 2.213 ± 0.085 | NA | 0.85 |
| “ | |  | |  | |  | | “ | | | BL\*BW | “ | OLS | PF | -2.197 ± 0.089 | 1.309 ± 0.03 | NA | **0.94** |
| “ | | NEM | | BIB,SCI,TIP | | (3:46) | | “ | | | BL | 3.55-23.65 | OLS | PF | -3.675 ± 0.23 | 2.212 ± 0.141 | NA | **0.92** |
| “ | |  | |  | |  | | “ | | | BL\*BW | “ | OLS | PF | -2.217 ± 0.205 | 1.288 ± 0.071 | NA | **0.94** |
| “ | | BRA | | ASI, DOL, EMP, RHA, STR, THE | | (6:80) | | “ | | | BL | 2.9-17.99 | OLS | PF | -3.374 ± 0.230 | 2.158 ± 0.101 | NA | **0.92** |
| “ | |  | |  | |  | | “ | | | BL\*BW | “ | OLS | PF | -2.2 ± 0.147 | 1.259 ± 0.049 | NA | **0.95** |
| “ | | CYC | | CAL, LAU, MUS, OTI, SYR, TAC | | (6:119) | | “ | | | BL | 2.9-15.65 | OLS | PF | -3.619 ± 0.212 | 2.632 ± 0.101 | NA | **0.92** |
| “ | |  | |  | |  | | “ | | | BL\*BW | “ | OLS | PF | -2.02 ± 0.131 | 1.298 ± 0.042 | NA | **0.94** |
| Hodar (1997) | | BRA | |  | | (26) | | Gaudix-Baza, Spain | | | HW | NA | OLS | PF | A=0.655 ± 0.105 | B=2.526 ± 0.139 | 0.47 | **0.933** |
|  | | NEM | |  | | (10) | | “ | | | HW | NA | OLS | PF | A=3.942 ± 0.259 | B=3.106 ± 0.278 | 0.55 | **0.94** |
| Ganihar (1997) | |  | | NA | | (#:20) | | Goa, India | | | BL |  | OLS | PF | -3.4294 ± 0.01994 | 2.5943 ± 0.0334 | 0.03 | **0.99** |
| Johnson and Strong (2000) | | ALL | | NA | | (75) | | Jamaica | | | BL | 1-12.5 |  | PF | -2.462 ± 0.196 | 1.881 ± 0.146 |  | 0.83 |
| “ | | NEM | | NA | | (21) | | “ | | | BL | 1-4.8 |  | PF | -2.562 ± 0.244 | 1.373 ± 0.207 |  | 0.836 |
| “ | | NEM exc. | | NA | | (54 | | “ | | | BL | 1.2-12.5 |  | PF | -2.105 ± 0.178 | 1.805 ± 0.124 |  | 0.895 |
| *Sabo et al. (2002)* | | *BRA* | |  | |  | | *California, USA* | | | *BL* | *N.P.* |  | *PF* | *A = 0.006 ± 0.007* | *B = 3.05 ± 0.36* |  | *0.85* |
| *“* | | *NEM* | |  | |  | | *“* | | | *BL* | *N.P.* |  | *PF* | *A = 0.1 ± 0.06* | *B = 1.57 ±0.2* |  | ***0.9*** |
| *“* | |  | | *Asilidae* | | *(1:9)* | | *“* | | | *BL* | *N.P.* |  | *PF* | *A = 0.38 ± 2.625* | *B = 1.5 ± 2.469* |  | *0.74* |
| *“* | |  | | *Bombyliidae* | | *(1:10)* | | *“* | | | *BL* | *N.P.* |  | *PF* | *A = 0.007 ± 0.011* | *B = 3.337 ±0.676* |  | ***0.95*** |
| *Brady and Noske (2006)* | | *NA* | | *NA* | | *(9 sp:78 spe)* | | *NT, AUS* | | | *B:* | *2-28* | *OLS* | *L* | *A= -0.041 ± 0.004* | *B = 0.010 ± 0.001* | *0.02* | *0.838* |
| Wardhaugh (2013) | |  | |  | | (#:16) | | QLD, AUS | | | BL |  | MA | PF | -3.29 ± 0.45 | 2.65 ± 0.36 | NA | 0.72 |
| Wardhaugh (2013) | |  | |  | | (#:16) | | QLD, AUS | | | BL \* BW |  | MA | PF | -1.91 ± 0.19 | 1.22 ± 0.11 | NA | 0.87 |

BIB =Bibionidae, SCI = Sciaridae, TIP = Tipulidae, ASI= Asilidae, DOL = Dolichopodidae, EMP = Empidae, RHA = Rhagionidae, STR = Stratiomyidae, THE = Therevidae, CAL = Calliphoridae, LAU = Lauxaniidae, MUS = Muscidae, OTI = Otitidae, SYR = Syrphidae, TAC = Tachinidae.NEM = Nematocera, BRA= Brachycera, CYC = Cyclorrapha

1. **Hymenoptera.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | | **Taxa** | | **Families** | | **Sample size** | | | **Biogeographical region** | | **Measure** | **Range in body length** | | **Reg. Type** | **Type** | **Equation** | | | |
|  |  | |  | |  | | **(Families: species)** |  | |  | | | **(mm)** |  | | ***B*0 ± S.E.** | ***B*1 ± S.E.** | **Resi. SE** | ***R2*** |
| Rogers et al. (1977) | |  | | \*\* | | (#:97) | | | Washington, USA | | BL | 0.7-27 | | OLS | PF | -3.871 ± 0.108 | 2.407 ± 0.06 | 0.55 | **0.97** |
| “ | |  | | FOR | | (#:34) | | | “ | | BL | 1.2-13.5 | | OLS | PF | -4.029 ± 0.171 | 2.572 ± 0.097 | 0.40 | **0.98** |
| Cane (1987) | |  | | Apidae | | (6:20) | | | USA | | ITD | 1-6 | | NL | EXP | A=0.77 | B=0.405 |  | **0.96** |
| *Schoener (1980)* | | *ALL* | | *\*\** | | *(#:174)* | | | *Dry forest, C. Rica* | | *BL* | *N.P* | | *OLS* | *PF* | *A = 0.043 ± 0.05* | *B = 2.07 ± 0.091* | *NA* | *0.87* |
| *“* | | *“* | | *“* | | *(#:122)* | | | *Rain forest, C. Rica* | | *BL* | *N.P.* | | *OLS* | *PF* | *A = 0.022 ± 0.056* | *B = 2.29 ± 0.137* | *NA* | *0.835* |
| *“* | | *“* | | *“* | | *(#:82)* | | | *Massachusetts* | | *BL* | *N.P.* | | *OLS* | *PF* | *A = 0.016 ± 0.072* | *B = 2.55 ± 0.107* | *NA* | *0.937* |
| *“* | | *“* | | *FOR* | | *(#:25)* | | | *Dry forest, C. Rica* | | *BL* | *N.P* | | *OLS* | *PF* | *A = 0.012 ± 0.113* | *B = 2.72 ± 0.26* | *NA* | ***0.907*** |
| *“* | | *“* | | *“* | | *(#:20)* | | | *Rainforest, C. Rica* | | *BL* | *N.P.* | | *OLS* | *PF* | *A = 0.21 ± 0.127* | *B = 2.31 ± 0.224* | *NA* | ***0.934*** |
| *“* | | *“* | | *“* | | *(#:13)* | | | *Massachusetts* | | *BL* | *N.P.* | | *OLS* | *PF* | *A = 0.034 ± 0.155* | *B = 2.19 ± 0.342* | *NA* | ***0.908*** |
| Gowing and Recher (1984) | |  | | \*\* | | (86) | | | NSW, Australia | | BL | 1-12 | | OLS | EXP | -2.860 ± 0.099 | 0.478 ± 0.016 | 0.48 | **0.918** |
| “ | |  | | FOR | | (68) | | | “ | | BL | 2-18 | | OLS | PF | -3.306 ± 0.258 | 2.489 ± 0.051 | 0.32 | **0.973** |
| Sample et al. (1993) | | ALL | | - | | (7:274) | | | West Virginia, USA | | BL | 2.81-34.91 | | OLS | PF | -4.284 ± 0.183 | 2.696 ± 0.083 | NA | 0.89 |
| “ | |  | | “ | | “ | | | “ | | BL \* BW | “ | |  | “ | -2.375 ± 0.08 | 1.456 ± 0.028 | NA | **0.95** |
| “ | |  | | Ichneumonidae | | (1: 106) | | | “ | | BL | 3.65-34.91 | |  | “ | -4.149 ± 0.262 | 2.464 ± 0.116 | NA | **0.9** |
| “ | |  | | “ | | “ | | | “ | | BL \* BW | “ | |  | “ | -2.497 ± 0.147 | 1.445 ± 0.053 | NA | **0.94** |
| “ | |  | | Braconidae | | (1:41) | | | “ | | BL | 2.81-15.42 | |  | “ | -3.854 ± 0.273 | 2.441 ± 0.147 | NA | **0.94** |
| “ | |  | | “ | | “ | | | “ | | BL \* BW | “ | |  | “ | -2.19 ± 0.142 | 1.445 ± 0.069 | NA | **0.96** |
| “ | |  | | Vespidae | | (1:19) | | | “ | | BL | 8.14-20.58 | |  | “ | -3.540 ± 0.544 | 2.782 ± 0.195 | NA | **0.96** |
| “ | |  | | “ | | “ | | | “ | | BL \* BW | “ | |  | “ | -1.537 ± 0.307 | 1.319 ± 0.07 | NA | **0.98** |
| “ | |  | | Formicidae | | (1:45) | | | “ | | BL | 3.62-17.41 | |  | “ | -4.727 ± 0.350 | 2.919 ± 0.11 | NA | **0.93** |
| “ | |  | | “ | | “ | | | “ | | BL \* BW | “ | |  | “ | -2.378 ± 0.265 | 1.473 ± 0.106 | NA | **0.9** |
| “ | |  | | Halictidae | | (1:21) | | | “ | | BL | 6-12.76 | |  | “ | -2.891 ± 0.386 | 2.302 ± 0.182 | NA | **0.95** |
| “ | |  | | “ | | “ | | | “ | | BL \* BW | “ | |  | “ | -2.758 ± 0.357 | 1.590 ± 0.119 | NA | **0.95** |
| “ | |  | | Pompilidae | | (1:15) | | | “ | | BL | 5.55-14.32 | |  | “ | -2.341 ± 0.873 | 2.006 ± 0.396 | NA | 0.81 |
| “ | |  | | “ | | “ | | | “ | | BL \* BW | “ | |  | “ | -1.946 ± 0.431 | 1.444 ± 0.154 | NA | **0.93** |
| Hodar (1997) | | ALL | |  | |  | | | Gaudix-Baza, Spain | | HW |  | | OLS | PF | A= 1.999 ± 0.112 | B= 2.09 ± 0.132 | 0.51 | **0.919** |
| “ | |  | | FOR – Workers | |  | | | “ | | HW |  | |  | “ | A= 0.552 ± 0.068 | B= 2.550 ± 0.116 | 0.19 | **0.982** |
| “ | |  | | FOR –Winged | |  | | | “ | | HW |  | |  | “ | A= 1.607 ± 0.127 | B= 2.752 ± 0.25 | 0.31 | **0.938** |
| Ganihar (1997) | |  | | NA \*\* | | (#:26) | | | Goa, India | | BL |  | | OLS | PF | -3.5917 ± 0.1646 | 2.6429 ± 0.1127 | 0.24 | **0.94** |
| Johnson and Strong (2000) | | ALL | |  | |  | | | Jamaica | | BL | 1.4-24.3 | | OLS | PF | -3.556 ± 0.183 | 2.193 ± 0.110 | NA | **0.923** |
|  | | FOR | |  | |  | | | “ | | BL | 1.6-9.9 | | OLS | PF | -3.730 ± 0.298 | 2.103 ± 0.238 | NA | **0.901** |
|  | | \*\* | |  | |  | | | “ | | BL | 1.4-24.3 | | OLS | PF | -3.295 ± 0.241 | 2.102 ± 0.132 | NA | **0.917** |
| *Sabo et al. (2002)* | | *ALL* | | *7\*\*\*\** | | *(7:54)* | | | *California, USA* | | *BL* | *N.P.* | | *NLL* | *PF* | *A= 0.56 ± 0.64* | *B= 1.56 ± 0.4* |  | *0.75* |
| *“* | |  | | *API* | | *(1:10)* | | | *“* | | *BL* | *N.P.* | | *NLL* | *PF* | *A= 0.006 ± 0.041* | *B= 3.407 ± 2.471* |  | *0.81* |
| *“* | |  | | *VES* | | *(1:19)* | | | *“* | | *BL* | *N.P.* | | *NLL* | *PF* | *A= 0.001 ± 0.002* | *B= 3.723 ± 0.798* |  | *0.95* |
| *Brady and Noske (2006)* | |  | | *FOR* | | *(8 sp:100)* | | | *NT, AUS* | | *BL* | *2-10* | | *OLS* | *P* | *0.001* | *2.330 ± 0.0151* | *0.49* | *0.708/0.956* |
| *“* | |  | | *\*\** | | *(9 sp:28)* | | | *NT, AUS* | | *BL* | *4-29* | | *OLS* | *P* | *6.783 ± 0.001* | *2.544 ± 0.26* | *0.57* | *0.786/0.905* |
| Wardhaugh (2013) | |  | |  | | (#:26) | | | Daintree QL AUS | | BL |  | | MA | PF | -4.3 ± 0.38 | 3 ± 0.24 | NA | 0.83 |
| Wardhaugh (2013) | |  | |  | | (#:26) | | | Daintree QL AUS | | BL \* BW |  | | MA | PF | -2.1 ± 0.09 | 1.34 ± 0.05 | NA | **0.97** |

ANT = Anthophoridae, API = Apidae, CHR = Chrysididae, FOR = Formicidae, ICH = Ichneumonidae, SPH = Sphecidae, VES = Vespidae. \*\* = excluded ants. \*\*\*\*Seven families = ANT, API, CHR, FOR, ICH, SPH, VES

1. **Lepidoptera**. \*\*\* = Multivariate regression using multiple length measures: length, width, wing area and wing length. See Garcia-Barros (2015) for parameters.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | | **Taxa** | | **Families** | | | **Sample size** | **Biogeographical region** | | | **Measure** | **Range in body length** | **Reg. Type** | | **Type** | **Equation** | | | |
|  |  | |  | |  | **(Families: species)** | | |  |  | | **(mm)** | |  | | ***B*0 ± S.E.** | ***B*1 ± S.E.** | **Resi. SE** | ***R2*** |
| Rogers et al. (1977) | |  | |  | | | (#:22) | Washington, USA | | | BL | 1.6-17 | OLS | | PF | -4.037 ± 0.133 | 2.903 ± 0.08 | 0.31 | **0.99** |
| Sample et al. (1993) | | ALL | |  | | | (#:384) | West Virginia, USA | | | BL | 2.76-40.73 |  | | PF | -5.036 ± 0.157 | 3.122 ± 0.064 |  | **0.93** |
| “ | | “ | |  | | | “ | “ | | | BL \* BW | “ |  | | “ | -2.607 ± 0.088 | 1.457 ± 0.024 |  | **0.95** |
| “ | |  | | MIC. | | | (#:46) | “ | | | BL | 2.76-10.6 |  | | “ | -4.913 ± 0.325 | 2.918 ± 0.169 |  | **0.93** |
| “ | |  | |  | | |  | “ | | | BL \* BW |  |  | | “ | -2.715 ± 0.199 | 1.395 ± 0.08 |  | **0.93** |
| “ | |  | | GEO | | | (1:58) | “ | | | BL | 6.45-21.70 |  | | “ | -4.172 ± 0.411 | 2.628 ± 0.167 |  | **0.9** |
| “ | |  | |  | | |  | “ | | | BL \* BW | “ |  | | “ | -2.343 ± 0.283 | 1.387 ± 0.084 |  | **0.91** |
| “ | |  | | ARC | | | (1:60) | “ | | | BL | 5.05-20.06 |  | | “ | -3.755 ± 0.242 | 2.658 ± 0.105 |  | **0.96** |
| “ | |  | |  | | |  | “ | | | BL \* BW | “ |  | | “ | -1.658 ± 0.148 | 1.222 ± 0.044 |  | **0.96** |
| Sage et al. (1982) | |  | |  | | | (#:25) | Texas, USA | | | BL | 4.9-22.9 |  | | PF |  |  |  | **0.92** |
| Hodar (1996) | | HET | |  | | | (10) | Gaudix-Baza, Spain | | | HW |  |  | | PF | A=2.053 ± 0.25 | B=2.804 ± 0.236 | 0.493 | 0.946 |
| “ | | ROP | |  | | | (10) | “ | | | HW |  |  | | “ | A=1.634 ± 0.46 | B=2.793 ± 0.446 | 0.485 | 0.831 |
| Ganihar et al. (1997) | |  | | NA | | | (#:10) | Goa, India | | | BL |  |  | | PF | -4.7915 ± 0.7507 | 2.8585 ± 0.2567 | 0.4568 | **0.93** |
| Johnson and Strong (2000) | |  | | NA | | | (40) | Jamaica | | | BL | 2.2-18.6 | OLS | | PF | -3.268 ± 0.255 | 2.243 ± 0.130 | NA | **0.942** |
| *Schoener (1980)* | |  | | *NA* | | | *(#:29)* | *Dry forest, Canas, Costa Rica* | | | *BL* | *N.P.* |  | | *PF* | *A= 0.026 ± 0.186735* | *B= 2.55 ± 0.571429* |  | ***0.958*** |
| *“* | |  | | *NA* | | | *(#:7)* | *Rainforest, Guipiles, Costa Rica* | | | *BL* | *N.P.* |  | | *“* | *A= 0.078 ± 0.139796* | *B= 1.32 ± 0.683673* |  | *0.749* |
| *“* | |  | | *NA* | | | *(#:18)* | *Massachusetts* | | | *BL* | *N.P.* |  | | *“* | *A= 0.014 ± 0.18673* | *B= 2.55 ± 0.571429* |  | *0.77* |
| *Brady and Noske (2006)* | |  | |  | | | *((6 sp: 28)* | *NT, AUS* | | | *BL* | *7-34* | *OLS* | | *PF* | *0.001* | *2.313 ± 0.223* | *0.396* | *0.805/0.938* |
| Wardhaugh (2013) | |  | | NA | | | (#:11) | Daintree QL AUS | | | BL |  | MA | | PF | -3.83 ± 0.41 | 2.77 ± 0.27 | NA | 0.83 |
| Wardhaugh (2013) | |  | | NA | | | (#:11) | Daintree QL AUS | | | BL \* BW |  | MA | | “ | -2.1 ± 0.21 | 1.37 ± 0.11 | NA | 0.88 |

HET = Heterocera, ROP = Ropalocera, MIC = Microlepidoptera, GEO = Geometridae, ARC = Arctiidae

Table S2. Predictive allometries for bee foraging distance and proboscis length (to be added). HW: Head width, IT: Intertegular distance.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source** | **Metric** | **Taxa** | **Measure** | **Equation** |
| van Nieuwstadt & Iraheta (1996) | Foraging distance | Stingless bees (Apidae:Meliponini) | Artificial nectar source |  |
|  |  |  | “ |  |
| Greenleaf et al. (2007) |  | Apidae | Max |  |
|  |  |  | Typical |  |
|  |  |  | Feeder |  |
|  |  |  | Comm |  |

**#SUPP# Preservative time**

**Method**

A key confounding factor which can affect predictive allometric models for insects is the time specimens spend within preservative (i.e. ethanol Leuven et al. 1985). As a trade-off between including greater species diversity and those that had been preserved, we assessed the impact of preservative time using Australian and German specimens (species n = 20), where there was considerable overlap in preserved and unpreserved specimens in species.

We fitted a linear-mixed effect model with y = Specimen weight ~ IT + Sex + preservative time with two random terms: country and species.

We also fitted a linear-mixed effect model with y = log(Specimen weight) ~ IT + Sex + preservative time with two random terms: country and species.

**Result**

Depending on model choice, there was a insignificant or significant effect of preservative time on specimen weight. Not logged: t-value 0.53, logged: t-value 4.1, - loss of 0.0001 per day

**##SUPP## Sample size variation**

ANY IDEAS FOR A TEST FOR THIS OR IF A TEST IS NEEDED? – I THINK PERHAPS WE LEAVE THIS PART OUT COMPLETELY

Sample size exhibited an interesting trend in relation to both ITD and dry weight. In bees, mean ITD and dry weight stabilised within the confidence intervals of the total sample size with >20 specimens per species (Figure S1A). For hoverflies, the lower overall sample sizes of each species limited inference of sample-size / mean stabilisation (FIGURE S1B).



Fig S1A. Intraspecific variation in IT and dry weight in relation to sample size in bees. Red line denotes the total trait mean and green lines represent 95% confidence intervals.



Fig S1B. Intraspecific variation in IT and dry weight in relation to sample size in hoverflies. Red line denotes the total trait mean and green lines represent 95% confidence intervals.

**References**

Angilletta Jr, M. J., T. D. Steury and M. W. Sears (2004). "Temperature, growth rate, and body size in ectotherms: fitting pieces of a life-history puzzle." Integrative and comparative biology **44**(6): 498-509.

Ashton, K. (2004). "Comparing phylogenetic signal in intraspecific and interspecific body size datasets." Journal of evolutionary biology **17**(5): 1157-1161.

Barton, K. and M. K. Barton (2018). "Package ‘MuMIn’."

Bates, D., M. Maechler, B. Bolker and S. Walker (2014). "lme4: Linear mixed-effects models using Eigen and S4." R package version **1**(7): 1-23.

Benke, A. C., A. D. Huryn, L. A. Smock and J. Bruce Wallace (1999). "Length-mass relationships for freshwater macroinvertebrates in North America with particular refernce to the Southeastern United States." Journal of the North American Benthological Society **18**(3): 308-343.

Blomberg, S. P., T. Garland Jr and A. R. Ives (2003). "Testing for phylogenetic signal in comparative data: behavioral traits are more labile." Evolution **57**(4): 717-745.

Boria, R. A., L. E. Olson, S. M. Goodman and R. P. Anderson (2014). "Spatial filtering to reduce sampling bias can improve the performance of ecological niche models." Ecological Modelling **275**: 73-77.

Boyce, M. S., P. R. Vernier, S. E. Nielsen and F. K. Schmiegelow (2002). "Evaluating resource selection functions." Ecological modelling **157**(2-3): 281-300.

Brady, C. J. and R. A. Noske (2006). "Generalised regressions provide good estimates of insect and spider biomass in the monsoonal tropics of Australia." Australian Journal of Entomology **45**(3): 187-191.

Brown, J. H., J. F. Gillooly, A. P. Allen, V. M. Savage and G. B. West (2004). "Toward a metabolic theory of ecology." Ecology **85**(7): 1771-1789.

Burgherr, P. and E. I. Meyer (1997). "Regression analysis of linear body dimensions vs. dry mass in stream macroinvertebrates." Archiv für Hydrobiologie **139**(1): 101-112.

Cane, J. H. (1987). "Estimation of bee size using intertegular span (Apoidea)

" Journal of Kansas Entomological Society **60**(1): 145-147.

Cariveau, D. P., G. K. Nayak, I. Bartomeus, J. Zientek, J. S. Ascher, J. Gibbs and R. Winfree (2016). "The Allometry of Bee Proboscis Length and Its Uses in Ecology." Plos One **11**(3).

Chown, S. L. and K. J. Gaston (2010). "Body size variation in insects: a macroecological perspective." Biological Reviews **85**(1): 139-169.

Davidowitz, G., L. J. D’Amico and H. F. Nijhout (2004). "The effects of environmental variation on a mechanism that controls insect body size." Evolutionary Ecology Research **6**(1): 49-62.

DeLong, J. P., B. Gilbert, J. B. Shurin, V. M. Savage, B. T. Barton, C. F. Clements, A. I. Dell, H. S. Greig, C. D. Harley and P. Kratina (2015). "The body size dependence of trophic cascades." The American Naturalist **185**(3): 354-366.

Dietze, M. C., A. Fox, L. M. Beck-Johnson, J. L. Betancourt, M. B. Hooten, C. S. Jarnevich, T. H. Keitt, M. A. Kenney, C. M. Laney, L. G. Larsen, H. W. Loescher, C. K. Lunch, B. C. Pijanowski, J. T. Randerson, E. K. Read, A. T. Tredennick, R. Vargas, K. C. Weathers and E. P. White (2018). "Iterative near-term ecological forecasting: Needs, opportunities, and challenges." Proceedings of the National Academy of Sciences.

Ehnes, R. B., B. C. Rall and U. Brose (2011). "Phylogenetic grouping, curvature and metabolic scaling in terrestrial invertebrates." Ecology Letters **14**(10): 993-1000.

Ganihar, S. R. (1997). "Biomass estimates of terrestrial arthropods based on body length." Journal of Bioscience **22**(2).

Garland, J., Theodore and A. R. Ives (2000). "Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods." The American Naturalist **155**(3): 346-364.

Garland, T., A. F. Bennett and E. L. Rezende (2005). "Phylogenetic approaches in comparative physiology." Journal of experimental Biology **208**(16): 3015-3035.

Gould, S. J. (1966). "Allometry and size in ontogeny and phylogeny." Biological Reviews **41**(4): 587-638.

Gowing, G. and H. F. Recher (1984). "Length-wight relationships for invertebrates from forests in south-eastern New South Wales." Austral Ecology **9**(1): 5-8.

Greenleaf, S. S., N. M. Williams, R. Winfree and C. Kremen (2007). "Bee foraging ranges and their relationship to body size." Oecologia **153**(3): 589-596.

Hagen, M. and Y. L. Dupont (2013). "Inter-tegular span and head width as estimators of fresh and dry body mass in bumblebees (Bombus spp.)." Insectes Sociaux **60**(2): 251-257.

Harris, D. J., S. D. Taylor and E. P. White (2018). "Forecasting biodiversity in breeding birds using best practices." PeerJ **6**: e4278.

Harrison, J. F., C. Klok and J. S. Waters (2014). "Critical PO 2 is size-independent in insects: implications for the metabolic theory of ecology." Current opinion in insect science **4**: 54-59.

Harvey, P. H. and M. D. Pagel (1991). The comparative method in evolutionary biology, Oxford university press Oxford.

Hedtke, S. M., S. Patiny and B. N. Danforth (2013). "The bee tree of life: a supermatrix approach to apoid phylogeny and biogeography." BMC Evolutionary Biology **13**(1): 138.

Henschel, J. R. and M. K. Seely (1997). "Mass-length relationships of Namib Tenebrionids." Madoqua **19**(2): 159-160.

Hijmans, R. J. (2012). "Cross‐validation of species distribution models: removing spatial sorting bias and calibration with a null model." Ecology **93**(3): 679-688.

Hodar, J. A. (1997). "The use of regression equations for the estimation of prey length and biomass in diet studies of insectivore vertebrates." Miscellania Zoologica **20**(2): 1-10.

Huxley, J. and J. S. Huxley (1993). Problems of relative growth.

Johnson, M. D. and A. M. Strong (2000). "Length-weight relationships of Jamaican arthropods." Entomological News **111**(4): 270-281.

Kapustjanskij, A., M. Streinzer, H. Paulus and J. Spaethe (2007). "Bigger is better: implications of body size for flight ability under different light conditions and the evolution of alloethism in bumblebees." Functional Ecology **21**(6): 1130-1136.

Karachle, P. K. and K. I. Stergiou (2012). Morphometrics and allometry in fishes, INTECH Open Access Publisher.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai, Montreal, Canada.

Leuven, R. S., T. C. Brock and H. A. van Druten (1985). "Effects of preservation on dry-and ash-free dry weight biomass of some common aquatic macro-invertebrates." Hydrobiologia **127**(2): 151-159.

Martin, C. A., R. Proulx, P. Magnan, C. Dytham and J. M. Lobo (2014). "The biogeography of insects' length-dry mass relationships." Insect Conservation and Diversity **7**(5): 413-419.

Martins, E. P. and T. Garland Jr (1991). "Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study." Evolution: 534-557.

Martins, E. P. and E. A. Housworth (2002). "Phylogeny shape and the phylogenetic comparative method." Systematic biology **51**(6): 873-880.

Mitchell, M., B. Muftakhidinov, T. Winchen, Z. Jędrzejewski-Szmek, T. G. Badger, badshah400 and A. Wilms. (2018). "Engauge Digitizer Version 10.6." from <http://markummitchell.github.io/engauge-digitizer>.

Olesen, J. M., J. Bascompte, Y. L. Dupont, H. Elberling, C. Rasmussen and P. Jordano (2010). "Missing and forbidden links in mutualistic networks." Proceedings of the Royal Society of London B: Biological Sciences: rspb20101371.

Oliveira, R. and C. Schlindwein (2010). "Experimental demonstration of alternative mating tactics of male Ptilothrix fructifera (Hymenoptera, Apidae)." Animal Behaviour **80**(2): 241-247.

Pagel, M. (1999). "Inferring the historical patterns of biological evolution." Nature **401**(6756): 877.

Paradis, E., J. Claude and K. Strimmer (2004). "APE: analyses of phylogenetics and evolution in R language." Bioinformatics **20**(2): 289-290.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, S. Heisterkamp, B. Van Willigen and R. Maintainer (2017). "Package ‘nlme’." Linear and nonlinear mixed effects models: 3-1.

Rall, B. C., G. Kalinkat, D. Ott, O. Vucic‐Pestic and U. Brose (2011). "Taxonomic versus allometric constraints on non‐linear interaction strengths." Oikos **120**(4): 483-492.

Ramalho, M., V. Imperatriz-Fonseca and T. Giannini (1998). "Within-colony size variation of foragers and pollen load capacity in the stingless bee Melipona quadrifasciata anthidioides Lepeletier (Apidae, Hymenoptera)." Apidologie **29**: 221-228.

Revell, L. J. (2012). "phytools: an R package for phylogenetic comparative biology (and other things)." Methods in Ecology and Evolution **3**(2): 217-223.

Rogers, L., R. Buschbom and C. Watson (1977). "Length-weight relationships of shrub-steppe invertebrates." Annals of the Entomological Society of America **70**(1): 51-53.

Rudolf, V. H. and N. L. Rasmussen (2013). "Ontogenetic functional diversity: size structure of a keystone predator drives functioning of a complex ecosystem." Ecology **94**(5): 1046-1056.

Sabo, J. L., J. L. Bastow and M. E. Power (2002). "Length-mass relationships for adult aquatic and terrestrial invertebrates in a California watershed." J. N. Am. Bethol. Soc. **21**(2): 336-343.

Sage, R. D. (1982). "Wet and dry-weight estimates of insects and spiders based on length." The American Midland Naturalist **108**(2): 407-411.

Sample, B. E., R. J. Cooper, R. D. Greer and R. C. Whitmore (1993). "Estimation of insect biomass by length and width." The American Midland Naturalist **129**(2): 234-240.

Schoener, T. W. (1980). "Length-weight regressions in tropical and temperate forest-understory insects." Annals of the Entomological Society of America **73**(1): 106-109.

Schramski, J. R., A. I. Dell, J. M. Grady, R. M. Sibly and J. H. Brown (2015). "Metabolic theory predicts whole-ecosystem properties." Proceedings of the National Academy of Sciences **112**(8): 2617-2622.

Spaethe, J. and L. Chittka (2003). "Interindividual variation of eye optics and single object resolution in bumblebees." Journal of Experimental Biology **206**(19): 3447-3453.

Speakman, J. R. (2005). "Body size, energy metabolism and lifespan." Journal of Experimental Biology **208**(9): 1717-1730.

Stang, M., P. G. Klinkhamer, N. M. Waser, I. Stang and E. van der Meijden (2009). "Size-specific interaction patterns and size matching in a plant–pollinator interaction web." Annals of Botany **103**(9): 1459-1469.

Stevens, V. M., A. Trochet, H. Van Dyck, J. Clobert and M. Baguette (2012). "How is dispersal integrated in life histories: a quantitative analysis using butterflies." Ecology letters **15**(1): 74-86.

Stone, M. (1974). "Cross-validatory choice and assessment of statistical predictions." Journal of the royal statistical society. Series B (Methodological): 111-147.

Streinzer, M., W. Huber and J. Spaethe (2016). "Body size limits dim-light foraging activity in stingless bees (Apidae: Meliponini)." Journal of Comparative Physiology A **202**: 643-655.

Team, R. C. (2018). "R: A language and environment for statistical computing."

Teder, T., T. Tammaru and T. Esperk (2008). "Dependence of phenotypic variance in body size on environmental quality." The American Naturalist **172**(2): 223-232.

Trites, A. W. and D. Pauly (1998). "Estimating mean body masses of marine mammals from maximum body lengths." Canadian Journal of Zoology **76**(5): 886-896.

Van Nieuwstadt, M. and C. R. Iraheta (1996). "Relation between size and foraging range in stingless bees (Apidae, Meliponinae)." Apidologie **27**(4): 219-228.

Velghe, K. and I. Gregory-Eaves (2013). "Body size is a significant predictor of congruency in species richness patterns: a meta-analysis of aquatic studies." PloS one **8**(2): e57019.

Wardhaugh, C. W. (2013). "Estimation of biomass from body length and width for tropical rainforest canopy invertebrates." Australian Journal of Entomology **52**(4): 291-298.

Wenger, S. J. and J. D. Olden (2012). "Assessing transferability of ecological models: an underappreciated aspect of statistical validation." Methods in Ecology and Evolution **3**(2): 260-267.

White, E. P., S. M. Ernest, A. J. Kerkhoff and B. J. Enquist (2007). "Relationships between body size and abundance in ecology." Trends in ecology & evolution **22**(6): 323-330.

Williams, N. M., E. E. Crone, T. a. H. Roulston, R. L. Minckley, L. Packer and S. G. Potts (2010). "Ecological and life-history traits predict bee species responses to environmental disturbances." Biological Conservation **143**(10): 2280-2291.

Woodward, G. and A. G. Hildrew (2002). "Body‐size determinants of niche overlap and intraguild predation within a complex food web." Journal of Animal Ecology **71**(6): 1063-1074.