Pollimetry: Predictive allometry for pollinating insects

Liam K. Kendall, Romina Rader, Vesna Gagic, Daniel Cariveau, Matthias Albrecht, Katherine Baldock, Andrea Holzschuh, Juanita Rodriguez, Louis Suter and Ignasi Bartomeus

**Abstract**

Allometric scaling laws have key implications for the conservation and management of pollinating insects in both managed and unmanaged ecosystems.

Body size (BS) can predict influential ecological traits yet available predictive models are outdated, rely upon geographically restricted sampling and have limited applicability for non-bee taxa.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 catalogued existing predictive allometries for pollinating insects (Hymenoptera (BS: 38), Diptera (BS: 26) and Lepidoptera (BS: 21) and improved upon pre-existing equations for estimating body size in key pollinating taxa (bees and hoverflies).

We measured dry weight and intertegular distance (ITD) of bees (species total: 298) and hoverflies (species total: 103) across three biogeographic regions: Australia, Europe and USA.

We then used linear mixed effect (LME) and phylogenetic generalised least squares (PGLS) models to construct a suite of state-of-the-art equations for estimating interspecific pollinator body size.

We used OLS regression to assess intraspecific ITD - body size predictions

Interspecific model validation was assessed using k-fold cross validation.

Overall differences between these models were minimal with PGLS models performing similarly to LME models. Intraspecific models found ITD a reliable proxy for body size in bees but not hoverflies, potentially due to sample size issues.

These highly applicable models form the R package 'pollimetry’ and provide an updated resource for allometric research concerning wild and managed pollinators globally.

**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; Dell et al. 2011, 2014; Velghe & Gregory-Eaves 2013). 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, primary productivity and pollination (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. 1976; Henschel & Seely 1997). Second, where research occurs in remote field sites, equipment limits can prevent direct measurements (Brady & Noske 2006). Thirdly, in diet/food web studies, body size estimates come from digested prey items (e.g. Hodar 1997). Lastly, 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 using a co-varying trait, as well as allowing body size proxies, has emerged across many biological disciplines.

Body length – body size models have been used extensively to predict body size for fish (e.g. Karachle & Stergiou 2012), mammals (e.g. Trites & Pauly 1998) and both aquatic (e.g. Burgherr & Meyer 1997; Benke 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) and this has led to the proliferation of multiple models for a wide range of taxa worldwide, especially insects. However, when compared, these models show significantly different coefficients both within- and between insect orders (Schoener 1980; Sample 1993; Ganihar 1997; Benke et al. 1999; Brady & Noske 2006), often due to biogeography (i.e. latitude, see Martins et al. 2014), and/or methodological influences such as sampling biases and model choice.

These differences highlight a need for consolidation and improvement in predictive allometric theory and practice. Predictive models require a robust and iterative framework in model choice, development and validation. Ordinary least squares (OLS) regression has been seen as ideal for prediction (INSREF from Cariveau) leading to a lack of incorporation of mixed effects and/or phylogenetic model structures (e.g. phylogenetic generalized least squares (PGLS) (Harvey & Pagel 1991)) despite their importance in explaining key ecological and evolutionary processes (Bolker?). Further, model validation techniques, such as cross-validation, which are common-place in statistics and medical sciences (eg. TWO REFS), have been overlooked in predictive ecology, with the exception of ecological distribution modelling (eg.Wenger and Olden 2012, Boria et al 2014). It is becoming clear these methods, in model building and testing, are necessary given the increasing burden of proof in biological prediction. Further, ‘iterative model-building’, whereby equations can be periodically updated rather than differentiated and replaced, represents an as-yet untested avenue for greater accuracy and wider application.

A number of key pollination traits exhibit allometric scaling. In bees, body size affects insect activity rates/periods (Strienzer et al. 2015), pollen load (e.g. Ramalho et al. 1998), foraging range (e.g. Greenleaf et al. 2007 and van Nieuwstadt & Iraheta 1996), proboscis length (Cariveau et al. 2016) and functional trait matching (Stang et al 2009; Olesen et al., 2010). Despite these influences, few predictive allometric 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 model was developed with a sample of 20 single females from solitary bee species in North America that represented six major bee families. It is now the most commonly used metric for estimating bee body size (Web of Science: 89 citations, Google scholar: 108 citations as of…? ) and has used in ecological (eg. 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 the 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).

The utility of Cane’s equation has not previously been tested beyond North American solitary bee species, except for in bumblebees (Hagen & Dupont 2013), in conjunction with biogeography or within more complex model structures that consider mixed effects or species’ phylogenies. Neither has ITD been assessed in other key pollinating taxa, such as hoverflies (Diptera: Syrphidae). Therefore, we aimed to develop predictive allometric equations within 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 a unified resource, an *R* package, entitled “pollimetry”.

**METHODOLOGY**

*Existing equations*

We selected three key pollinating insect orders: Diptera, Hymenoptera and Lepidoptera and collated all known predictive allometric models using a systematic literature search.

*Specimen collection and measurements*

Only recently curated (<5 years)- or fresh- undamaged specimens were included. For every included specimen, we obtained preservative time, sample location (latitude and longitude), collection method (pan trap, sweeping, malaise trap) and taxonomic designation. Cane (1989)’s original data was obtained using Engauge Digitizer version (INSREF)

*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.

Intertegular distance was measured in millimetres using a stereo-microscope, either mounted with a calibrated scale or microscope camera.

**Data analysis - Model structure**

For each taxon, we constructed species mean datasets stratified for measurer and country of origin. We used a power function in model formulation which is typical of predictive allometry:

Ln(Y) = ln(a) + b\*ln(IT) + c\*ln(IT)

We extended this formula to include multiple interactions with IT: sex, biogeographic region and taxonomic family. We constructed linear mixed effect models to predict body size, over linear regression as species overlap between measurers and within biogeographic regions required a more-complex model structure. Both measurer and species were included as random terms.

**Incorporating phylogeny**

We explored the influence of phylogenetic relatedness on predicting pollinator body size using a simplified mean dataset, with a single species mean per region. Sex was not considered in these models. Overlapping species: in bees; European honeybee (*Apis mellifera*) and the sweat bee (*Halictus rubicundus*) present in multiple regions, were removed from their introduced regions, Australia and North America respectively. For bees, we used an apoid genera tree (Hedtke et al. 2013). Non-represented genera were pruned, using the *ape* package (Paradis et al., 2004) and species added to genera tips as pure-birth subtrees using *phytools* (Revell et al. 2017). As such, we made the explicit assumption that phylogenetic patterns in body size were assessed at and above the genera level.

For hoverflies, we used genera tree from Skevington et al. unpublished.

We assessed the significance of phylogenetic signal of both body size and ITD using Pagel’s λ (Pagel 1999) and the ‘phylosig’ function (Revell et al. 2017). To assess if incorporating phylogeny improved body size predictions, we then determined the relationship between dry weight and ITD and biogeographic region using phylogenetic generalized least squares (PGLS) regression. λ was fitted at an initial value of 0.5 and optimized by maximum likelihood.

**Model selection and cross-validation**

We first fitted the full model with all predicted explanatory variables, for LME: ITD in interaction with family, region and sex. For PGLS: IT in interaction with region. We then performed model selection assessing all subset models using the ‘dredge’ function within the R library MuMIn (Barton 2017). The best fitting models were then ranked by Akaike Information Criterion (AIC). Given our predictive framework, we iteratively removed terms region and sex from LME models for both taxa for wider utility. Lastly, we considered ITD in isolation.

**Cross-validation**

We implemented k-fold cross validation to test overall model performance and compare prediction error. 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), and then compared against the 5-k set (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 *R2* and AIC across the five sets. For PGLS models, λ for each model was fixed at the optimised value from the full dataset. Simplified k-fold cross validation was used to test accuracy of Cane’s (1989) original equation. We tested his equation against every test set individually to derive a mean RMSE across sets.

**Intraspecific predictions and variance**

We assessed intraspecific predictions and sample size variation in trait measurements. For the five most abundant species of both bees and hoverflies irrespective of sex ~~(Bees:~~ *~~Homalictus urbanus~~* ~~(n = 251),~~ *~~Lasioglossum pauxillum~~* ~~(n = 113),~~ *~~Bombus lucorum~~* ~~(n = 111),~~ *~~Andrena flavipes~~* ~~(n = 75) and~~ *~~Lasioglossum lanarium~~* ~~(n = 68); Hoverflies:~~ *~~Helophilus parallelus~~* ~~(n = 19),~~ *~~Sphaerophoria macrogaster~~* ~~(n = 17),~~ *~~Episyrphus balteatus~~* ~~(n = 15),~~ *~~Melanostoma mellinum~~* ~~(n = 12) and~~ *~~Syritta pipiens~~* ~~(n = 12)),~~ we tested the utility of ITD in predicting intraspecific female body size variation using species-level OLS regression. Furthermore, we plotted species trait means independently against increasing sample size to estimate the adequate sample size whereby variance stabilised within confidence intervals of the actual sample size. ANY IDEAS FOR A TEST FOR THIS OR IF A TEST IS NEEDED?

**Results**

*Existing 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. Surprisingly, there were no equations for Syrphidae.

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 1989 & 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 6 models for bee foraging distance and three models for bee proboscis length.

**Species and specimen distribution**

In total, we measured 298 bee species from Australia, North America and Europe, including Cane’s (1989) original 20 species and 103 hoverfly species from Australia and Europe. Five out of six bee families and both syrphid subfamilies were represented.

**Interspecific model selection and performance**

For bees, our non-phylogenetic 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). Family and region were more informative than sex however for hoverflies, including subfamily was less important than region and sex across the best fitting models. The hoverfly ITD-only model ranked higher than the best-fitting model that included subfamily and ITD (Table 1B, Model 5).

Phylogenetic signal was highly significant (*p* <0.001) for bee ln body size (λ: 0.71) and ln ITD (λ: 0.75). PGLS models exhibited similar trends to LME models with region-specific models best-explaining the body size ~ ITD relationship. Further, PGLS models showed considerable decrease in AIC relative to standard GLS models (ITD \* Region: Δ 25.47, λ 0.54 ; ITD + Region: Δ 25.89, λ 0.54; ITD Δ 43.28, λ: 0.6).

For hoverflies – TBC – waiting on COI phylogeny

Cross-validation indicated that high accuracy persisted across all tested models for both bees and hoverflies. Overall 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 1, Figure 2). In hoverflies, there was less overall dispersion in prediction error. Interestingly, the ‘best’ LME model differed on the basis of RMSE, *R2* and AIC in both data sets. Incorporating phylogeny into model-fitting resulted in marginal increase 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. All new models were considerably better performing than Cane’s (1989) original equation.

**Intra-specific variation**

Across the five most abundant species of bees and hoverflies (females only), intraspecific predictions of body size using ITD were mixed (Figure 2). All bee species exhibited a significant relationships between ITD and dry weight, however adjusted *R2*varied considerably from 0.02 in *Homalictus urbanus* to 0.46 for *Lasioglossum lanarium* (Table 3, Figure 3). In contrast to bees, Three hoverfly species, Austrosyrphus spp, *H. parallelus and M. 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 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 either body length, body length \* body width (in the case of pre-existing equations) and ITD using Cane’s (1989) equation or our new equations. Also included are Greenleaf et al’s (2007) and van Nieuwstadt & Iraheta’s (1996) allometric equations for estimating foraging distance in bees using ITD or head width, as well as Cariveau et al. (2016)’s allometric equations for estimating bee tongue length.

**#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. 1982??). 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 = ln(specimen weight) ~ preservative time with two random terms: country and species.

**Result**

Across twenty species, preservative time was found to exhibit an effect upon specimen weight, accounting for a loss of 0.006mg per day preserved (t-value: -3.050).

**DISCUSSION**

Traditional predictive allometry has been limited in its applicability due to sampling biases and statistical

Herein, we defined a novel iterative framework in which we described and tested a suite of predictive allometric models which considered different model structures and influences upon body size variation.

We demonstrated clear and unequivocal proof of the utility of the intertegular distance (ITD) in predicting body size in two key pollinating taxa: bees and hoverflies, as first demonstrated by Cane (1989) for bees. 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 or phylogeny we improved model performance and overcame the limitations of traditional predictive allometry. These three predictors represent fundamental 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 differences in body size proportions and mass accumulation between all three regions. Importantly, region was as or more important than both sex or taxonomy in our LME models.

Martin et al (2014) contributed differences in body-length allometric coefficients between geographic regions to a latitudinal gradient, suggesting comparable geographic regions should exhibit similar allometric trends. However, previous studies have found both similarities (Gowing & Recher 1984 and Rogers et al. 1977) and differences in allometric co-efficients between regions (Schoener 1980 and Rogers et al. 1977) and only one prior predictive allometric study examined multiple biogeographic regions in concert (Schoener 1980). These differences likely rise from differing biogeographic diversification patterns as well as sampling biases. ~~This raises doubts regarding the biological reasons behind these differences.~~ Both bees and hoverflies are highly diverse groups; in Australia alone, we included <1% of total estimated diversity of 2000 bee species (INSREF) and X% of the described hoverfly species (REF). Therefore, allometric co-efficient differences between regions are likely a result of both sampling biases as well as differing distribution patterns. Indeed, ITD may be more robust to these influences than body length given the subtlety of regional differences in BS~ITD we observed. These limitations suggest that for predictive allometry to exhibit greater accuracy, specimens from different biogeographical regions should be incorporated into model development and validation.

Sexual size dimorphism (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). This SSD is hypothesised to be a result of greater fitness nad increased fecundity inferred by larger body size in females (Teder and Tammaru 2005: Stillwell 2010). Sex was retained as an integral predictor either in addition or in interaction with ITD suggesting 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.

In bees, such dimorphism is attributed to physical requirements of nest provisioning and nest construction. As such, it is in more pronounced in ground nesting and/or provisioning taxa than non-provisioning and/or cavity nesting taxa as well as eusocial species (Shreeves and Field 2008; 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 are few examples of specialised morphological sexual dimorphism, other than eye size. Despite this, our results suggest morphological differences are significant. Therefore, as ITD is sensitive to sexual dimorphism in both bees and hoverflies, consideration of specimen’s gender is integral for more accurate prediction of body size and allometric traits.

*Intra-specific variation and sample size*

Previous studies have cast doubt over the utility of predictive models to accurately describe intraspecific allometric traits (e.g. Hagen & Dupont 2013; Cariveau et al. 2016). bOMBUS refernce. Our results provide reasonable proof that such an aim is species-specific and error-ridden. Further, given the lack of utility for hoverflies, except for *Helophilus parallelus,* this suggests it is sensitive to sample size. If considerations are made at the population level, sample sizes of >20 individuals are adequate to accurately estimate a species-level allometric co-efficient and extrapolate to a population based off ITD. However, if sample sizes are not adequate, use of an interspecific model or direct measurement are advisable. Our sample sizes of hoverfly species were insufficient to properly test either hypothesis. Given showed a significant trend suggests this may be possible across hoverflies with adequate sample size. In any case, the interspecific model will provide adequate estimations of body size distributions within a given population.

Complex model structures within an iterative framework provide a new and much needed re-invigoration for the field of predictive allometry. LME structures allowed us to accurately account for species overlap within regions to assess macro-ecological trends. Phylogenetic signal in body size distributions has been inferred in a number of vertebrate and invertebrate groups (e.g. ), including bees. Incorporating phylogeny considerably decreased AIC in our predictive PGLS models relative to GLS models. This suggests that PGLS methods can be both more informative and more accurate. Failing to account for these phylogenetic patterns heightens the risk of inaccurate predictions, which is a key tenet of the PGLS method (Martins 1991; Martins et al. 2002; Garland et al. 2005). Where phylogenetic information is unavailable, including taxonomy represents an adequate compromise. Including taxonomic family within our LME bee models improved predictive accuracy however in hoverflies, including subfanily was less informative in describing interspecific body size variation. Taxonomy has been used as a phylogenetic proxy before (Cariveau et al. 2016) but our results suggest it should only be used where correct or up-to-date phylogenies are unavailable, and its use should be assessed in different taxa.

An important yet underutilised aspect of predictive ecology is estimating model performance on untested data. By dividing our data into sequential validation and testing sets, we were able to assess how accuracy differs between sets, each comprising a different random subset of our entire dataset. Such an approach is used in distribution modelling in order to….

How is it used in distribution modelling? Developed for what ends

The Akaike Information Criterion (AIC) is seen as an appropriate metric in order to avoid model over-fitting as well as judge model performance. Cross-validation is important given the differential results it can provide between the full and subsetted datasets. For example, considering only AIC in cross validation of our hoverfly models, the ITD only model was better performing within random subsetted data in contrast with our full dataset. Few examples exist for direct comparison of this. By considering multiple metrics (AIC, RMSE and CV *R2*) a better predictive picture is produced and disseminating the best model becomes a hypothesis-driven formula. 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.

~~However, two issues remain from this approach, i) how to define ‘the best’ cross-validated model and ii), how to adequate compare cross-validation results with the full dataset.~~

~~Example 1~~

~~Example 2~~

Model averaging has been suggested where models differ within a delta of 10. In both our studied taxa, this wouold have resulted in averaging across X and X models respectively.

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 highlighting the need for 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 potential allometric traits at either the intra- or inter-specific level. The consequence of size is ubiquitous within pollination research yet few have utilised allometric theory in pollinating taxa beyond bees. The prescribed iterative framework heralds a dynamic new direction for predictive allometry and provides greater prediction through hypothesis-led model choice, testing and investigation in allometric pollination research.

Table 1. 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 |
| Garcia-Barros (2015) | |  | |  | | |  | Worldwide | | | \*\*\* |  | MU | | - | - | - | - | **0.96** |

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

Table 2. Predictive allometries for bee foraging distance. HW: Head width, IT: Intertegular distance.

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

Table 3. AIC and delta 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.

1. Bees

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Explanatory variables | AIC | Δ |
| LME | 1. ITD \* Family + ITD \* Region + ITD \* Sex | **430.1** | 0 |
|  | 2. ITD \* Family + ITD \* Region + Sex | 433.0 | 2.9 |
|  | 3. ITD \* Family + ITD \* Region | 445.9 | 15.8 |
|  | 4. ITD \* Family + ITD \* Sex | 493.3 | 63.1 |
|  | 5. ITD \* Family | 509.2 | 79.1 |
|  | 6. ITD | 546.6 | 116.5 |
| PGLS | ITD \* Region | **195.1** | 0 |
|  | ITD + Region | 215.7 | 20.6 |
|  | ITD | 233.2 | 38.1 |

1. Hoverflies

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Explanatory variables | AIC | Δ |
| LME | 1. ITD + Sex | **194.7** | 0 |
|  | 2. ITD + Region + Sex | 194.8 | 0.1 |
|  | 3. ITD + Region + Sex + Subfamily | 195.2 | 0.5 |
|  | 4. ITD + Sex + Subfamily | 195.3 | 0.6 |
|  | 5. ITD \* Sex | 195.4 | 0.7 |
|  | 6. ITD \* Sex + Region | 195.5 | 0.8 |
|  | 7. ITD + Subfamily | 199.6 | 4.9 |
|  | 8. ITD | 199.6 | 4.9 |

Table 4. 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.

1. Bees

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | | Explanatory variables | RMSE | *R2* | AIC |
| LME | 1. ITD \* Family + ITD \* Region + ITD \* Sex | | **0.3816629** | **0.8903881** | 396.205 |
|  | 2. ITD \* Family + ITD \* Region + Sex | | 0.383455 | 0.8893434 | **394.8568** |
|  | 3. ITD \* Family + ITD \* Region | | 0.3877454 | 0.8865717 | 400.3984 |
|  | 4. ITD \* Family + ITD \* Sex | | 0.4154217 | 0.8708062 | 431.536 |
|  | 5. ITD \* Family | | 0.4221665 | 0.8663327 | 436.4804 |
|  | 6. ITD | | 0.4317461 | 0.8588888 | 446.232 |
| PGLS | 1. ITD \* Region | | **0.3776816** | **0.894051** | **157.675** |
|  | 2. ITD + Region | | 0.3900165 | 0.886211 | 175.5921 |
|  | 3. ITD | | 0.3892253 | 0.886775 | 174.6972 |

1. Hoverflies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Explanatory variables | RMSE | *R2* | AIC |
| LME | 1. ITD + Sex | 0.4187792 | 0.8176169 | 167.9012 |
|  | 2. ITD + Region + Sex | **0.417564** | 0.8179376 | 169.993 |
|  | 3. ITD + Region + Sex + Subfamily | 0.4303686 | 0.8214873 | 173.6362 |
|  | 4. ITD + Sex + Subfamily | 0.4300453 | 0.8211913 | 171.2439 |
|  | 5. ITD \* Sex | 0.4212705 | 0.8154102 | 169.9698 |
|  | 6. ITD \* Sex + Region | 0.4208831 | 0.8158122 | 172.054 |
|  | 7. ITD + Subfamily | 0.4379373 | **0.8325788** | 171.2974 |
|  | 8. ITD | 0.4381055 | 0.8271961 | **167.3067** |

Table 5. 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-efficient, *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 |
| Bees | *Andrena flavipes* | 21.07 (1,57) | 1.308 | 2.029 | 0.257 | <0.001 |
|  | *Bombus lucorum* | 81.15 (1,101) | 1.412 | 1.966 | 0.44 | <0.001 |
|  | *Homalictus urbanus* | 6.055 (1,209) | -0.164 | 1.166 | 0.024 | 0.014 |
|  | *Lasioglossum lanarium* | 53.87 (1,61) | 0.702 | 2.13 | 0.46 | <0.001 |
|  | *Lasioglossum pauxillum* | 43.92 (1,110) | 0.402 | 3.112 | 0.279 | <0.001 |
| Hoverflies | *Austrosyrphus spp.* | 12.7 (1,30) | 0.087 | 2.032 | 0.274 | 0.001 |
|  | *Helophilus parallelus* | 14.84 (1,17) | 0.286 | 2.485 | 0.435 | 0.001 |
|  | *Sphaerophoria macrogaster* | 0.04 (1,8) | 0.361 | 0.195 | -0.11 | N.S. |
|  | *Episyrphus balteatus* | 0.08 (1,8) | 1.334 | 0.885 | -0.11 | N.S. |
|  | *Melanostoma scalare* | 6.38 (1,7) | -2.172 | 7.619 | 0.4 | 0.03 |

**Figure 1. Bee genera tree from Hedtke et al. (2013) with species subtrees added. Colour denotes 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.



Fig 4A. 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 4B. 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.