

# Analysis for Melin et al: ‘Allometric relationships shape foreleg evolution of long-legged oil bees (Melittidae: Rediviva)’

Annalie Melin, Res Altwegg, John C. Manning and Jonathan F. Colville

28 November 2020

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Data preparation</b>	<b>2</b>
<b>3</b>	<b>Estimating the parameters of the static allometric relationships for each species</b>	<b>3</b>
<b>4</b>	<b>Comparing the static and evolutionary allometric relationships</b>	<b>6</b>
4.1	Variation in allometric parameters when <i>R. neliana</i> is excluded . . . . .	10
<b>5</b>	<b>Testing whether static and evolutionary allometric slopes differ</b>	<b>10</b>
<b>6</b>	<b>Do these results change when we include <i>R. gigas</i>?</b>	<b>12</b>
<b>7</b>	<b>Does accounting for phylogeny affect our results?</b>	<b>13</b>
<b>8</b>	<b>Estimating the parameters of the static allometric relationships for each population in <i>R. neliana</i></b>	<b>18</b>
<b>9</b>	<b>Is variation in allometric relationships explained by spur length of the host flowers?</b>	<b>22</b>
<b>10</b>	<b>Is allometric variation among <i>R. neliana</i> populations explained by spur length of the host flowers?</b>	<b>27</b>
	<b>References</b>	<b>32</b>

## 1 Introduction

This document is an appendix to the paper ‘Allometric relationships shape foreleg evolution of long-legged oil bees (Melittidae: Rediviva)’ by Annalie Melin, Res Altwegg, John C. Manning and Jonathan F. Colville, detailing the statistical analyses presented there. The data are published on UCT’s data portal (DOI: 10.25375/uct.13297640).

This study examines narrow-sense allometry of foreleg length in a group of oil-collecting bee species from South Africa. Narrow-sense allometry is defined as

$$Y = a \times X^b \tag{1}$$

where  $Y$  is foreleg length, and  $X$  is a linear measure of body size (inter-tegula distance) in our case. The allometric parameters  $a$  and  $b$  determine the shape of the allometric relationship.

On the log scale, the allometric power law becomes linear,  $\ln(Y) = \ln(a) + b \times \ln(X)$ , and we can estimate the parameters  $\ln(a)$  and  $b$  using linear regression techniques, where  $\ln(a)$  is estimated by the intercept and  $b$  is estimated by the slope.

Different types of allometry are recognised. For us here, two are important: 1) *static allometry* is the allometric relationship measured within species across individuals at the same developmental state (adults in our case), and 2) *evolutionary allometry* is the allometric relationship measured across species.

## 2 Data preparation

We'll need the following R packages:

```
library(ape)
library(phyloilm)
library(geiger)
library(nlme)
library(RColorBrewer)
```

The data set consists of data on 869 individual bees belonging to 25 species in the genus *Rediviva*. Each row corresponds to an individual bee and the next two columns give total leg length and a linear measure of body size (inter-tegula distance). The column 'spur\_length' contains the average spur length of the flowers of the plant species this particular bee species is known to visit. For two species, *R.neliana* and *R.pallidula*, we have enough data to be able to distinguish between populations. The columns 'Taxa' and 'population\_spur\_length' identify the population an individual bee came from, for these two species, and the spur length of the flowers that they visit in these particular locations.

```
beemorph <- read.csv("Rediviva_allom_dataset.csv")
summary(beemorph)
```

```
##      Taxa      leg_length      body_size      spur_length
## Length:869      Min.      : 5.515      Min.      :1.822      Min.      : 0.000
## Class :character 1st Qu.: 7.907      1st Qu.:2.704      1st Qu.: 4.075
## Mode  :character Median : 9.383      Median :2.907      Median : 7.295
##                      Mean  :10.371      Mean  :2.898      Mean   : 7.368
##                      3rd Qu.:12.038      3rd Qu.:3.128      3rd Qu.: 9.929
##                      Max.   :22.665      Max.   :4.378      Max.   :19.429
##
## population_code   population_spur_length
## Length:869      Min.      : 5.320
## Class :character 1st Qu.: 7.070
## Mode  :character Median : 8.810
##                      Mean   : 8.684
##                      3rd Qu.:10.550
##                      Max.    :13.880
##                      NA's    :668
```

Create some variables that make life easier later:

```
beemorph$log.leg <- log(beemorph$leg_length)
beemorph$log.body <- log(beemorph$body_size)
beemorph$Taxa <- as.factor(beemorph$Taxa)
beemorph$SPECIES.no <- as.numeric(beemorph$Taxa)
```

So we have measurements of leg length and body size for a number of individuals for each of 25 species of the genus *Rediviva* and want to estimate the parameters of the static allometric relationship and the evolutionary allometric relationship. The former is the ordinary linear regression line of log(leg length) on log(body size) **within** species and the latter is the same relationship **between** species .

### 3 Estimating the parameters of the static allometric relationships for each species

In a first step, we examine the static allometry for each species, using maximum likelihood methods:

```
m1 <- lm(log.leg ~ Taxa * log.body, data=beemorph)
anova(m1)

## Analysis of Variance Table
##
## Response: log.leg
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## Taxa       24  72.102   3.0043   924.1195 < 2.2e-16 ***
## log.body    1   3.671   3.6708  1129.1422 < 2.2e-16 ***
## Taxa:log.body 24   0.474   0.0198    6.0809 < 2.2e-16 ***
## Residuals   819   2.663   0.0033
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA table above shows that there is variation among taxa in both the allometric intercept and the allometric slope even though the Mean Squares for the former was much larger than for the latter. The model explained 96.6% of the variation in leg length.

We slightly re-parameterise Model `m1` above to estimate the parameters of the static allometry for each species directly and collect the parameter estimates with confidence intervals, from Model `m2`:

```
m2 <- lm(log.leg ~ -1 + Taxa + Taxa:log.body, data=beemorph)
nspecies <- length(unique(beemorph$Taxa)) # the number of species we have
a <- exp(m2$coefficients[1:nspecies])
b <- m2$coefficients[(nspecies+1):(nspecies*2)]
allometries <- data.frame(cbind(a, b))
allometries$lcl.a <- exp(confint(m2)[1:nspecies,1])
allometries$ucl.a <- exp(confint(m2)[1:nspecies,2])
allometries$lcl.b <- confint(m2)[(nspecies + 1):(nspecies*2),1]
allometries$ucl.b <- confint(m2)[(nspecies + 1):(nspecies*2),2]
allometries$loga <- m2$coefficients[1:nspecies]
allometries$loga.se <- sqrt(diag(vcov(m2)))[1:nspecies]
allometries$b.se <- sqrt(diag(vcov(m2)))[(nspecies+1):(nspecies*2)]
```

The models above fitted separate allometric curves to each species, i.e.  $\ln(\text{leg}) = \ln(a) + b \times \ln(\text{body})$ . The slopes estimate  $b$  for each species and exponentiating the intercepts estimates  $a$  for each species. Table S1 below gives the parameter estimates with confidence intervals and Figure S1 plots the fitted relationships.

The variance in slopes among species was 0.0755 which is smaller than the mean of the squared standard error of the species-specific slope estimates, 0.1084, suggesting that the variation in slopes is largely due to sampling variance.

Table S1: Summary of morphological data and allometric parameters for 25 species of *Rediviva* bees. N: sample size (number of individuals); LegLength: mean total length of foreleg length [mm]; LL.SD: standard deviation of leg length; BodySize: inter-tegula distance as a measure of body size; BS.SD: standard deviation of body size; a and b: parameters a of the allometric relationships; lcl.a, ucl.a: lower and upper 95% confidence limits of parameter a; lcl.b, ucl.b: lower and upper 95% confidence limits of parameter b.

	N	LegLength	LL.SD	BodySize	BS.SD	a	b	lcl.a	ucl.a	lcl.b	ucl.b
<i>Rediviva albifasciata</i>	29	5.86	0.17	2.51	0.15	4.60	0.26	3.31	6.38	-0.09	0.62
<i>Rediviva alonsoae</i>	11	8.10	0.38	3.33	0.16	4.93	0.41	2.06	11.77	-0.31	1.14
<i>Rediviva aurata</i>	30	9.07	0.51	2.75	0.14	4.08	0.79	2.71	6.15	0.38	1.19
<i>Rediviva autumnalis</i>	19	13.92	0.43	3.17	0.14	10.34	0.26	5.32	20.11	-0.32	0.83
<i>Rediviva brunnea</i>	15	11.78	0.54	3.18	0.11	9.27	0.21	3.52	24.39	-0.63	1.04
<i>Rediviva colorata</i>	39	10.49	0.65	3.00	0.24	9.29	0.11	7.06	12.25	-0.14	0.36
<i>Rediviva emdeorum</i>	45	20.14	1.39	3.36	0.16	8.76	0.69	5.61	13.70	0.32	1.05
<i>Rediviva gigas</i>	8	10.82	0.47	4.20	0.10	3.43	0.80	0.29	40.23	-0.92	2.52
<i>Rediviva intermedia</i>	19	9.10	0.28	3.26	0.16	5.16	0.48	2.75	9.69	-0.05	1.01
<i>Rediviva intermixta</i>	18	8.05	0.27	2.96	0.17	5.58	0.34	3.33	9.36	-0.14	0.81
<i>Rediviva longimanus</i>	39	17.43	0.91	3.19	0.19	8.91	0.58	6.29	12.63	0.28	0.88
<i>Rediviva macgregori</i>	35	12.86	0.64	3.19	0.16	7.05	0.52	4.58	10.84	0.15	0.89
<i>Rediviva micheneri</i>	41	13.79	0.63	2.84	0.15	7.05	0.64	5.02	9.90	0.32	0.97
<i>Rediviva neliana</i>	147	8.79	1.74	2.46	0.37	3.35	1.06	3.16	3.55	1.00	1.13
<i>Rediviva nitida</i>	31	9.29	0.76	3.02	0.19	3.56	0.87	2.51	5.04	0.55	1.18
<i>Rediviva pallidula</i>	64	10.20	0.53	2.94	0.17	6.54	0.41	5.05	8.48	0.17	0.65
<i>Rediviva parva</i>	50	7.43	0.52	2.85	0.18	3.26	0.78	2.50	4.27	0.53	1.04
<i>Rediviva peringueyi</i>	19	9.74	0.39	3.09	0.17	6.90	0.31	4.01	11.86	-0.17	0.79
<i>Rediviva rhodosoma</i>	24	12.44	0.47	3.08	0.17	6.34	0.60	3.90	10.30	0.17	1.03
<i>Rediviva ruficornis</i>	30	9.54	0.36	2.78	0.09	5.95	0.46	3.18	11.15	-0.15	1.07
<i>Rediviva rufipes</i>	49	7.47	0.32	2.93	0.14	5.33	0.31	3.68	7.73	-0.03	0.66
<i>Rediviva rufocincta</i>	66	8.05	0.52	2.75	0.14	3.56	0.80	2.66	4.77	0.52	1.09
<i>Rediviva saetigera</i>	23	7.09	0.27	2.94	0.13	4.37	0.45	2.40	7.94	-0.11	1.00
<i>Rediviva transkeiana</i>	7	10.01	0.06	2.74	0.10	9.20	0.08	2.78	30.43	-1.10	1.27
<i>Rediviva whiteheadi</i>	11	10.88	0.22	3.15	0.09	11.48	-0.05	2.60	50.67	-1.34	1.25

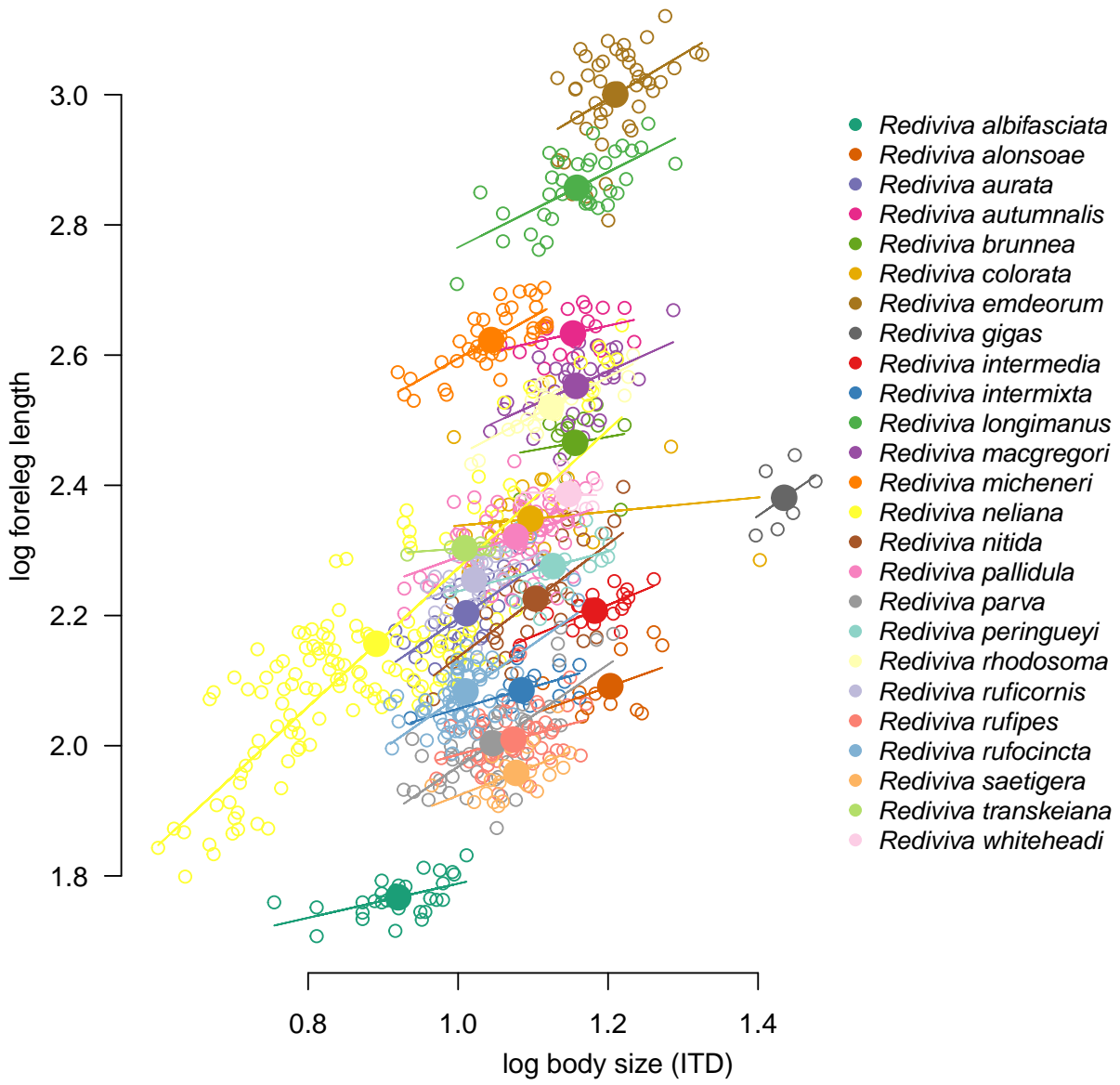


Figure S1: Allometric relationship of leg length against body size in *Rediviva* species. The solid circles indicate the species means.

## 4 Comparing the static and evolutionary allometric relationships

In a next step, we want to estimate the evolutionary allometric relationship and compare it to the static allometric relationship. We also want to estimate the variance among species in the coefficients of their static allometric relationship. We use random effects to estimate these variances.

We use multi-level models (also known as contextual models (Heisler and Damuth 1987; Egset et al. 2011)), to separately estimate the within-species and between-species slope, following van de Pol and Wright (2009):

$$l_{ij} = \beta_0 + \beta_w(b_{ij} - b_{.j}) + \beta_b b_{.j} + \epsilon_{ij} \quad (2)$$

where  $l_{ij}$  is the logarithm of the leg length measurement of the  $i^{th}$  individual of species  $j$ ,  $b_{ij}$  is the logarithm of the body size of that individual,  $b_{.j}$  is the logarithm of the mean body size across all individuals of species  $j$ ,  $\beta_0$  is the intercept,  $\beta_w$  is the within-species effect, i.e. the static allometric slope,  $\beta_b$  is the between-species effect, i.e. the evolutionary allometric slope, and  $\epsilon_{ij}$  are the residuals. The residuals are assumed to follow a normal distribution with a mean of zero and variances to be estimated.

The model in equation 2 assumes a common intercept and static allometric slope across all species. In the next model, we allow the intercept to vary among species by adding species random effects,  $\mu_{0j}$ . These random effects are assumed to follow a normal distribution with a mean of zero and variances to be estimated.

$$l_{ij} = \beta_0 + \beta_w(b_{ij} - b_{.j}) + \beta_b b_{.j} + \mu_{0j} + \epsilon_{ij} \quad (3)$$

Next, we allow the static allometric slopes to vary among species by adding a random slope to Model 3:

$$l_{ij} = (\beta_0 + \mu_{0j}) + (\beta_w + \mu_{wj})(b_{ij} - b_{.j}) + \beta_b b_{.j} + \epsilon_{ij} \quad (4)$$

The code chunk below calculates  $b_{.j}$  and  $b_{ij} - b_{.j}$ :

```
# calculate mean body size per species:
mean.body <- aggregate(beemorph$log.body, by = list(beemorph$Taxa), FUN=mean)
colnames(mean.body) <- c("Taxa", "mean.body")

# add mean body size back onto data frame:
beemorph <- merge(beemorph, mean.body, by.x = "Taxa", by.y = "Taxa", all.x=T)
# difference in body size within species:
beemorph$diff.body <- beemorph$log.body - beemorph$mean.body
```

The evolutionary allometry describes how a trait scales with body size across species. From Figure S1 it is clear that *R. gigas* is an outlier in that respect. This species has the largest body size in this data set but has average leg length. As it also is ecologically quite different from the other species, we exclude it from the main analysis but for completeness report results also with this species included.

```
beemorph.r <- beemorph[beemorph$Taxa != "Rediviva gigas",]
beemorph.r$Taxa <- factor(beemorph.r$Taxa)
beemorph.r$SPECIES.no <- as.numeric(beemorph.r$Taxa)
```

We are now ready to fit the models:

```
m3 <- lm(log.leg ~ diff.body + mean.body, data=beemorph.r) # static allometry same across species
summary(m3)
```

```
##
## Call:
## lm(formula = log.leg ~ diff.body + mean.body, data = beemorph.r)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.54867 -0.19005 -0.00064  0.14524  0.53487
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.37723    0.08918   4.230 2.59e-05 ***
## diff.body    0.84896    0.10571   8.031 3.18e-15 ***
## mean.body    1.81666    0.08437  21.533 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2385 on 858 degrees of freedom
## Multiple R-squared:  0.381, Adjusted R-squared:  0.3796
## F-statistic: 264.1 on 2 and 858 DF, p-value: < 2.2e-16

m4 <- lme(log.leg ~ diff.body + mean.body, random =~1|Taxa, data=beemorph.r) # random intercept
summary(m4)

## Linear mixed-effects model fit by REML
## Data: beemorph.r
##      AIC      BIC  logLik
## -2205.44 -2181.667 1107.72
##
## Random effects:
## Formula: ~1 | Taxa
##      (Intercept)  Residual
## StdDev:   0.2451031 0.06115609
##
## Fixed effects: log.leg ~ diff.body + mean.body
##              Value Std.Error DF   t-value p-value
## (Intercept) 0.1780596 0.6772183 836   0.262928  0.7927
## diff.body    0.8489607 0.0271087 836  31.316885  0.0000
## mean.body    1.9577526 0.6214584  22   3.150255  0.0046
## Correlation:
##      (Intr) dff.bd
## diff.body  0.000
## mean.body -0.997  0.000
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -5.281023741 -0.546657864 -0.007954349  0.532025096  3.694688325
##
## Number of Observations: 861
## Number of Groups: 24

m5 <- lme(log.leg ~ diff.body + mean.body, random =~diff.body|Taxa, data=beemorph.r)
# random intercept and slope
summary(m5)

## Linear mixed-effects model fit by REML
## Data: beemorph.r
##      AIC      BIC  logLik
## -2302.875 -2269.593 1158.438
##
```

```

## Random effects:
## Formula: ~diff.body | Taxa
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev      Corr
## (Intercept) 0.24610390 (Intr)
## diff.body   0.22844234 0.385
## Residual    0.05691609
##
## Fixed effects: log.leg ~ diff.body + mean.body
##           Value Std.Error DF   t-value p-value
## (Intercept) -0.1713919 0.6488697 836 -0.264139 0.7917
## diff.body    0.5399429 0.0637053 836  8.475640 0.0000
## mean.body    2.2792734 0.5952827  22  3.828893 0.0009
## Correlation:
##           (Intr) dff.bd
## diff.body  0.069
## mean.body -0.997 -0.047
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -3.51719191 -0.52455906 -0.01995853  0.56459790  3.59964321
##
## Number of Observations: 861
## Number of Groups: 24

```

The residual standard error decreases markedly, from 0.238 in Model **m3** to 0.061 in Model **m4**, when we included random intercepts. Including random slopes reduced the residual standard error to 0.057 in Model **m5**. This is a relatively small reduction, which agrees with our earlier result that the variation in slopes among species is likely largely due to sampling variance. AIC favoured Model **m5** over Model **m4**, however (-2302.88 compared to -2205.44).



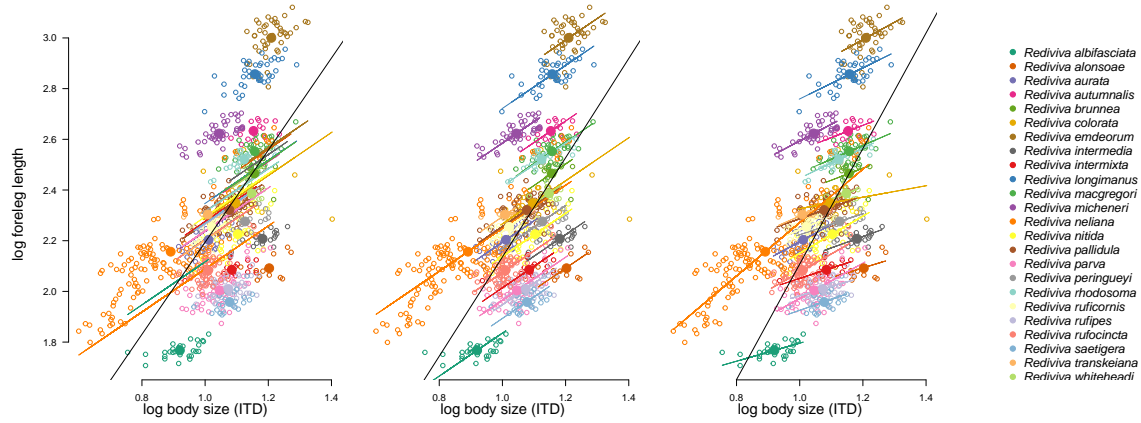


Figure S2: Allometric relationship of leg length against body size in *Rediviva* species. The black line shows the fitted evolutionary allometric relationship.

Does the evolutionary allometric slope ( $\beta_b$  in Equations 2 - 4; `mean.body` in the R output above) differ from isometry? The estimates are 1.8 (Model m3), 1.958 (Model m4) and 2.279 (Model m5). Do these estimates of  $\beta_b$  differ from 1? We can easily test this for the estimates of  $\beta_b$  from Models m3 to m5 against the null hypothesis  $H_0 : \beta_b = 1$ .

```
# m3
(tstat.m3 <- (coefficients(m3)[3]-1)/sqrt(vcov(m3)[3,3])) # t statistic

## mean.body
## 9.679821

pt(tstat.m3,df=summary(m3)$df[2], lower.tail = F)*2 # p value (2-sided alternative)

## mean.body
## 4.204085e-21

# m4
(tstat.m4 <- (summary(m4)$tTable[3,1]-1)/summary(m4)$tTable[3,2]) # t statistic

## [1] 1.541137
pt(tstat.m4,df=summary(m4)$tTable[3,3], lower.tail = F)*2 # p value (2-sided alternative)

## [1] 0.1375452

# m5
(tstat.m5 <- (summary(m5)$tTable[3,1]-1)/summary(m5)$tTable[3,2]) # t statistic

## [1] 2.149018
pt(tstat.m5,df=summary(m5)$tTable[3,3], lower.tail = F)*2 # p value (2-sided alternative)

## [1] 0.04289424
```

According to Models m3 and m5, the evolutionary allometric slope is significantly different from 1. In Model m4,  $\beta_b$  was estimated with a large standard error, leading to a relatively high P value and weak evidence against  $H_0$ . We note that Model m5 was best supported by the data and therefore conclude that –on balance– the data support the notion that the evolutionary allometric slope is larger than 1.

What proportion of the between species variation in foreleg length is explained by the variation in mean body size between species? I.e. does the evolutionary allometric relationship examined above explain a large

proportion of the variation in mean leg length? If it does, this would suggest that body size constrained the evolution of leg length. To examine this question, we re-fitted the model with random intercept (Equation 3) without entering mean body size as predictor.

$$l_{ij} = \beta_0 + \beta_w(b_{ij} - b_{.j}) + \mu_{Tj} + \epsilon_{ij} \quad (5)$$

In this model, the variance of the random effects  $\mu_{Tj} - \sigma_{Tj}^2$  – estimates the total among-species variance in foreleg length whereas the variance of the random effects  $\mu_{0j}$  (in Equation 3) –  $\sigma_{0j}^2$  – estimates the among-species variance after accounting for mean body size. Comparing  $\sigma_{Tj}^2$  to  $\sigma_{0j}^2$ , we can then estimate the proportion of among-species variance in foreleg length that is explained by species-specific body size:

$$\frac{\sigma_{Tj}^2 - \sigma_{0j}^2}{\sigma_{Tj}^2}$$

```
m4.T <- lme(log.leg ~ diff.body, random = ~1|Taxa, data=beemorph.r) # random intercept
```

The variance of the random intercepts increased from 0.06 in Model m4 to 0.083 in Model m4.T and mean body size therefore explained 28% of the variance in the foreleg length among species.

#### 4.1 Variation in allometric parameters when *R. neliana* is excluded

*R. neliana* had the steepest static allometric relationship of all species and we are going to examine this species in more detail later. Here, we quickly want to check how much the estimated variation in static allometric slopes changes if we exclude this species.

Re-running the analyses in the previous section without *R. neliana*, we obtained qualitatively similar results. The residual standard error decreases markedly, from 0.225 in Model m3 to 0.045 in Model m4, when we included random intercepts. Including random slopes reduced the residual standard error to 0.044 in Model m5 (Figure S3). The variance in slopes among species was 0.0638 which is smaller than the mean of the squared standard error of the species-specific slope estimates, 0.1129, suggesting that the variation in slopes is largely due to sampling variance.

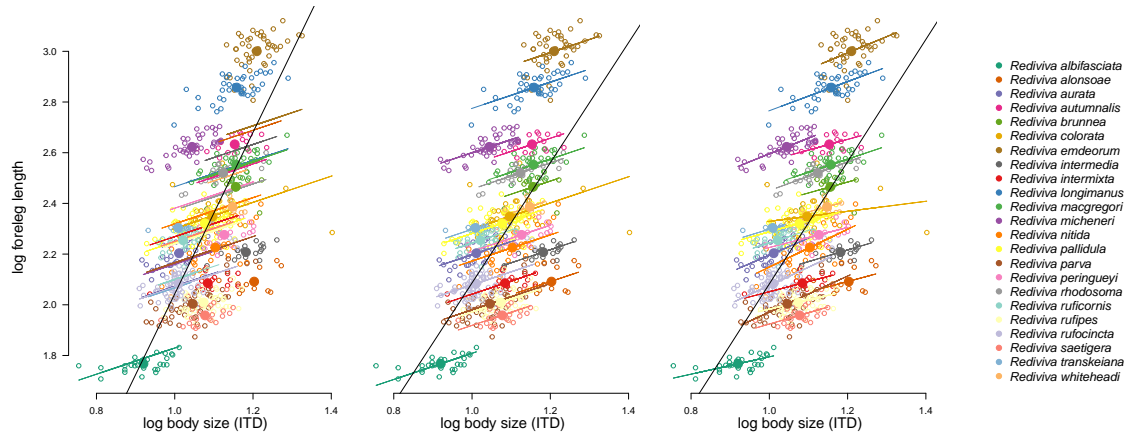


Figure S3: Allometric relationship of leg length against body size in *Rediviva* species, excluding *R. neliana*. The black line shows the fitted evolutionary allometric relationship.

## 5 Testing whether static and evolutionary allometric slopes differ

We reparameterised the multi-level model in Equations 2 to 4 to allow us to directly test whether the evolutionary allometric slope differs from the (mean) static allometric slope (Pol and Wright 2009):

$$l_{ij} = \beta_0 + \beta_w b_{ij} + (\beta_b - \beta_w) b_{.j} + \epsilon_{ij} \quad (6)$$

$$l_{ij} = \beta_0 + \beta_w b_{ij} + (\beta_b - \beta_w) b_{.j} + \mu_{0j} + \epsilon_{ij} \quad (7)$$

$$l_{ij} = (\beta_0 + \mu_{0j}) + (\beta_w + \mu_{wj}) b_{ij} + (\beta_b - \beta_w) b_{.j} + \epsilon_{ij} \quad (8)$$

The models are very similar to the ones in Equations 2 to 4 except that we include the individual body size measures instead of the difference to the species mean. This way, the coefficient for species-specific mean body size estimates the difference between the evolutionary and static allometric slopes (see van de Pol and Wright (2009) for details).

```
m6 <- lm(log.leg ~ log.body + mean.body, data=beemorph.r) # static allometry same across species
m7 <- lme(log.leg ~ log.body + mean.body, random = -1|Taxa, data=beemorph.r) # random intercept
m8 <- lme(log.leg ~ log.body + mean.body, random = -diff.body|Taxa, data=beemorph.r)
# random intercept and slope
```

Note that Models m6 to m8 are just reparameterisations of models m3 to m5 and thus identical to those in terms of their fit to the data. The estimated difference between the evolutionary and static allometric slopes ( $\delta = \beta_b - \beta_w$ ) with confidence intervals are shown in Table S2

Table S2: Difference ( $\delta$ ) between the evolutionary and static allometric slopes estimated by three different models across *Rediviva* species. lcl and ucl are the lower and upper limits of the 95% confidence interval.

Model	delta	lcl	ucl
m6	0.97	0.70	1.23
m7	1.11	-0.18	2.40
m8	1.74	0.49	2.99

The coefficient estimating the difference between the evolutionary and static allometric slopes is positive, indicating that the evolutionary allometric slope is steeper than the static allometric slope. The estimates broadly agree across models but Model m7 yielded a wide confidence interval that included 0.

## 6 Do these results change when we include *R. gigas*?

*R. gigas* was a clear outlier in the general relationship between body size and leg length across species and we therefore excluded *R. gigas* above. Here, we repeat the analysis with *R. gigas* included.

We refit Models m3 to m5 to test whether the evolutionary allometric slope differs from isometry. The estimated evolutionary slopes, i.e.  $\beta_b$ , are now 1.617 (se = 0.08,  $t = 7.666$ ,  $P = 0$  testing  $H_0 : \beta_b = 1$ ) for Model m3, 1.216 (se = 0.494,  $t = 0.436$ ,  $P = 0.667$  testing  $H_0 : \beta_b = 1$ ) for Model m4 and 1.322 (se = 0.49,  $t = 0.656$ ,  $P = 0.518$  testing  $H_0 : \beta_b = 1$ ) for Model m5. Only Model m3, which was poorly supported by the data, suggests some evidence against  $H_0$  and therefore conclude that –on balance – when including *R. gigas*, we have no clear evidence that the evolutionary allometric slope differs from 1.

We re-fitted models m6 to m8 including *R. gigas* and Table S3 below gives the estimated difference between the evolutionary and static allometric slopes (delta =  $\beta_b - \beta_w$ ) with confidence intervals.

Table S3: Difference (delta) between the evolutionary and static allometric slopes estimated by three different models across *Rediviva* species including *R. gigas*. lcl and ucl are the lower and upper limits of the 95% confidence interval.

Model	delta	lcl	ucl
m6	0.77	0.50	1.03
m7	0.37	-0.66	1.39
m8	0.78	-0.25	1.81

Including *R. gigas*, the analysis still suggested that the evolutionary allometric slope is steeper than the static allometric slope but the difference was smaller and the confidence interval overlapped zero for Models m7 and m8. In conclusion, including the outlier species *R. gigas* affects the results presented earlier, albeit not the general pattern. Figure S4 below shows, however, that the fitted evolutionary allometric line is a poor description of the general relationship between log body size and log leg length across species.

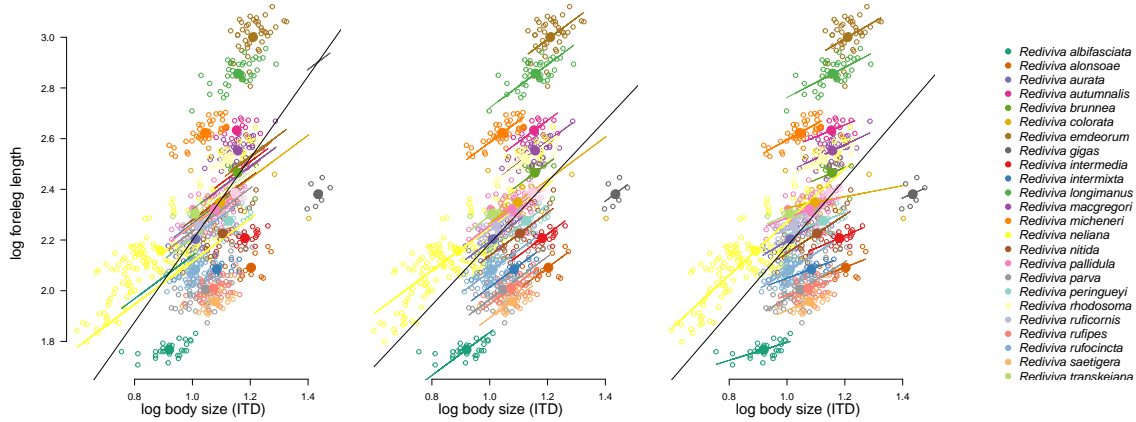


Figure S4: Allometric relationship of leg length against body size in *Rediviva* species, including *R. gigas*. The black line shows the fitted evolutionary allometric relationship.

## 7 Does accounting for phylogeny affect our results?

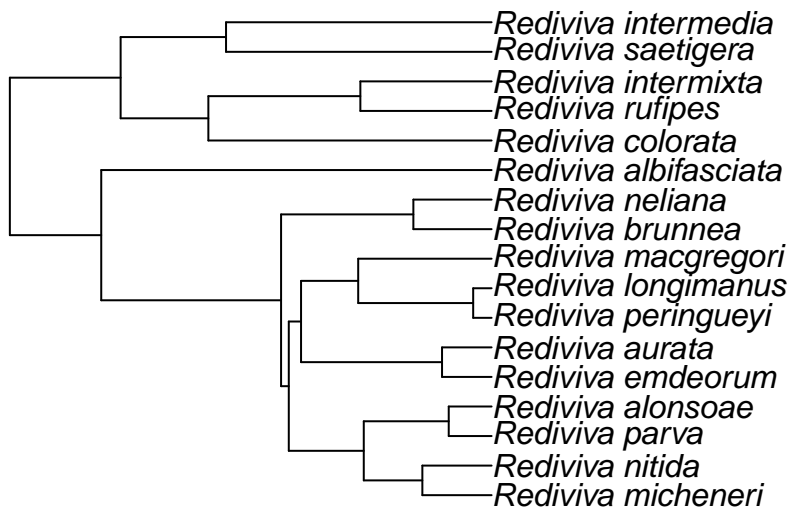
Our analyses above assumed independent residuals but clearly, the species are phylogenetically related and this relatedness could cause non-independence in the data. Our maximum likelihood estimates of the parameters of the static allometric relationships in Table S1 should not be affected because we estimated separate parameters for each species. However, our estimate of the evolutionary allometric parameters could be affected. In this section, we therefore re-run some key models while accounting for phylogenetic non-independence. We do not have complete phylogenetic information for all species examined so far, so this re-analysis is based on a reduced set of species, which is the main reason why we did not include phylogeny throughout.

A phylogeny published by Kahnt et al (2017) included most of the species in our sample. The code chunk below reads the tree and renames the tips so that the species names match those in the data frame. We drop *R. gigas* from the phylogenetic tree:

```
beetree <- read.tree("Rediviva simplified treeUSE.nex")
beetree$tip.label <- gsub("Rediviva_", "Rediviva ", beetree$tip.label)

beetree.r <- drop.tip(beetree, c("Rediviva gigas")) ## drop R. gigas

plot(beetree.r)
```



```
beetree.r$tip.label <- gsub(" ", "", beetree.r$tip.label) # remove spaces in species names
```

There are seven species in our morphological data set that don't appear in the phylogeny and we remove all observations of these species:

```
species.remove <- c("Rediviva autumnalis", "Rediviva pallidula", "Rediviva rhodosoma",
                    "Rediviva ruficornis", "Rediviva rufocincta", "Rediviva transkeiana",
                    "Rediviva whiteheadi")

beemorph.phy <- beemorph.r[!beemorph.r$Taxa %in% species.remove,]
```

Our data set contains multiple individuals per species and we therefore need to add tips to the phylogenetic tree:

```
beemorph.phy$species <- gsub(" ", "", beemorph.phy$Taxa)
rownames(beemorph.phy) <- gsub(" ", "", paste0(beemorph.phy$species, 1:NROW(beemorph.phy)))

species <- unique(beemorph.phy$species)
```

```

## because there are multiple rows per species:
nice.tree.multi <- beetree.r
for (g in 1:NROW(species)){
  if(!is.null(nice.tree.multi$edge.length)){
    flat.tree <- read.tree(text =
      paste0("(", paste(rownames(beemorph.phy)[grep(species[g], rownames(beemorph.phy))],
        collapse = ":0.02, "), ":0.02);"))
  } else {
    flat.tree <- read.tree(text =
      c("(", paste(rownames(beemorph.phy)[grep(species[g], rownames(beemorph.phy))],
        collapse = ", "), ");"))
  }

  nice.tree.multi <- bind.tree(nice.tree.multi, flat.tree,
    where = which(nice.tree.multi$tip.label == species[g]),
    position = 0, interactive = FALSE)
}

```

Next, we use phylogenetic generalized least squares models. We first add phylogenetic model structures to the errors using model m3, which is the simplest multi-level model without random effects:

```

m3.bm <- phylolm(log.leg ~ diff.body + mean.body, phy = nice.tree.multi, model = "BM",
  data = beemorph.phy)
summary(m3.bm)

```

```

##
## Call:
## phylolm(formula = log.leg ~ diff.body + mean.body, data = beemorph.phy,
##   phy = nice.tree.multi, model = "BM")
##
##      AIC   logLik
## -1506.4    757.2
##
## Raw residuals:
##      Min       1Q   Median       3Q      Max
## -0.42575 -0.07941  0.17859  0.45240  0.65894
##
## Mean tip height: 28.94673
## Parameter estimate(s) using ML:
## sigma2: 0.2111338
##
## Coefficients:
##              Estimate      StdErr t.value p.value
## (Intercept) -0.122356   3.513058 -0.0348  0.9722
## diff.body    0.885533   0.030477 29.0559 <2e-16 ***
## mean.body    2.128502   3.084062  0.6902  0.4903
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-squared: 0.5701    Adjusted R-squared: 0.5687
m3.OUr <- phylolm(log.leg ~ diff.body + mean.body, phy = nice.tree.multi, model = "OUrandomRoot",
  data = beemorph.phy)
summary(m3.OUr)

```

```
##
## Call:
## phylolm(formula = log.leg ~ diff.body + mean.body, data = beemorph.phy,
##       phy = nice.tree.multi, model = "OUrandomRoot")
##
##      AIC  logLik
## -1550.6   780.3
##
## Raw residuals:
##      Min      1Q   Median      3Q      Max
## -0.48847 -0.17793  0.09323  0.34111  0.59753
##
## Mean tip height: 28.94673
## Parameter estimate(s) using ML:
## alpha: 1.48124
## sigma2: 0.2230513
##
## Coefficients:
##              Estimate   StdErr t.value   p.value
## (Intercept) 0.174178 0.811901  0.2145  0.830202
## diff.body    0.885533 0.030867 28.6888 < 2.2e-16 ***
## mean.body    1.934129 0.742226  2.6058  0.009379 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-squared: 0.5657      Adjusted R-squared: 0.5644
##
## Note: p-values and R-squared are conditional on alpha=1.48124.
m3.OUf <- phylolm(log.leg ~ diff.body + mean.body, phy = nice.tree.multi, model = "OUfixedRoot",
                 data = beemorph.phy)
summary(m3.OUf)

##
## Call:
## phylolm(formula = log.leg ~ diff.body + mean.body, data = beemorph.phy,
##       phy = nice.tree.multi, model = "OUfixedRoot")
##
##      AIC  logLik
## -1550.6   780.3
##
## Raw residuals:
##      Min      1Q   Median      3Q      Max
## -0.48847 -0.17793  0.09323  0.34111  0.59753
##
## Mean tip height: 28.94673
## Parameter estimate(s) using ML:
## alpha: 1.48124
## sigma2: 0.2230513
##
## Coefficients:
##              Estimate   StdErr t.value   p.value
## (Intercept) 0.174178 0.811902  0.2145  0.830202
## diff.body    0.885533 0.030867 28.6888 < 2.2e-16 ***
## mean.body    1.934129 0.742226  2.6058  0.009379 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-squared: 0.5657      Adjusted R-squared: 0.5644
##
## Note: p-values and R-squared are conditional on alpha=1.48124.
m3.lambda <- phylolm(log.leg ~ diff.body + mean.body, phy = nice.tree.multi, model = "lambda",
                    data = beemorph.phy)
summary(m3.lambda)

##
## Call:
## phylolm(formula = log.leg ~ diff.body + mean.body, data = beemorph.phy,
##         phy = nice.tree.multi, model = "lambda")
##
##      AIC  logLik
## -1535.6   772.8
##
## Raw residuals:
##      Min      1Q   Median      3Q      Max
## -0.42575 -0.07952  0.17842  0.45212  0.65895
##
## Mean tip height: 28.94673
## Parameter estimate(s) using ML:
## lambda : 0.9894676
## sigma2: 0.0133414
##
## Coefficients:
##              Estimate      StdErr t.value    p.value
## (Intercept) -0.121285   0.879727 -0.1379   0.890389
## diff.body    0.885533   0.030867 28.6885 < 2.2e-16 ***
## mean.body    2.127609   0.772462  2.7543  0.006049 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-squared: 0.566 Adjusted R-squared: 0.5646
##
## Note: p-values and R-squared are conditional on lambda=0.9894676.
```

AIC suggests that the best model structure is the one assuming an Ornstein-Uhlenbeck process (Models m3.OUr and m3.OUf). However, all models return a similar estimate for the evolutionary allometric slope ( $\beta_b$  in Equation 2 and mean.body in the output above ranged from 1.934 to 2.129), which is also comparable to the slope that was estimated by the model without phylogenetic structure (m3: 1.817) in an earlier section.

We saw earlier, however, that Model m3 was not a good description of our data. We next add a phylogenetic correlation structure to Models m4 and m5. These models include random effects and are more complex statistical models. We only managed to get the models assuming a Brownian Motion process to converge.

```
m4.bm <- lme(log.leg ~ diff.body + mean.body, random = ~1|Taxa, data=beemorph.phy,
            correlation = corBrownian(1, nice.tree.multi))
summary(m4.bm)
```

```
## Linear mixed-effects model fit by REML
## Data: beemorph.phy
##      AIC      BIC   logLik
```



```

##      -5817.773 -5795.489 2913.886
##
## Random effects:
## Formula: ~1 | Taxa
##      (Intercept)   Residual
## StdDev:    0.2837536 0.07824462
##
## Correlation Structure: corBrownian
## Formula: ~1 | Taxa
## Parameter estimate(s):
## numeric(0)
## Fixed effects: log.leg ~ diff.body + mean.body
##      Value Std.Error DF   t-value p-value
## (Intercept) 0.1728923 0.8626126 622  0.200429  0.8412
## diff.body    0.6746445 0.0287997 622 23.425432  0.0000
## mean.body    1.9366506 0.7884406  15  2.456305  0.0267
## Correlation:
##      (Intr) dff.bd
## diff.body  0.000
## mean.body -0.997  0.000
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -3.44538322 -0.43478179 -0.04354433  0.44087324  3.42865107
##
## Number of Observations: 640
## Number of Groups: 17
m5.bm <- lme(log.leg ~ diff.body + mean.body, random =~diff.body|Taxa, data=beemorph.phy,
             correlation = corBrownian(1, nice.tree.multi))
summary(m5.bm)

## Linear mixed-effects model fit by REML
## Data: beemorph.phy
##      AIC      BIC    logLik
## -5960.922 -5929.725 2987.461
##
## Random effects:
## Formula: ~diff.body | Taxa
## Structure: General positive-definite, Log-Cholesky parametrization
##      StdDev      Corr
## (Intercept) 0.28478004 (Intr)
## diff.body    0.25217861 0.531
## Residual    0.06872903
##
## Correlation Structure: corBrownian
## Formula: ~1 | Taxa
## Parameter estimate(s):
## numeric(0)
## Fixed effects: log.leg ~ diff.body + mean.body
##      Value Std.Error DF   t-value p-value
## (Intercept) -0.3468319 0.7848926 622 -0.441885  0.6587
## diff.body    0.5376452 0.0804316 622  6.684499  0.0000
## mean.body    2.4142344 0.7169023  15  3.367592  0.0042
## Correlation:

```

```
##          (Intr) dff.bd
## diff.body 0.121
## mean.body -0.996 -0.086
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.90794724 -0.48515084 -0.04479196  0.47746928  2.97036221
##
## Number of Observations: 640
## Number of Groups: 17
```

Again, the estimated evolutionary allometric slopes were similar to the ones obtained earlier from the models assuming no phylogenetic dependence structure: 1.937 in Model `m4.bm` compared to 1.958 in Model `m4` and 2.414 in Model `m5.bm` compared to 2.279 in Model `m5`. We conclude that our earlier results were not sensitive to violation of the assumption that the errors were phylogenetically independent.

## 8 Estimating the parameters of the static allometric relationships for each population in *R. neliana*

We saw earlier that *R. neliana* had a steeper allometric slope than the other species. Is it possible that our *R. neliana* sample is made up of individuals that belong to several isolated populations on independent evolutionary tracks? In this section, we examine variation within and between populations of *R. neliana*. For this section, we refer to the within-population allometric relationship as the static allometric relationship and to the between-population allometric relationship as the evolutionary allometric relationship.

We first select the *R. neliana* individuals from the full data set and tabulate them by population:

```
beemorph.n <- beemorph[beemorph$Taxa == "Rediviva neliana", c(2,3,5:8)]
beemorph.n$population_code <- factor(beemorph.n$population_code)
beemorph.n$pop.no <- as.numeric(beemorph.n$population_code)
table(beemorph.n$population_code)
```

```
##
##  R. neliana_1 R. neliana_11 R. neliana_18  R. neliana_2 R. neliana_21
##           20           15           15           15           15
## R. neliana_22 R. neliana_3  R. neliana_4  R. neliana_6  R. neliana_8
##           7           15           15           15           15
```

Following identical procedures as above, we first estimate the static allometric relationship for each population separately:

```
m1.n <- lm(log.leg ~ population_code * log.body, data=beemorph.n)
anova(m1.n)
```

```
## Analysis of Variance Table
##
## Response: log.leg
##
##              Df Sum Sq Mean Sq  F value    Pr(>F)
## population_code      9 3.9149  0.43499 471.6891 < 2.2e-16 ***
## log.body              1 0.8239  0.82385 893.3554 < 2.2e-16 ***
## population_code:log.body  9 0.0239  0.00266   2.8845  0.003913 **
## Residuals          127 0.1171  0.00092
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA table above shows that there is variation among populations in both the allometric intercept

and the allometric slope even though the Mean Squares for the former was much larger than for the latter. The model explained 97.6% of the variation in leg length. Table S4 shows the maximum likelihood estimates for the parameters  $a$  and  $b$  (Equation 1) of the population-specific allometric relationships.

The variance in slopes among populations was 0.0478 which is not much larger than the mean of the squared standard error of the population-specific slope estimates 0.0234 suggesting that the variation in slopes is largely due to sampling variance.

Table S4: Summary of morphological data and allometric parameters for 10 populations of *R. neliana* bees. N: sample size (number of individuals per population); LegLength: mean total length of foreleg length [mm]; LL.SD: standard deviation of leg length; BodySize: inter-tegula distance as a measure of body size; BS.SD: standard deviation of body size; a and b are the parameters of the allometric relationships; lcl.a, ucl.a: lower and upper 95% confidence limits of parameter a; lcl.b, ucl.b: lower and upper 95% confidence limits of parameter b.

	N	LegLength	LL.SD	BodySize	BS.SD	a	b	lcl.a	ucl.a	lcl.b	ucl.b
R. neliana_1	20	8.54	0.42	2.69	0.15	4.74	0.60	3.72	6.04	0.35	0.84
R. neliana_11	15	8.47	0.66	2.21	0.19	4.51	0.79	3.86	5.27	0.60	0.99
R. neliana_18	15	9.17	0.97	2.52	0.28	3.98	0.90	3.46	4.58	0.75	1.05
R. neliana_2	15	7.48	1.09	2.35	0.34	3.25	0.98	2.96	3.56	0.87	1.09
R. neliana_21	15	13.00	0.50	3.20	0.14	6.94	0.54	4.61	10.46	0.19	0.89
R. neliana_22	7	10.15	0.29	2.48	0.11	6.59	0.48	3.96	10.97	-0.08	1.04
R. neliana_3	15	7.56	0.84	2.34	0.34	4.02	0.74	3.67	4.40	0.64	0.85
R. neliana_4	15	7.27	0.30	2.12	0.10	4.26	0.71	3.30	5.50	0.37	1.05
R. neliana_6	15	8.10	0.22	2.35	0.08	6.37	0.28	4.36	9.30	-0.16	0.73
R. neliana_8	15	9.01	0.99	2.31	0.27	4.23	0.90	3.75	4.78	0.76	1.05

We fit the models described in equations 2 to 4 to the *R. neliana* data:

```
m3.n <- lm(log.leg ~ diff.body + mean.body, data=beemorph.n) # static allometry same across species
summary(m3.n)

##
## Call:
## lm(formula = log.leg ~ diff.body + mean.body, data = beemorph.n)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.199955 -0.076061  0.003661  0.082739  0.166169
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.07708    0.06135  17.555  <2e-16 ***
## diff.body    0.82854    0.08584   9.652  <2e-16 ***
## mean.body    1.21159    0.06831  17.738  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09404 on 144 degrees of freedom
## Multiple R-squared:  0.739, Adjusted R-squared:  0.7354
## F-statistic: 203.9 on 2 and 144 DF, p-value: < 2.2e-16

m4.n <- lme(log.leg ~ diff.body + mean.body, random=~1|population_code, data=beemorph.n)
# random intercept
summary(m4.n)
```

```

## Linear mixed-effects model fit by REML
## Data: beemorph.n
##      AIC      BIC    logLik
##   -525.8002 -510.9511 267.9001
##
## Random effects:
## Formula: ~1 | population_code
##      (Intercept)   Residual
## StdDev:  0.09881118 0.03220614
##
## Fixed effects: log.leg ~ diff.body + mean.body
##              Value Std.Error DF   t-value p-value
## (Intercept) 1.0483477 0.25280300 136   4.146896  0.0001
## diff.body    0.8285352 0.02939846 136  28.182945  0.0000
## mean.body    1.2570502 0.28226277   8   4.453475  0.0021
## Correlation:
##      (Intr) dff.bd
## diff.body  0.000
## mean.body -0.992  0.000
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.22130307 -0.67146969 -0.06919938  0.67961895  2.49822234
##
## Number of Observations: 147
## Number of Groups: 10
m5.n <- lme(log.leg ~ diff.body + mean.body, random =~diff.body|population_code, data=beemorph.n)
# random intercept and slope
summary(m5.n)

## Linear mixed-effects model fit by REML
## Data: beemorph.n
##      AIC      BIC    logLik
##   -528.1827 -507.394 271.0913
##
## Random effects:
## Formula: ~diff.body | population_code
## Structure: General positive-definite, Log-Cholesky parametrization
##              StdDev      Corr
## (Intercept) 0.09885416 (Intr)
## diff.body    0.13291295 0.011
## Residual    0.03049193
##
## Fixed effects: log.leg ~ diff.body + mean.body
##              Value Std.Error DF   t-value p-value
## (Intercept) 1.0462855 0.25281557 136   4.138533  0.0001
## diff.body    0.7690741 0.05632445 136  13.654356  0.0000
## mean.body    1.2593852 0.28227695   8   4.461524  0.0021
## Correlation:
##      (Intr) dff.bd
## diff.body  0.002
## mean.body -0.992 -0.001
##
## Standardized Within-Group Residuals:

```

```
##           Min           Q1           Med           Q3           Max
## -2.0080129 -0.6828817  0.1063575  0.6916677  2.2161583
##
## Number of Observations: 147
## Number of Groups: 10
```

The residual standard error decreased markedly, from 0.094 in Model `m3.n` to 0.032 in Model `m4.n`, when we included random intercepts. Including random slopes reduced the residual standard error to 0.03 in Model `m5.n`. This is a relatively small reduction, which agrees with our earlier result that the variation in slopes among species is likely largely due to sampling variance. AIC favoured Model `m5` over Model `m4`, however (-528.18 compared to -525.8).

Using the fitted Models `m3.n` to `m5.n`, we test whether the evolutionary allometric slope differs from isometry. The estimated evolutionary slopes, i.e.  $\beta_b$ , are 1.212 (se = 0.068,  $t = 3.098$ ,  $P = 0.002$  testing  $H_0 : \beta_b = 1$ ) for Model `m3.n`, 1.257 (se = 0.282,  $t = 0.911$ ,  $P = 0.389$  testing  $H_0 : \beta_b = 1$ ) for Model `m4.n` and 1.259 (se = 0.282,  $t = 0.919$ ,  $P = 0.385$  testing  $H_0 : \beta_b = 1$ ) for Model `m5.n`. Only Model `m5.n`, which was poorly supported by the data, suggests some evidence against  $H_0$  and we therefore conclude that we have no strong evidence that the evolutionary allometric slope among populations of *R. neliana* differs from 1.

We next examined  $R^2$  for the evolutionary allometric relationship as we did in the between-species comparison earlier.

```
m4.T.n <- lme(log.leg ~ diff.body, random = ~1|population_code, data=beemorph.n) # random intercept
```

The variance of the random intercepts increased from 0.01 in Model `m4.n` to 0.03 in Model `m4.T.n` and mean body size therefore explained 67.8% of the variance in the foreleg length among species.

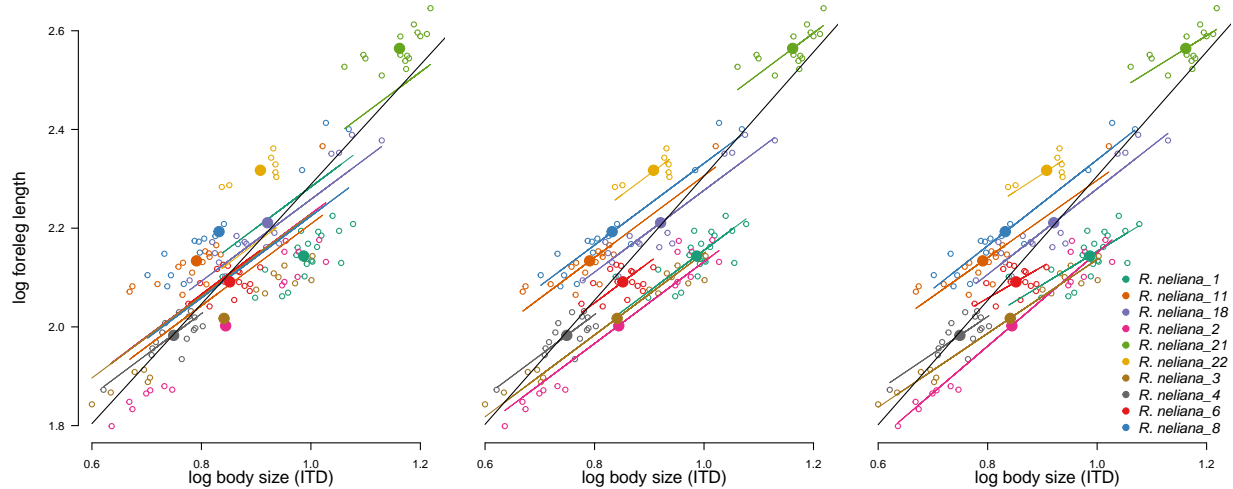


Figure S5: Allometric relationship of leg length against body size in *R. neliana* populations. The black line shows the fitted evolutionary allometric relationship.

As we saw when comparing evolutionary and static allometry across species, the allometric slope is steeper among populations than within. Is this difference statistically supported?

The coefficient estimating the difference between the evolutionary and static allometric slopes was positive (Table S5), indicating that the evolutionary allometric slope is steeper than the static allometric slope but the difference is much smaller than in the comparison across species and Models `m7` and `m8` both yielded a confidence interval that included 0.

In summary, when examining within- and among-population allometric relationships in *R. neliana*, we found qualitatively similar patterns as we did when examining within- and among-species allometric relationship

across the *Rediviva* genus but these patterns were weaker and not always statistically supported.

Table S5: Difference (delta) between the evolutionary and static allometric slopes estimated by three different models across *R. neliana* populations. lcl and ucl are the lower and upper limits of the 95% confidence interval.

Model	delta	lcl	ucl
m6	0.38	0.17	0.60
m7	0.43	-0.23	1.08
m8	0.49	-0.17	1.15

## 9 Is variation in allometric relationships explained by spur length of the host flowers?

For all species except *R. neliana*, the allometric exponent (slope) was estimated to be smaller than one. And when we considered within-population allometry for *R. neliana*, this was also true for all populations of that species. An allometric exponent  $< 1$  implies that small individuals have relatively long legs for their size and large individuals have relatively short legs for their size. What are the selective forces acting on this relationship and what could explain the variation among species – and in the case of *R. neliana*, variation among population – in these allometric parameters?

These bees collect oil from flowers using their forelegs, so spur length of the flowers from which these bees collect the oil is a candidate environmental driver likely to act as selective force. We generated a list of oil host plants for each *Rediviva* species, combined this with spur length data and calculated a mean spur length across all host plants per bee species.

Figure S6 shows leg length as a function of spur length (using a weighted mean across all plant species each bee species is known to visit). *R. gigas* is included in this plot. Leg length tracks spur length with most species having legs that are longer than the spur length of the flowers they visit.

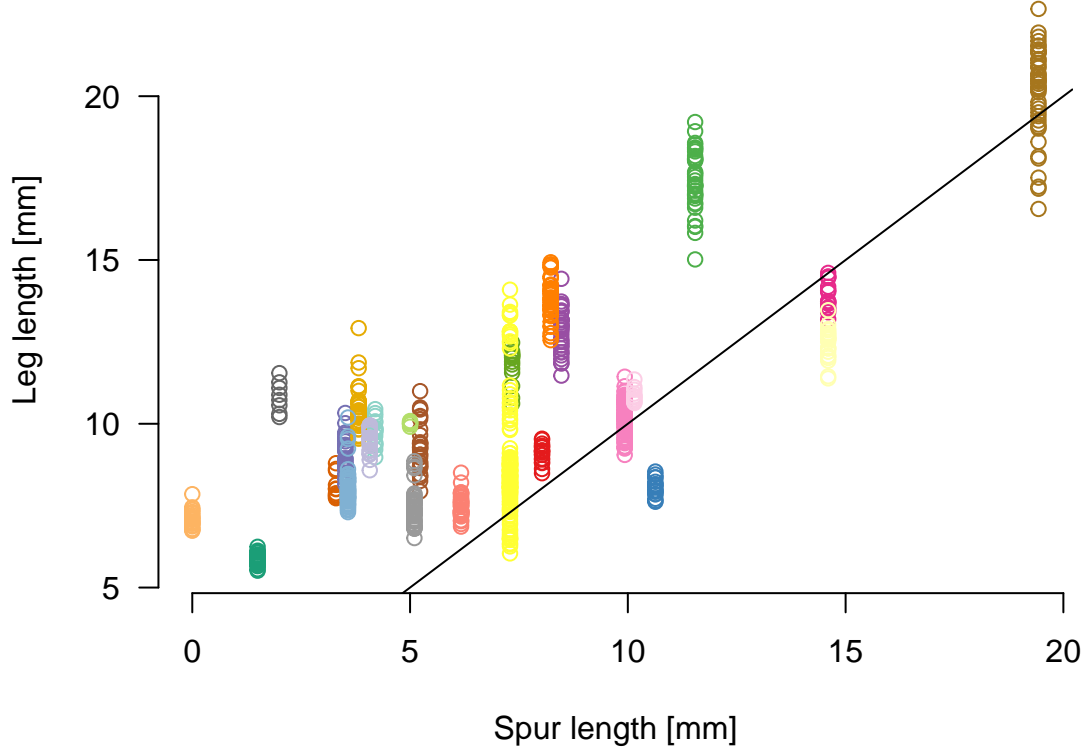


Figure S6: Leg length of *Rediviva* individuals of different species against mean spur length of their host plant flowers. The solid black line is the 1:1 relationship. Colours indicate different species using the same colour scheme as in Figure S1.)

We slightly re-parameterise the model described in Equation 3 by subtracting the overall mean log body size ( $\beta_{..}$ ) from the species-specific mean log body size ( $\beta_{.j}$ ) so that the intercept estimates the log leg length for a bee of average log body size.

$$l_{ij} = (\beta_0 + \mu_{0j}) + \beta_w(b_{ij} - b_{.j}) + \beta_b(b_{.j} - b_{..}) + \epsilon_{ij} \quad (9)$$

We then add log spur length,  $s$ , as an explanatory variable to this model:

$$l_{ij} = (\beta_0 + \mu_{rj}) + \beta_w(b_{ij} - b_{.j}) + \beta_b(b_{.j} - b_{..}) + \beta_s s + \epsilon_{ij} \quad (10)$$

Spur length varies among species and the random effects  $\mu_{rj}$  in Equation 10 therefore estimate the remaining variation among species in the allometric intercept after accounting for variation in spur length.

We estimate the proportion of variance of the allometric intercept that is explained by spur length as:

$$\frac{\sigma_{0j}^2 - \sigma_{rj}^2}{\sigma_{rj}^2}$$

where  $\sigma_{0j}^2$  is the estimated variance of the random effects  $\mu_{0j}$  and  $\sigma_{rj}^2$  is the estimated variance of the random effects  $\mu_{rj}$ . We excluded *R. gigas* from this analysis because this species was an outlier with a large influence on the estimated evolutionary allometric relationship.

```
m9 <- lme(log.leg ~ diff.body + I(mean.body-mean(log.body)), random=~1|Taxa, data=beemorph.r) # random
summary(m9)
```

```
## Linear mixed-effects model fit by REML
## Data: beemorph.r
##      AIC      BIC    logLik
## -2205.44 -2181.667 1107.72
##
## Random effects:
## Formula: ~1 | Taxa
##      (Intercept)   Residual
## StdDev:    0.2451031 0.06115609
##
## Fixed effects: log.leg ~ diff.body + I(mean.body - mean(log.body))
##                                     Value Std.Error DF t-value p-value
## (Intercept)                        2.2388955 0.0543921 836 41.16216 0.0000
## diff.body                          0.8489607 0.0271087 836 31.31688 0.0000
## I(mean.body - mean(log.body)) 1.9577526 0.6214584 22 3.15026 0.0046
## Correlation:
##                                     (Intr) dff.bd
## diff.body                          0.000
## I(mean.body - mean(log.body)) -0.389 0.000
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -5.281023741 -0.546657864 -0.007954349 0.532025096 3.694688325
##
## Number of Observations: 861
## Number of Groups: 24
```

```
m10 <- lme(log.leg ~ diff.body + I(mean.body-mean(log.body)) + log(spur_length+0.5),
            random=~1|Taxa, data=beemorph.r) # adding spur length
summary(m10)
```

```
## Linear mixed-effects model fit by REML
## Data: beemorph.r
##      AIC      BIC    logLik
## -2210.385 -2181.864 1111.192
##
## Random effects:
## Formula: ~1 | Taxa
##      (Intercept)   Residual
## StdDev:    0.1960997 0.06115595
##
## Fixed effects: log.leg ~ diff.body + I(mean.body - mean(log.body)) + log(spur_length + 0.5)
##                                     Value Std.Error DF t-value p-value
## (Intercept)                        1.8703386 0.1099693 836 17.007819 0.0000
## diff.body                          0.8489607 0.0271087 836 31.316955 0.0000
## I(mean.body - mean(log.body)) 1.1922890 0.5400352 21 2.207799 0.0385
## log(spur_length + 0.5)          0.2134248 0.0584735 21 3.649937 0.0015
## Correlation:
```



```
##                               (Intr) dff.bd I(-m(
## diff.body                    0.000
## I(mean.body - mean(log.body)) 0.215  0.000
## log(spur_length + 0.5)       -0.918  0.000 -0.389
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -5.27703519 -0.54397154 -0.01176452  0.53283239  3.69351607
##
## Number of Observations: 861
## Number of Groups: 24
```

There was strong evidence that leg length was positively related to spur length Figure S7 and variance of the random intercepts decreased from 0.06 in Model m9 to 0.038 in Model m10. Spur length therefore explained 36% of the variance in the intercept among species.

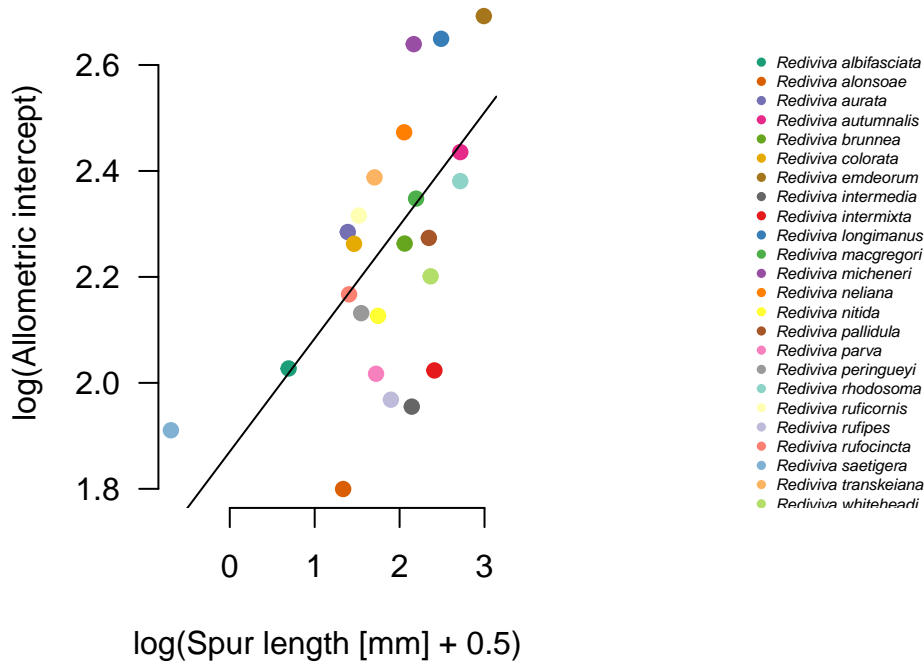


Figure S7: Relationship of allometric intercept to spur length of the host plants for *Rediviva* bee species. The allometric intercept shown are the Best Linear Unbiased Predictors for the species-specific intercepts estimated by model m9. The regression line shows the fitted relationship from model m10.

So far, we analysed narrow-sense allometry as a power law (Equation 1) and we used log-transformed values for body size, leg length and spur length. An argument could be made that selection on leg length through variation in spur length follows arithmetic relationships instead: 1 mm change in spur length should trigger a 1 mm response in leg length. We therefore re-fitted Model m10 on the arithmetic scale:

$$l_{ij}^a = (\beta_0 + \mu_{rj}) + \beta_w(b_{ij}^a - b_{..}^a) + \beta_b(b_{.j}^a - b_{..}^a) + \beta_s s^a + \epsilon_{ij} \quad (11)$$

The superscripts <sup>a</sup> indicate that these variables were on the arithmetic scale, rather than the logarithmic scale used so far.

```
# calculate mean body size per species on arithmetic scale:
mean.body.a <- aggregate(beemorph.r$body_size, by = list(beemorph.r$Taxa), FUN=mean)
colnames(mean.body.a) <- c("Taxa", "mean.body.a")

# add mean body size back onto data frame:
beemorph.r <- merge(beemorph.r, mean.body.a, by.x = "Taxa", by.y = "Taxa", all.x=T)

# difference in body size within species:
beemorph.r$diff.body.a <- beemorph.r$body_size - beemorph.r$mean.body.a

m11 <- lme(log_length ~ diff.body.a + I(mean.body.a - mean(body_size)) + spur_length,
           random =~1|Taxa, data=beemorph.r) # adding spur length
summary(m11)

## Linear mixed-effects model fit by REML
## Data: beemorph.r
##      AIC      BIC    logLik
## 1880.656 1909.176 -934.3279
##
## Random effects:
## Formula: ~1 | Taxa
##      (Intercept) Residual
## StdDev:      2.025004 0.6631789
##
## Fixed effects: log_length ~ diff.body.a + I(mean.body.a - mean(body_size)) + spur_length
##
##              Value Std.Error DF   t-value p-value
## (Intercept)      6.508818 0.8215969 836   7.922155  0.0000
## diff.body.a       3.015978 0.1081592 836  27.884618  0.0000
## I(mean.body.a - mean(body_size)) 2.724555 2.0723341 21   1.314728  0.2028
## spur_length       0.508026 0.1074654 21   4.727347  0.0001
## Correlation:
##
##              (Intr) dff.b. I(..-m
## diff.body.a      0.000
## I(mean.body.a - mean(body_size)) 0.252  0.000
## spur_length     -0.840  0.000 -0.507
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -5.824013443 -0.557119795 -0.003935144  0.503045983  3.814275300
##
## Number of Observations: 861
## Number of Groups: 24

# testing whether the effect of spur length is different from 1
tstat.m11 <- (summary(m11)$tTable[4,1]-1)/summary(m11)$tTable[4,2] # t statistic
p.spur <- pt(tstat.m11,df=summary(m11)$tTable[4,3], lower.tail = T)*2 # p value (2-sided alternative)
```

In this analysis, the coefficient for spur length is easier to interpret: it estimates the change in leg length in millimeters, for every millimeter change in spur length. If leg length tracked spur length perfectly, we would expect this coefficient to be close to 1. However, the estimated effect was 0.51, which is considerably less than 1 ( $t = -4.58$ ,  $P = 1.6 \times 10^{-4}$ , Figure S8).

Comparing the variance of the random intercepts in Model m11 to an equivalent model without spur length,

we find that spur length explained 49.4% of the variance in the intercept among species when analysing allometry on the arithmetic scale.

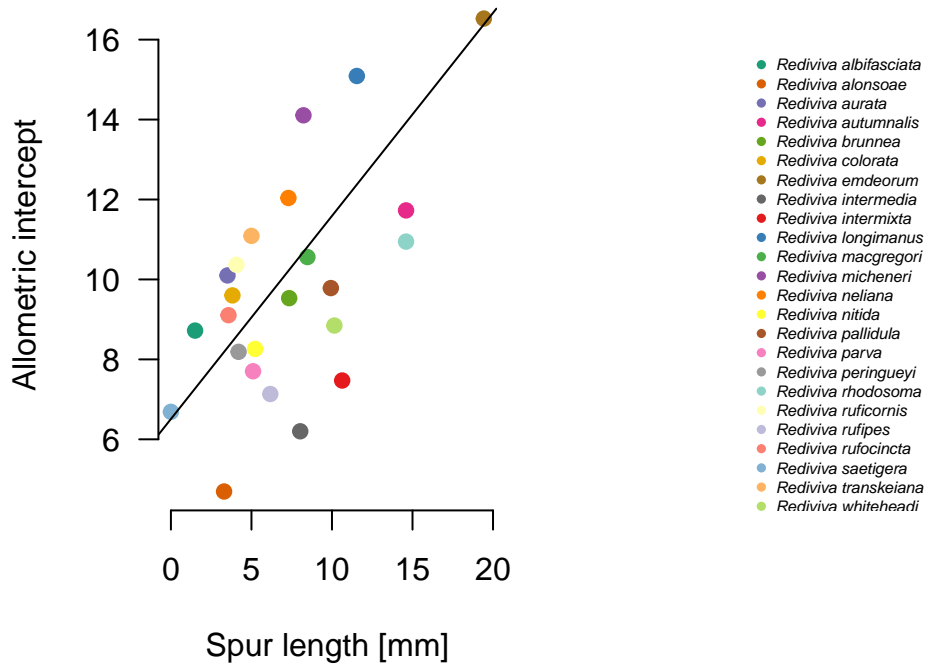


Figure S8: Relationship of allometric intercept (i.e. estimated species-specific leg length from Model m11) to spur length of the host plants for *Rediviva* bee species, fitted on the arithmetic scale.

## 10 Is allometric variation among *R. neliana* populations explained by spur length of the host flowers?

We next apply the analyses of the section above to *R. neliana* and examine whether variation in the allometric parameters among populations is related to spur length of the population-specific host plants.

Figure S9 shows leg length as a function of spur length. Leg length tracks spur length reasonably well across populations.

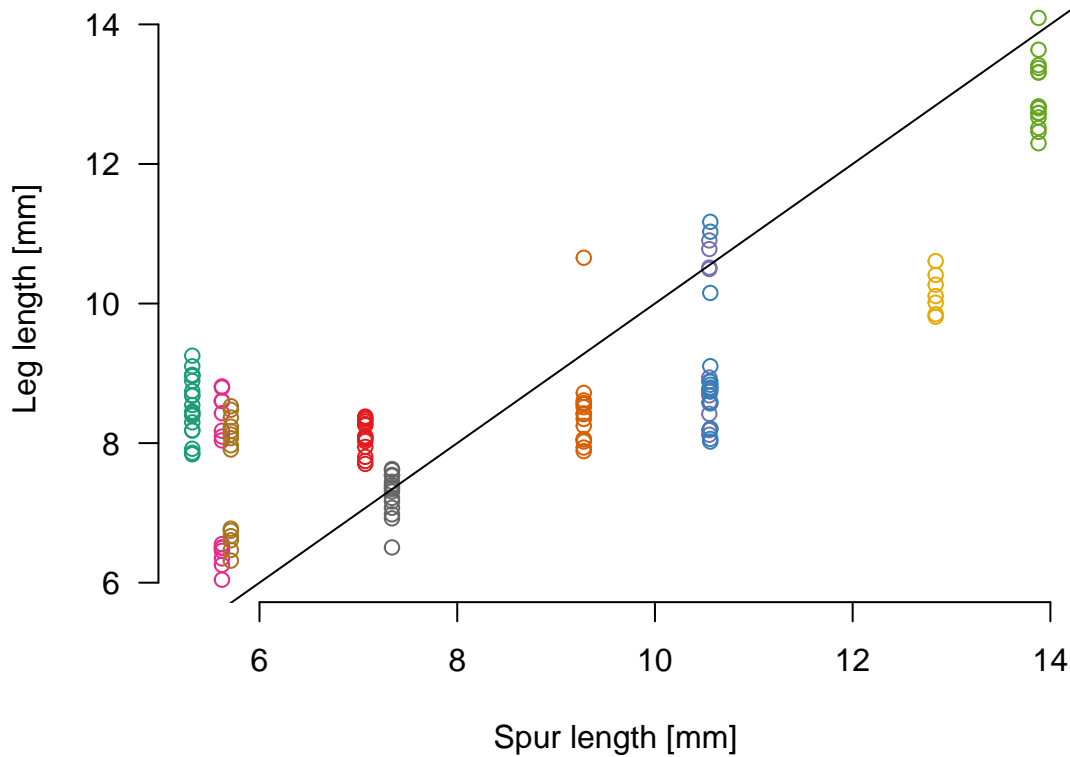


Figure S9: Leg length of *R. neiliana* individuals from different populations against spur length of their host plant flowers. The solid black line is the 1:1 relationship. Colours indicate different populations using the same colour scheme as in Figure S5.)

```
m9.n <- lme(log.leg ~ diff.body + I(mean.body - mean(log.body)), random = ~1 | population_code,
            data=beemorph.n) # random intercepts and slopes
summary(m9.n)
```

```
## Linear mixed-effects model fit by REML
## Data: beemorph.n
##      AIC      BIC    logLik
## -525.8002 -510.9511 267.9001
##
## Random effects:
## Formula: ~1 | population_code
##      (Intercept)  Residual
## StdDev:  0.09881118 0.03220614
##
## Fixed effects: log.leg ~ diff.body + I(mean.body - mean(log.body))
##
##              Value Std.Error DF t-value p-value
## (Intercept)    2.1683868 0.03137374 136 69.11472  0.0000
## diff.body      0.8285352 0.02939846 136 28.18295  0.0000
## I(mean.body - mean(log.body)) 1.2570502 0.28226277 8 4.45348 0.0021
## Correlation:
```

```

##                               (Intr) dff.bd
## diff.body                    0.000
## I(mean.body - mean(log.body)) 0.021 0.000
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.22130307 -0.67146969 -0.06919938 0.67961895 2.49822234
##
## Number of Observations: 147
## Number of Groups: 10

m10.n <- lme(log.leg ~ diff.body + I(mean.body-mean(log.body)) + log(population_spur_length+0.5),
             random = ~1|population_code, data=beemorph.n) # adding spur length
summary(m10.n)

## Linear mixed-effects model fit by REML
## Data: beemorph.n
##      AIC      BIC    logLik
##   -541.5314 -523.7543 276.7657
##
## Random effects:
## Formula: ~1 | population_code
##      (Intercept)   Residual
## StdDev: 0.02348257 0.03219929
##
## Fixed effects: log.leg ~ diff.body + I(mean.body - mean(log.body)) + log(population_spur_length +
##                               Value Std.Error DF t-value p-value
## (Intercept)                1.5161072 0.06041296 136 25.09573      0
## diff.body                   0.8285352 0.02939221 136 28.18894      0
## I(mean.body - mean(log.body)) 0.9190710 0.07718875   7 11.90680      0
## log(population_spur_length + 0.5) 0.2985221 0.02748450   7 10.86147      0
## Correlation:
##                               (Intr) dff.bd I(.-m(
## diff.body                    0.000
## I(mean.body - mean(log.body)) 0.396 0.000
## log(population_spur_length + 0.5) -0.991 0.000 -0.397
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.16171198 -0.63306065 -0.04199296 0.66023437 2.54849768
##
## Number of Observations: 147
## Number of Groups: 10

```

The variance of the random intercepts decreased from 0.01 in Model m9 to 0.001 in Model m10 and spur length therefore explained 94.4% of the variance in the intercept among species. The relationship between intercepts and spur length was strongly positive (Figure S10).

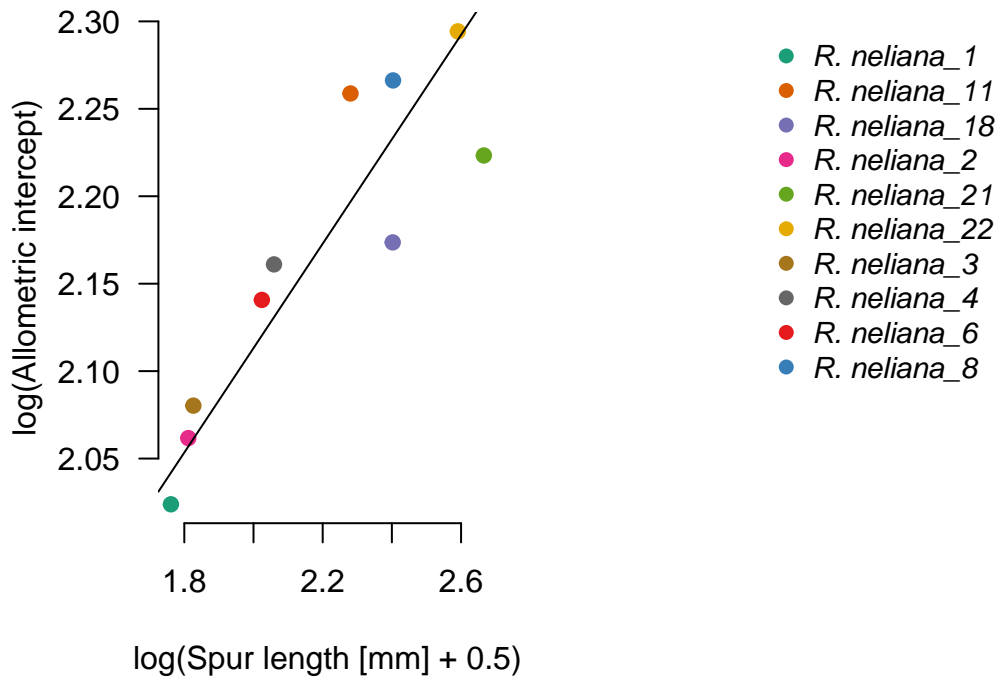


Figure S10: Relationship of allometric intercept to spur length of the host plants for *Rediviva neliana* populations. The allometric intercept shown are the Best Linear Unbiased Predictors for the population-specific intercepts estimated by model m9. The regression line shows the fitted relationship from model m10.

```
# calculate mean body size per population on arithmetic scale:
mean.body.a <- aggregate(exp(beemorph.n$log.body), by = list(beemorph.n$population_code), FUN=mean)
colnames(mean.body.a) <- c("Taxa", "mean.body.a")

# add mean body size back onto data frame:
beemorph.n <- merge(beemorph.n, mean.body.a, by.x = "population_code", by.y = "Taxa", all.x=T)
# difference in body size within species:
beemorph.n$diff.body.a <- exp(beemorph.n$log.body) - beemorph.n$mean.body.a

m11.n <- lme(exp(log.leg) ~ diff.body.a + I(mean.body.a - mean(body_size)) + population_spur_length,
             random = ~1 | population_code, data=beemorph.n) # adding spur length
summary(m11.n)

## Linear mixed-effects model fit by REML
## Data: beemorph.n
##      AIC      BIC    logLik
##  96.24288 114.0199 -42.12144
##
## Random effects:
## Formula: ~1 | population_code
##      (Intercept) Residual
## StdDev:   0.2013907 0.2918618
```

```
##
## Fixed effects: exp(log.leg) ~ diff.body.a + I(mean.body.a - mean(body_size)) + population_spur_
##
##               Value Std.Error DF t-value p-value
## (Intercept)    6.179117 0.24712044 136 25.00448      0
## diff.body.a    2.889377 0.11122753 136 25.97717      0
## I(mean.body.a - mean(body_size)) 3.271283 0.26950322 7 12.13820      0
## population_spur_length    0.308210 0.02703073 7 11.40223      0
## Correlation:
##               (Intr) dff.b. I(..-m
## diff.body.a      0.000
## I(mean.body.a - mean(body_size)) 0.480 0.000
## population_spur_length    -0.961 0.000 -0.495
##
## Standardized Within-Group Residuals:
##               Min      Q1      Med      Q3      Max
## -2.28662680 -0.57060922 -0.03948262 0.56448828 2.66546075
##
## Number of Observations: 147
## Number of Groups: 10
# testing whether the effect of spur length is different from 1
tstat.m11 <-(summary(m11.n)$tTable[4,1]-1)/summary(m11.n)$tTable[4,2] # t statistic
p.spur <- pt(tstat.m11,df=summary(m11.n)$tTable[4,3], lower.tail = T)*2 # p value (2-sided alternative)
```

As for the between-species comparison, we re-fitted the multi-level model (Equation 11) to leg length and body length on the arithmetic scale. If leg length tracked spur length perfectly, we would expect this coefficient to be close to 1 on this scale. However, the estimated effect was 0.31, which is considerably less than 1 ( $t = -25.59$ ,  $P = 0$ , Figure S11).

Comparing the variance of the random intercepts in Model m11 to an equivalent model without spur length, we find that spur length explained 94.9% of the variance in the intercept among populations when analysing allometry on the arithmetic scale.

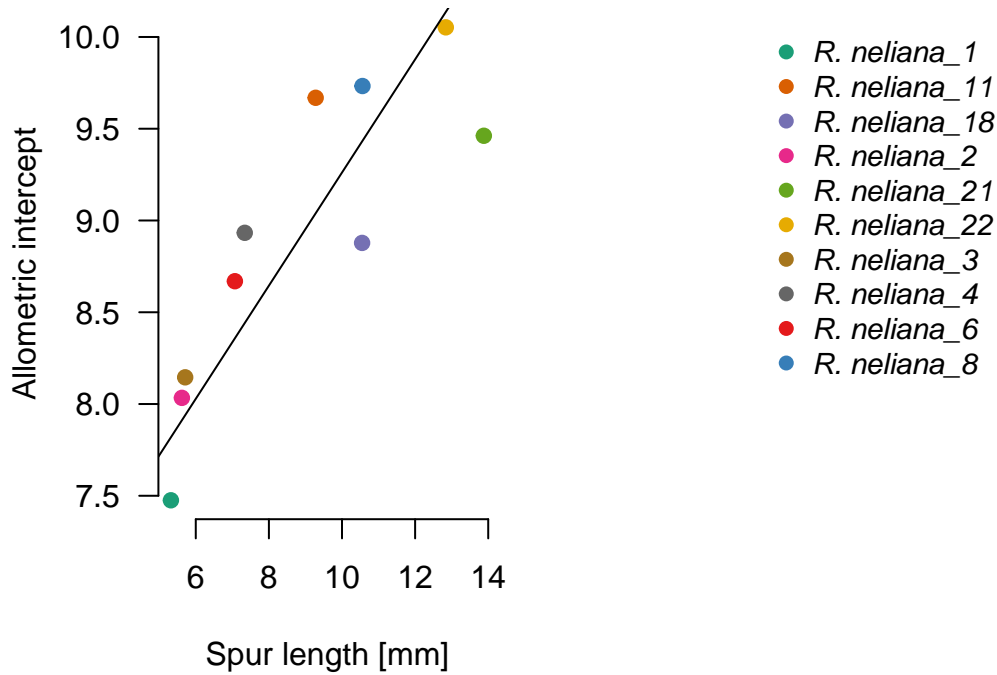


Figure S11: Relationship of allometric intercept (i.e. estimated population-specific leg length from Model m11) to spur length of the host plants for *Rediviva neliana* populations, fitted on the arithmetic scale.

## References

- Egset, C. K., G. H. Bolstad, G. Rosenqvist, J. A. Endler, and C. Pélabon. 2011. Geographical variation in allometry in the guppy (*Poecilia reticulata*). *Journal of Evolutionary Biology* 24:2631–2638.
- Heisler, I. L., and J. Damuth. 1987. A Method for Analyzing Selection in Hierarchically Structured Populations. *American Naturalist* 130:582–602.
- Kahnt, B., G. A. Montgomery, E. Murray, M. Kuhlmann, A. Pauw, D. Michez, R. J. Paxton, and B. N. Danforth. 2017. Playing with extremes: Origins and evolution of exaggerated female forelegs in south african rediviva bees. *Molecular Phylogenetics and Evolution* 115:95–105.
- Pol, M. van de, and J. Wright. 2009. A simple method for distinguishing within- versus between-subject effects using mixed models. *Animal Behaviour* 77:753–758.