Tutorial for: Sexual selection as the evolutionary driver of sex-difference in adult lifespan in mammals and birds

Johanna Staerk, Dalia A. Conde, Morgane Tidière, Orsolya Vincze, Samuel Pavard, Mathieu Giraudeau, Zjef Pereboom, Mads F. Bertelsen, Simeon Q. Smeele, Rita da Silva, Andras Liker, Balázs Vági, Jean-François Lemaître, Jean-Michel Gaillard, Tamás Székely, Fernando Colchero

Contents

Brief introduction

Here we present code to visualize the life tables presented on the manuscript, to calculate life expectancy from Siler mortality parameters, and run the Bayesian PGLS analyses.

The .zip file includes three directories, namely 01code, 02data, and 03tutorial. The 01code directory includes two R script files, ALEdifferencesFunctions.R and ALEdifferencesCode.R. This last is the script you should work with, unless you choose to change any of the analysis functions.

Load functions, libraries, and data

First you will need to point the R code in ALEdifferencesCode.R to the path where you have saved the directories:

```
# Set working directory:
setwd("path to directory...")

# Source functions:
source("01code/ALEdifferencesFunctions.R")
```

The second line calls the analysis functions while also loading the relevant R libraries. The following lines load the different data tables used for this tutorial. There are several tables under the directory `02data/tables/`:

- `LifeTables.csv`: life tables obtained from the BaSTA analysis per species and sex.

- `SilerParamsFromMatur.csv`: estimated Siler mortality parameters per species and sex;
- `ALEdifferences.csv`: posterior mean, SD, and lower and upper quantiles for life expectancy per sex and for ALE differences;
- `PLGSdata.csv`: data used to run the BPGLS including the ALE differences as response variable, and the predictors used for analysis.
- `indivTestDat.csv`: Individual level data for BaSTA simulated from the Siler mortality parameters for female and male chimpanzees (**Pan troglodytes**).

In the attached R script, the tables are called in as:

```
# Load life tables:
lifeTabs <- read.csv(file = "02data/tables/LifeTables.csv",
                     header = TRUE, stringsAsFactors = FALSE)

# Load Siler mortality parameters from maturity:
thetaMat <- read.csv(file = "02data/tables/SilerParamsFromMatur.csv",
                     header = TRUE, stringsAsFactors = FALSE)

# Load estimated ALE differences:
leDiff <- read.csv(file = "02data/tables/ALEdifferences.csv",
                   header = TRUE, stringsAsFactors = FALSE)

# Load data for BPGLS:
pglsDat <- read.csv(file = "02data/tables/PLGSdata.csv",
                    header = TRUE, stringsAsFactors = FALSE)

# Test data for BaSTA analysis:
survDat <- read.csv(file = "02data/tables/indivTestDat.csv",
                     header = TRUE, stringsAsFactors = FALSE)
```

Under the directory `02data/rdata/` we include the file `fullphylo.RData` with the combined phylogeny for birds and mammals, which can be loaded as

```
# Load phylogeny:
load("02data/rdata/fullphylo.RData")
```

Plotting life tables

All the life tables used in the study are stored in the lifeTabels.csv object. As an example, we show how to plot the life table survival for female African elephants (*Loxodonta africana*):

```
# ================================== #
# ==== PLOT LIFE TABLE SURVIVAL: ====
# ================================== #
# Chose species:
species <- "Pan troglodytes"
```

The code that follows (lines 46-72) produces the plot below:

Calculate life expectancy

Theoretical background

Here we use parametric mortality functions to calculate life expectancy, which is a summary statistics of the distribution of ages at death. Thus, let X be a random variable for ages at death, with realizations $x \geq 0$.

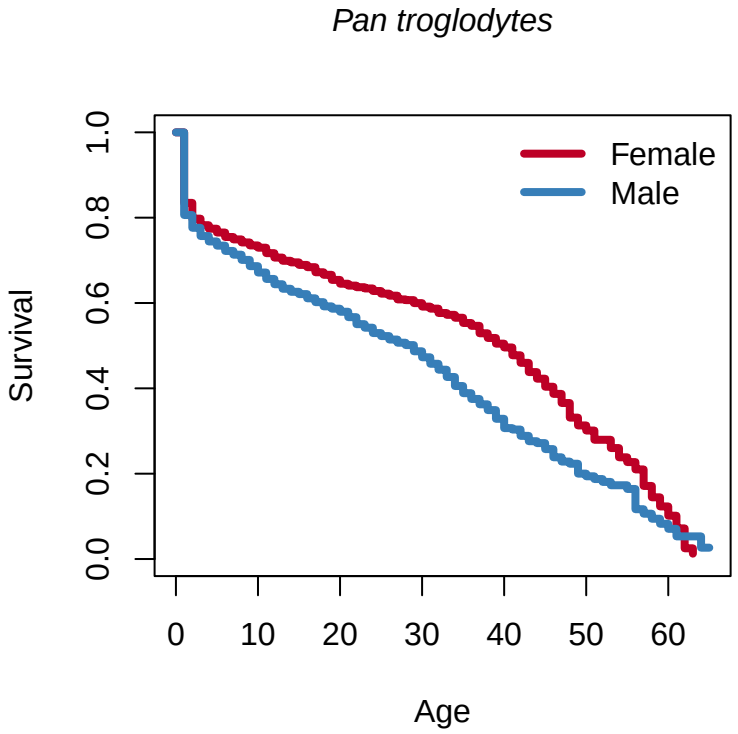

Figure 1: Example of life table survival plot for female and male chimpanzees in ZIMS.

The age-specific hazard rate or mortality function is

$$\mu(x) = \lim_{\Delta x \to 0} \frac{\Pr[x < X < x + \Delta x | X > x]}{\Delta x}. \tag{1}$$

From the hazard rate or mortality function in Eq.~1, we calculate the cumulative hazard, given by

$$U(x) = \int_0^x \mu(t)dt, \tag{2}$$

from which the survival or cumulative survival function is

$$S(x) = e^{-U(x)}, \tag{3}$$

with complement given by the cumulative distribution function $F(x) = 1 - S(x)$ The probability density function of ages at death is given by

$$f(x) = \mu(x)S(x). \tag{4}$$

As the name suggests, life expectancy is the expected value of the age at death and, as an expectation, it is calculated from the distribution of ages at death as

$$\begin{aligned} e_0 &= \int_0^\infty x f(x)dx \\ &= \int_0^\infty S(x)dx. \end{aligned} \tag{5}$$

Summary variables from Siler model

Here, we use the Siler mortality function (Siler 1979), given by

$$\mu(x|\boldsymbol{\theta}) = \exp(a_0 - a_1 x) + c + \exp(b_0 + b_1 x), \tag{6}$$

where $\boldsymbol{\theta}^\top = [a_0, a_1, c, b_0, b_1]$ is a vector of Siler mortality parameters, where $a_0, b_0 \in \mathbb{R}$ and $a_1, c, b_1 \geq 0$.

The cumulative hazard of the Siler model is

$$U(x|\boldsymbol{\theta}) = \frac{e^{a_0}}{a_1}(1 - e^{-a_1 x}) + cx + \frac{e^{b_0}}{b_1}(e^{b_1 x} - 1). \tag{7}$$

We use the R package BaSTA (Colchero and Clark 2012, Colchero et al. 2012, Colchero et al. 2021) for inference on age-specific survival for both sexes. As an example, we provide a test dataset on which to run BaSTA:

```
# ======================================== #
# ==== AGE-SPECIFIC SURVIVAL ANALYSIS: ====
# ======================================== #
# Find life history data for chosen species:
idSps <- which(lifehist$species == species)

# Find age at maturity:
alpha <- floor(max(c(lifehist$AFB_Female[idSps], lifehist$AFB_Male[idSps])))

# Run BaSTA on survival data:
out <- basta(object = survDat, dataType = "census", model = "GO",
             shape = "bathtub", formulaMort = ~ Sex - 1, minAge = alpha,
             nsim = 4, parallel = TRUE, ncpus = 4)
```

The results of BaSTA can be visualized as with other inferences functions with functions `summary` or `print`:

```
# Print output to the console:
print(out, digits = 3)
```

```
>
> Call:
> Model                    : GO
> Shape                    : bathtub
> Minimum age              : 8
> Covars. structure        : fused
> Cat. covars.             : SexFemale, SexMale
> Cont. covars.            :
>
> Coefficients:
>                 Mean  StdErr Lower95%CI Upper95%CI SerAutocorr UpdateRate
> a0.SexFemale -3.2199 0.73890   -4.73414    -1.8452    -0.01431      0.254
> a0.SexMale   -2.9482 0.69934   -4.40681    -1.6448     0.05609      0.235
> a1.SexFemale  6.7059 3.27535    1.55301    14.3569     0.00877      0.253
> a1.SexMale    3.9525 2.96446    0.33787    11.2997     0.05910      0.242
> c.SexFemale   0.0104 0.00131    0.00784     0.0131     0.25015      0.234
> c.SexMale     0.0176 0.00252    0.01282     0.0218     0.53007      0.253
> b0.SexFemale -9.9646 0.74011  -11.55663    -8.5701     0.90442      0.249
> b0.SexMale   -9.0484 1.02867  -11.08935    -7.0209     0.89520      0.259
> b1.SexFemale  0.1609 0.01560    0.13059     0.1940     0.90545      0.255
> b1.SexMale    0.1352 0.02152    0.09198     0.1770     0.89516      0.258
> lambda        0.0295 0.00358    0.02302     0.0367     0.04592      0.239
>              PotScaleReduc
> a0.SexFemale          1.00
> a0.SexMale            1.00
> a1.SexFemale          1.00
> a1.SexMale            1.00
```

```
> c.SexFemale           1.00
> c.SexMale             1.01
> b0.SexFemale          1.00
> b0.SexMale            1.00
> b1.SexFemale          1.00
> b1.SexMale            1.00
> lambda                1.00
>
> Convergence:
> All parameters converged properly.
>
> DIC = 5822.41
```

while different plots can be produced, such as plots of goodness of fit as:

```
# Plot goodness of fit plots:
plot(out, plot.type = 'gof')
```

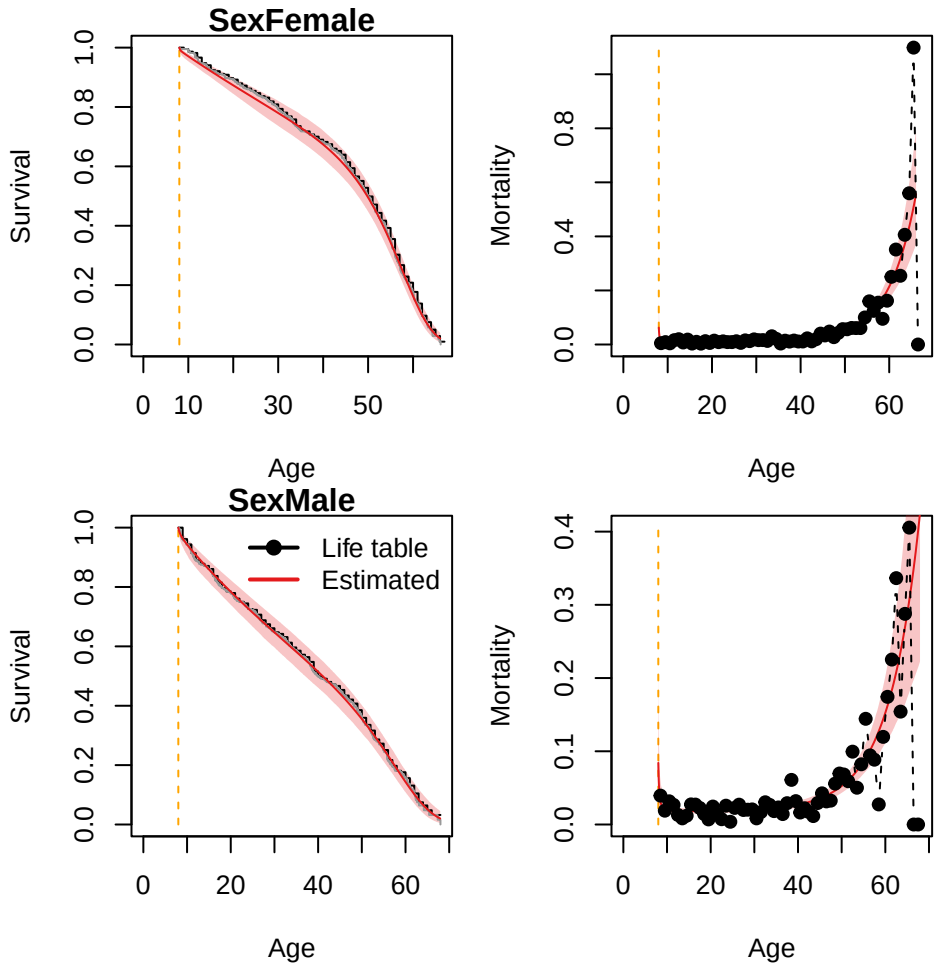

Figure 2: Goodness of fit plot from BaSTA output on the test data provided with this tutorial.

To calculate the vector of posterior estimates of life expectancy, we use a left-hand quadrature to find the integral in Eq.~5 for each vector of Siler parameters per MCMC iteration after burn-in and thinning. Thus, in the R script section on Calculating life expectancy, we show how to estimate it for female chimpanzees in ZIMS, but the code can be adapted to any of the species or sexes:

```
# ============================================================ #
# ==== CALCULATE POSTERIOR LIFE EXPECTANCY FROM BaSTA OUTPUT: ====
# ============================================================ #
# Calculate posterior for ALE per sex and of ALE differences:
exFull <- CalcPostALE(out, ncpus = 4)

# Calculate summary statistics (posterior mean, SD, and quantiles):
quants <- CalcQuants(object = exFull)
```

The results in the `quants` object can be printed to the console as

```
print(quants, digits = 4)
```

```
>               Mean      SD    Lower   Upper zeroOverlap
> ALE_Female 36.5315 0.88673 34.66901 38.2232          NA
> ALE_Male   31.0053 1.04874 28.74186 33.0874          NA
> ALEdiff     0.1507 0.03605  0.06892  0.2195    2.89e-05
```

Bayesian PGLS

As we describe in the methods section of the manuscript, we implemented weighted Bayesian PGLS (BPGLS) with the R package BayesPGLS between the ALE differences, δ_e, and life-history predictors associated with our hypotheses. We obtained phylogenetic trees from Upham et al. (2019 a, b) for mammals and Jetz et al. (2012) for birds. We computed the maximum clade credibility using the phangorn R package (Schliep, 2011) from a sample of 100 trees based on the birth-death node-dated trees for mammals and the Ericson All Species Tree using the VertLife Phylogeny subsets tool. For BPGLS with both classes combined, we grafted the two phylogenies by means of the R package geiger (Pennell et al. 2014) assuming a median divergence time of 319 MY, as provided by timetree.org.

Package BayesPGLS uses MCMC with direct sampling for the regression parameters with Metropolis-Hastings to estimate Pagel's λ_p ($0 \leq \lambda_p \leq 1$), which provides an estimate of the intensity of the phylogenetic signal (Pagel, 1999). Given that we used estimated ALE differences from Bayesian survival trajectory analyses on species with different sample sizes and other sources of uncertainty, we used the posterior standard errors of δ_e, σ_δ, to calculate weights for the regression. The priors for the regression parameters were all normally distributed with mean 0 and variance 100.

Below is an example of a BPGLS for artiodactyls testing the effect of sexual size dimorphism and mating system on ALE differences:

```
# ================================================ #
# ==== EVOLUTIONARY PREDICTORS OF ALE DIFFS.: ====
# ================================================ #
# Subset data to only include Artiodactyls:
idOrder <- which(pglsDat$Order == "Artiodactyla")
pglsDatOrder <- pglsDat[idOrder, ]

# Run BPGLS:
outpgls <- RunBayesPGLS(formula = exDiff ~ log(MaleBM) + log(FemaleBM) + MS,
                        data = pglsDatOrder, weights = "weights",
                        phylo = fullphylo, ncpus = 6, nsim = 6)
```

Which can be visualized by printing the results to the console as

```
print(outpgls)
```

```
> Model:
```

```
> exDiff ~ log(MaleBM) + log(FemaleBM) + MS
>
> Coefficients:
>                   Mean     SD   Lower    Upper zeroCoverage Rhat
> (Intercept)    -0.3221 0.1544 -0.6288 -0.02317        0.037    1
> log(MaleBM)     0.1610 0.0673  0.0275  0.29316        0.017    1
> log(FemaleBM) -0.1428 0.0692 -0.2802 -0.00686        0.039    1
> MSYes         -0.0689 0.0727 -0.2119  0.07163        0.343    1
> sigma          0.0746 0.0200  0.0439  0.12124         <NA>    1
> lambda         0.8221 0.0767  0.6411  0.93630         <NA>    1
>
> DIC = -379.561
```

while different plots can be produced, such as a plot of the posterior densities for the parameters

```
plot(outpgls, plot.type = 'density')
```

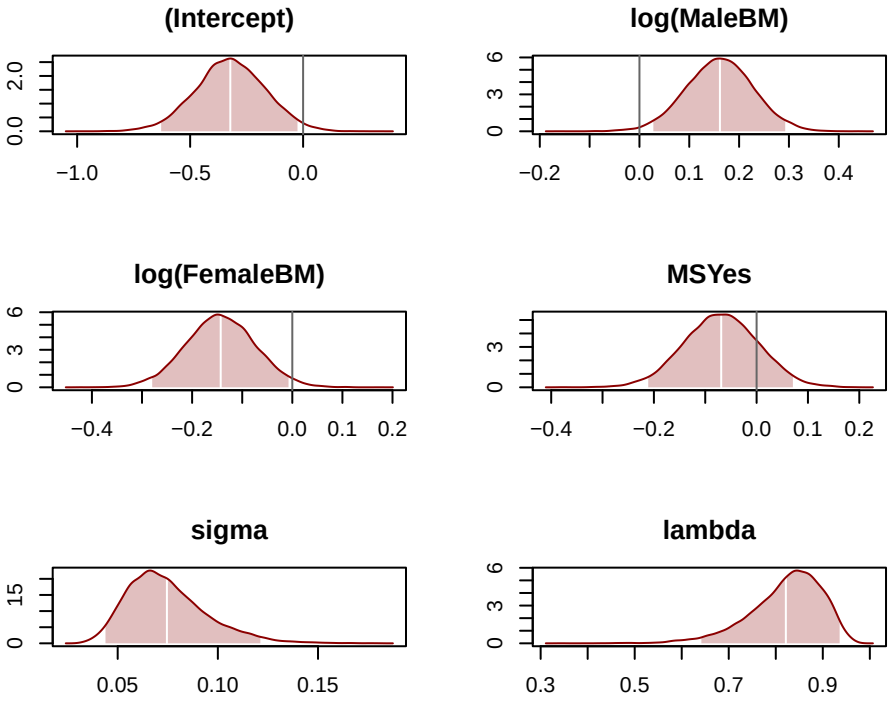

Figure 3: Posterior densities of the regression and Pagel s lambda parameters from BayesPGLS

Effect of allometry on regressors

As we show in the methods section, it is possible to account for the effect of allometry within a regression framework. Let Y and X be two variables allometrically related as

$$Y = \alpha X^{\gamma}, \tag{8}$$

where $\alpha, \gamma > 0$ are the scale and power coefficients for the power function. By taking the log on both sides of Eq. 8 we have that

$$\log Y = \log \alpha + \gamma \log X. \tag{9}$$

A new variable that accounts for the allometric relation between X and Y can be estimated as

$$A = \log \alpha = \log Y - \gamma \log X. \tag{10}$$

The variable A in Eq. 10 is commonly used for sexual size dimorphism, where Y is the male body mass and X is the female body mass, or to estimate relative masses, for instance where Y is testes mass and X is male body mass (e.g., testes mass relative to body mass).

Let Z be a random variable for a given response, in our case ALE differences, which we assume is linearly related to our variable A. We have that, given the linear model, the expected value of Z is

$$\begin{aligned} \hat{Z} &= \eta + \beta A \\ &= \eta + \beta \left[\log Y - \gamma \log X\right] \\ &= \eta + \beta \log Y - \beta\gamma \log X, \end{aligned} \tag{11}$$

where $\eta, \beta \in \mathbb{R}$ are the intercept and slope of the linear regression. Alternatively, we can estimate the expected value of Z as

$$\hat{Z} = \eta + \beta_1 \log Y + \beta_2 \log X \tag{12}$$

where η is the intercept as in Eq. 11, and $\beta_1, \beta_2 \in \mathbb{R}$ are slope parameters. By equating Eqs. 11 and 12 it is simple to show that $\beta_1 = \beta$ and $\beta_2 = -\beta\gamma$, and thus $\gamma = -\beta_1/\beta_2$. In other words, the effect of A is given by β_1 while γ can be obtained from the ratio between the two slope coefficients.

In the following code section, we show that this relationship holds, and that both β and γ can be retrieved within a regression framework on simulated data. The data are simulated as

```
# Number of observations:
n <- 1000

# Alpha:
al <- 1

# Gamma:
gam <- 0.8

# Simulate X (e.g., male body mass)
x <- runif(n = n, 1, 100)

# Variability in allometry:
del <- 2

# Simulate Y (e.g., testes mass)
y <- exp(log(al) + gam * log(x) + rnorm(n = n, mean = 0, sd = del))

# Simulate predictor (e.g., relative testes mass, RL):
xy <- log(y) - gam * log(x)

# Regression coefficients between response and predictor (e.g., ALE diffs ~ RL)
sig <- 0.5
eta <- 0.5
bet <- 1

# Simulate response Z (e.g., ALE diffs.):
z <- eta + bet * xy + rnorm(n = 100, mean = 0, sd = sig)

# Run regression between Z and log Y, log X:
reg1 <- lm(z ~ log(y) + log(x))

# Run regression between Z and XY (e.g., RL):
reg2 <- lm(z ~ xy)
```

We show that the parameters are retrieved when the variables X and Y are included as predictors. Here are the outputs for β and β_1:

```
> Beta for simulation: 1
```

```
> Beta from regression 2: 1.003
```

```
> Beta from regression 1: 1.003
```

While here are the outputs for γ and $-\beta_2/\beta_1$:

```
> Gamma for simulation: 0.8
```

```
> Estimated gamma: 0.795
```

References

Colchero F. *et al.*, The long lives of primates and the 'invariant rate of ageing' hypothesis. Nature Communications. **12**, 3666 (2021).

Colchero, F., Clark, J. S., Bayesian inference on age-specific survival for censored and truncated data. Journal of Animal Ecology. 81, 139-149 (2012).

Colchero, F., Jones, O. R., Rebke, M., BaSTA: an R package for Bayesian estimation of age-specific survival from incomplete mark–recapture/recovery data with covariates. Methods Ecol Evol. 3, 466-470 (2012).

Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N, Pearse W (2023). caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 1.0.3, https://CRAN.R-project.org/package=caper.

Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, M. E. Alfaro, and L. J. Harmon (2014) geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics, 30, 2216-2218.

Revell, L. J. (2024) phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). PeerJ, 12, e16505.

Siler, W. A competing-risk model for animal mortality. Ecology. **60**, 750-757 (1979).