Modelling allometry: statistical and biological considerations – a Reply to Packard

Christophe Pélabon1, Morgane Tidière1, Jean-François Lemaître2, Jean-Michel Gaillard2

1 Department of Biology; Centre for Biodiversity Dynamics, NTNU, Norwegian University of Science and Technology, 7491 Trondheim, Norway.

2 Université Lyon 1; CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F–69622, Villeurbanne, France.

**Abstract**: Allometry studies describe how phenotypic traits increase relative to the increase of the size of the organism. Because the increase in organism’s size is due to growth, a multiplicative process, allometric relationships are often analysed on a proportional scale (e.g. a log-log scale) to account for the multiplicative nature of the data. Still, the log-transformation of the data when estimating allometric relationships has been the subject of debate. In a series of replies to various case studies of allometry, G.C. Packard has repeatedly criticized this approach under the premise that the log-transformation of the data alters the estimate of the allometric exponent, and obscures the biological meaning of the allometric parameters. Recently, Packard (2018) reanalysed data from our study on horn length allometry in Bovids (Tidière *et al.*, 2017) and reached conclusions that contrasted with those reported in our original study. Echoing many authors before us, we argue here that log-transformation of the data in allometric studies is justified by the expected distribution of the residual variation in ontogenetic, static, and evolutionary allometry. We also point out that allometric slopes thus obtained have a direct biological interpretation in terms of elasticities. Finally, we show that Packard’s criticism is based on qualitative and not quantitative assessment of the models fitted on different scales, and his conclusions disregard statistical and biological evidence supporting models fitted on a log-log scale.

Additional keywords: scale, measurement error, biological error, elasticity, fitness function

**Introduction**

Allometric relationships describe the covariation generally observed between morphological, physiological, and life history traits, and the size of the organisms (Huxley, 1924, 1932; Huxley & Tessier, 1936). To explain these relationships, Huxley (1932) hypothesized that both the size of a trait *y* and the size of the rest of the organism *x*, are controlled by a common growth factor. It follows that the relative growth of *y* and *x* are linked by a power function

 *y = x*, (1)

where ** is a constant and ** the allometric exponent (see also Savageau, 1979).

The allometric exponent **, measures the increase in trait size relative to the size of the organism, and can inform thereby on the presence of developmental or genetic constraints in ontogenetic and static allometry (i.e. allometry measured during ontogeny within individuals and, at similar developmental stage among individuals of different size within a population, respectively ; Lande, 1979, 1985). In evolutionary allometry (i.e. allometry across populations or species), ** may reflect either the effect of different selective optima of the trait size relative to the body size across populations or species (i.e. selective constraint) or the effect of developmental (genetic) correlations between trait size and body size that constrain population and species divergence (the allometry-as-a-constraint hypothesis, Lande, 1979, 1985; Voje *et al.*, 2014; see Pélabon *et al.*, 2014 for a review).

 Following Huxley (1932), allometric relationships are generally analysed on log-transformed data. On a log-log scale, the power function becomes linear:

 (2)

where *log()* is the allometric intercept, and ** is the allometric slope.

 The use of log-transformed data for studying allometric relationship has been long debated, however (e.g. Smith 1980, 1984, Bales 1996; Sartori and Ball 2009, Packard 2015), possibly starting with D’Arcy Thompson (1942) re-analysing data from Huxley without log-transformation (cited in Smith, 1980). During the past years, the debate has been fed by G.C. Packard and colleagues in a series of more than 20 papers where they reanalysed data from published studies by fitting non-linear models to the data on the arithmetic scale. The main justification provided by these authors for doing so is that log-transformation alters the data so that predictions from the models on the arithmetic scale (i.e. after back-transformation) are no longer accurate, and the interpretation of the parameters is unclear.

 In our study of the evolutionary allometry of horn length and body mass in Bovids (Tidière *et al.*, 2017), we found that the evolutionary allometry in males is non-linear on a log-log scale, with a decreasing allometric slope when the size of the species increases. This pattern, influenced by the phylogenetic relationships among species, resulted from an absence of horn length increase with an increase in mean mass among species larger than 350 kg. In females, the evolutionary allometry between horn length and body mass was linear, and was also influenced by the phylogenetic relationship among species. Furthermore, sexual dimorphism in horn length in species with both sexes carrying horns reached a maximum for horn length of ca. 32 cm in females and subsequently decreased with increasingly larger female horns. We interpreted these results as reflecting a change in selection pressures on horn size in males from small to large species, and proposed that in the largest species, horn length is replaced with body mass as the target of sexual selection in males. Packard (2018) challenged these conclusions. After arguing that the log-transformation of the data altered their distribution “*in ways that may conceal flaws in the original data*” (text in italics from Packard, 2018) and that allometric models should be validated with graphical analyses on the arithmetic scale, he adjusted a three-parameter power function () that, according to him, better fitted the data. From his reanalysis, Packard concluded that horn length of males and females had the same allometric exponent **, and the same intercept *y0*, and only differed by the constant, **. Several authors have already pointed out the problems with Packard’s arguments against the use of log-transformed data (see e.g. Kerkhoff & Enquist, 2009; Cawley & Janacek, 2010; Ballantyne, 2013; Glazier, 2013; Mascaro *et al.*, 2014; Lemaître *et al.*, 2015; Xiao *et al.*, 2011). Here, we summarize the most salient points of disagreement, and develop some new arguments in support of the use of log-transformed data in allometry analyses.

Due to the assumptions associated with the distribution of the residuals in least square regression (normal distribution and homoscedasticity of the residuals), the debate on the use of log-transformed data in allometry studies has mostly focused on the scale of the residual variation (normal *vs.* log-normal). Here, we first discuss the different sources of residual variation in ontogenetic, static and evolutionary allometry, and we argue that a log-normal distribution of the residuals is generally expected, therefore justifying the log-transformation of the data when studying allometry. We then show that the allometric slope estimated from models fitted on log-transformed data has a particularly meaningful biological interpretation. Finally, we point out that Packard’s approach is qualitative and does not provide a proper quantitative assessment of the different approaches.

**Meaning and distribution of the residual variance**

The main criticism of allometric studies based on log-transformed data formulated by Packard and colleagues states that the error term in the model becomes, by definition, multiplicative on the arithmetic scale. Effectively, additive error on a log-log scale:

, with , (3)

becomes multiplicative on the arithmetic scale:

 . (4)

According to Packard and colleagues, this is a problem because the use of a proportional scale decreases the relative contribution of the large species to the fit of the model (Packard *et al.*, 2009). Consequently, when back-transformed to the arithmetic scale, models generally poorly fit the distribution of the largest species (Packard & Birchard, 2008; Packard, 2011). As an alternative, these authors propose to fit a non-linear regression on the arithmetic scale that assumes additive errors (equation 3 in Packard, 2017). Then, one should test for the normality and homoscedasticity of the residuals. If the distribution of the residuals deviates from normality and/or shows signs of heteroscedasticity, specific variance-mean relationship should be modelled to account for this deviation.

First, it is noteworthy that in the case of heteroscedastic residuals with a log-normal distribution on the arithmetic scale, fitting a non-linear model on an arithmetic scale with log-normal residuals should provide the same results as fitting a linear model on a log-log scale with normally distributed residuals. The allometric exponent ** should equal to the slope of the linear regression, and the intercept of the linear regression should equal to the logarithm of the constant **in the non-linear model on the arithmetic scale. In this case, both models are equally valid, but such a correspondence does not hold anymore for three-parameter non-linear models. However, modelling *ad hoc* distributions for the residual variance to fit non-linear models on the arithmetic scale, as advocated by Packard, does not account for the theoretical foundation of allometric relationships, and it may lead to overlook information concealed in the distribution of the residuals when this one deviates from theoretical expectations.

Deviation from the allometric slope (i.e. residuals) results from the combination of biological and measurement errors (Riska, 1991). In ontogenetic allometry assessed with transversal data and in static allometry, biological errors correspond to any developmental variation among individuals that is caused by stochastic (i.e. developmental noise) or environmental perturbations of the growth of the trait *y* relative to the growth of the trait *x*. Such perturbations may affect either the constant **, or the allometric exponent **, and can be modelled as:

, (5)

where *i* is the error affecting **, and *i* is the error effecting **. Deviations from the power function generated by *i* will equal , and their distribution will depend on the distribution of *i*. Because ** corresponds to the value of *y* when *x* equals 1 (whatever the unit of *x*), the distribution of *i* depends on the distribution of the error in *y*. Within a population, traits are often normally distributed, and we expect *i* to be normally distributed as well. The error affecting the allometric exponent ** will generate deviations from the power function equal to , that is, equivalent to the *exp(i)* in eq. 4. If *i* is normally distributed, the residuals of the allometric relationship will be log-normally distributed. The biological error being the product of these two errors, its distribution will correspond to the product of a normal and a log-normal distribution and will be log-normal.

Measurement error is the error made during the measurement procedure and its magnitude depends on the precision of the measuring device. If the trait *y* is always measured with the same instrument and assuming that the accuracy of this instrument does not depend on the size of the attribute measured, we expect the measurement error to be normally distributed with a mean of zero and a variance *2* on the measurement scale, that is, on an arithmetic scale. Including both types of error in the power function gives:

 , (6)

where is the measurement error. The log-transformation of the data renders the biological error additive, with a constant variance as *x* increases, but diminishes the impact of measurement error on the residual variance when *x* increases. As long as the measurement error is much smaller than the biological error, as it is often the case, the distribution of the residuals will be approximately log-normal. However, if the measurement error becomes non-negligible compared to the biological error, the distribution may differ from a log-normal distribution, and we could expect the residual variance to decrease with increasing *x*.

Evolutionary allometry may result from developmental constraints affecting the trajectory of the population or species divergence (Lande, 1979; 1985; Voje *et al.*, 2014). In this case, the biological error results from the adaptation of populations or species to specific environmental conditions that drive the size of the trait *y* relative to the size of the trait *x* away from the expected allometric trajectory. Alternatively, evolutionary allometry may result from correlated selection generating adaptive ridge (Armbruster & Schweagerle, 1996; Armbruster *et al.*, 2014). In both cases, the scale of the biological error depends on the scale of the fitness landscape on which populations or species evolve. At first sight, we may consider that variation in fitness associated with the variation in relative trait size across populations or species is on a proportional scale. For example, we do not expect the same change in fitness for a change in beak depth of 0.5 mm in *Geospiza fuliginosa* (mean beak depth ca. 7 mm) and in *G. magnirostris* (mean beak depth ca. 16 mm, means from Schluter & Grant, 1984), but we rather expect that a 5% change in beak depth in both species has a comparable effect on their fitness. In this case, we expect a log-normal distribution of the residuals. In some cases, however, similar differences in absolute size may generate similar effects on fitness across populations or species of different sizes. For example, the fit between floral reproductive organs (i.e. anther and stigma) and insect pollinators may generate a fitness landscape that is defined by the absolute size of the pollinators, generating an adaptive ridge defined on the arithmetic scale (Armbruster *et al.*, 2009), and residuals normally distributed on the arithmetic scale. Therefore, the scale of the biological error in evolutionary allometry is an empirical question, but multiplicative error (i.e. log-normally distributed) may often be a reasonable starting point.

Measurement errors in evolutionary allometry are too rarely estimated. This may stem from the complexity of these measurement variances that combine different sources of errors. First, both the *x* and *y* traits are often measured with different tools for species of very different size. In this case, it is likely that the magnitude of the measurement error varies across the size range. Measurement error can also result from sampling variance that generates errors in the estimation of population or species means when these means are estimated independently for *x* and *y*, as it is often the case when data for the two traits come from different studies. Whether or not these different sources of error and/or their combination translate into residual variance proportional to the trait measured is generally unknown. However, as for ontogenetic and static allometry, measurement errors in evolutionary allometry are generally assumed to be small compared to biological errors.

These considerations suggest that theory generally predicts multiplicative error in allometric relationships as long as the measurement error is negligible compared to the biological error and that variation in fitness is related to proportional variation in trait size across populations or species in evolutionary allometry. Importantly, our capacity to distinguish between normal and log-normal distribution of the residuals depends on the range of variation of the *x* trait. With limited variation, as it is often the case in static allometry, normal and log-normal distribution may be very difficult to distinguish, as illustrated in the study by Gingerich (2000). Therefore, although a log-normal distribution is the theoretical expectation for the residuals in allometric studies, it may be difficult to distinguish it from a normal distribution when variation in the traits is limited.

Consequently, we agree with Packard and colleagues that the distribution of the residuals should be carefully examined when performing any allometric study. However, we disagree with these authors when they advocate modelling any type of residual variance after fitting a non-linear regression on the arithmetic scale, because this *ad hoc* modelling of the error distribution solely based on statistical arguments considers the distribution of the residuals as non-informative and does not account for biology-grounded theoretical expectations. Instead, we consider the log-normal distribution of the residuals as a null hypothesis which can be disproved, and we see deviation from this null hypothesis as a valuable source of information that can help understanding the nature of the residual variance.

In our study (Tidière *et al.*, 2018), after fitting a quadratic model on a log-log scale, residuals were normally distributed and independent from the predictor variable, as expected from the theory. Although this was acknowledged by Packard (2018) who used log-normally distributed residuals to fit his models, it was apparently not sufficient to justify the log-transformation of the data.

**Interpreting the allometric slope on a log-log scale**

Packard and colleagues have repeatedly stated that allometric relationships should not be interpreted on a log-log scale because the main purpose of allometric studies is to make prediction of the size of a trait (i.e. on the measurement scale) when only body size is known (Packard, 2011; 2018). Although this procedure is common in palaeontology (Grabowski *et al.*, 2015), this perspective ignores a large body of literature where the allometric slope ** is the main focus of the study (Eberhard *et al.*, 1998; Bonduriansky, 2007; Pélabon *et al.*, 2014; Voje, 2016). On a log-log scale, the allometric slope represents the proportional change in a trait *y* that occurs in response to a proportional change in trait *x*, that is, the percent change in *y* for a 1% change in *x*. Such proportional sensitivity indices have been called elasticities by microeconomists and they are particularly useful in biology, especially when traits have different units or dimensions (De Kroon *et al.*, 1986; Van Tienderen, 1995, 2000). To illustrate the use of elasticities in interpreting biological variation, let us consider an example of ontogenetic allometry between horn size and body mass in a hypothetical bovid species. Say the allometric coefficient is 1.3 in males, and 1.05 in females. If we try to illustrate the difference between sexes in horn growth on the arithmetic scale, we need to consider at which body mass the comparison is done. Indeed, for an increase of one kg between 14 and 15 kg, male horns will increase by 0.86 mm while they will increase by 0.41 mm in females. However, for a same increase of one kg between 41 and 42 kg, the horn growth in males and females will be 1.16 and 0.44mm, respectively (we consider similar intercepts for the ease of comparison). Using a log-log scale, these figures can be interpreted as follow: for a 1% increase in body mass, the horn length will increase by 1.3% in males, but only by 1.05% in females, whatever the size of the animals.

 For complex allometries, as long as the traits are expressed on a log-log scale and that the model describing their relationship is a linear model, it is possible to define an allometric slope between *x* and *y* that can be interpreted as an elasticity. This can be achieved by using broken-stick regression (i.e. piecewise linear function), where each segment represents the elasticity of the two traits between the limit defined by the model (e.g. Lemaître et al. 2014). Alternatively, one can use a quadratic regression (on a log-log scale) that allows estimating the elasticity of the two traits for a given value of log(*x*) by calculating the derivative of the function (Tidière et al., 2017).

 To account for non-linear allometry, Packard (2018) used a three-parameter power function: . With such a function, the slope of the derivative at specific values of *x* does not correspond anymore to the derivative of the quadratic function on a log-log scale, and it cannot be interpreted as an elasticity. Furthermore, as noticed by Packard himself, the parameter *Y0* has no biological interpretation. Yet, in an attempt to provide an interpretation of this parameter, Packard (2018) explained that the important parameter is not *Y0*, but the *x* intercept (i.e. when the curve passes through the *x* axis) that he interpreted as the size under which Bovids could not wear horns. To support his claim, Packard stated that the smallest species with horns in both sexes had this expected mass. This interpretation appears to us as being “*ad hoc*” and not convincing. It implies that the size of horns relative to body size in Bovids is governed by the same process in both sexes and in all species, ignoring obvious differences in selection on horn length among species of different size and the fact that in many species females are not wearing horns.

**Qualitative vs. quantitative assessment of the alternative models**

Packard (2018) claimed that models on arithmetic scale fit better the data than those fitted on a log-log scale. However, this claim is only based on visual (qualitative) assessment of the fit of the models and not on proper quantitative comparison between models fitted on different scales (see also Packard, 2012 for a similar issue). In some earlier criticisms of models fitted on a log-log scale, Packard (e.g. Packard & Birchard, 2008) mistakenly used r-square values for comparing the models on the two scales. However, due to the non-linear transformation that affects the distribution of the residuals, r-square values, as well as AIC values, cannot be used to compare models on different scales without correction (see Ballantyne, 2013 for appropriate correction of AIC values). To avoid this problem, Packard (2018) only compared visually models fitted on arithmetic scale or log-log scale. This procedure is disputable, especially because the assessment is done on graphs on the arithmetic scale, while the model is fitted with a log-normally distributed error.

To compare quantitatively the fit of the different models, we used the correlation between the observed and the predicted values for horn length provided by the different models. Because Packard fitted his model with a log-normal error, we compared observed and predicted values using log-transformed values, and calculated the r-square between the observed and predicted values for the two types of model (Fig. 1). This analysis revealed that, although the difference is small, our models better fit the data (higher r-square). The poorer fit of Packard’s model is also revealed by the slightly wider distribution of the residuals of the non-linear model that results from the inability of this model to predict the particularly small horn length of males from the largest species (Fig. 2). Packard acknowledged this problem when he wrote “*the distribution of data for males is slightly peculiar…*”. But surprisingly, instead of questioning his model, Packard criticized the “*suitability of the data for use in an allometric analysis*”. Such a practice does not match the commonly accepted way of modelling biological processes where the model should be fitted to the data and not the other way round, and represents another example of the predominance of statistics at the expense of the meaning in the analysis of biological data (Houle *et al*. 2011). This problem is clearly illustrated by the inconsistencies between the predictions drawn from Packard’s model and the patterns observed in the data. Indeed, based on his non-linear model, Packard concluded that both horn length and sexual dimorphism in horn length increase monotonically with increasing body mass in Bovids. The data show, however, whatever the scale on which they are represented, that these conclusions are both erroneous. The longest horns are observed for medium-sized Bovid species, while for species above 230 kg (*Hippotragus niger*), there is no more increase in horn length with increasing body mass. Similarly, the maximum sexual dimorphism in horn length is displayed by species with female horn length of about 32 cm, and subsequently decreases with increasing horn length (Fig. 4 in Tidière *et al*. 2018). These patterns are captured by the models fitted on a log-log scale for males and females (Tidière *et al.*, 2017). We are fully aware that the quadratic model we fitted (Tidière et al. 2017) is not suitable for describing allometry, but the statistical significance of the quadratic term allows us to conclude safely that the proportional increase in horn length relative to the proportional increase in mass (i.e. the allometry) changes when the species body mass increases. This suggests that the selective pressures that determine the optimal horn size for a given body mass in Bovids change when increasing species body mass.

**Conclusions**

The two analyses (Tidière *et al.*, 2017; Packard, 2018) reveal different conceptions of modelling biological processes. Packard’s approach, mostly based on statistical considerations, aims at fitting allometric data with any model as long as these models are on an arithmetic scale because it is the scale on which traits are measured. The interpretation of the model parameters is performed *a posteriori* and unexpected deviations (in this case the particularly high residuals for the males of the largest species) are interpreted as outliers that may be discarded. In the approach we advocate, the allometric model (Eq. 1) serves as a “null model” that describes the expected proportional increase of two traits sharing a common factor affecting their growth (Huxley 1932, Savageau 1979). This model provides a theoretical justification for the log-normal distribution of the error and the use of a proportional scale (i.e. a log-log scale) to analyse allometry. In this context, the allometric slope is a key parameter that can be interpreted as an elasticity between the two traits. Such a dimensionless number (*sensu* Charnov 1993) provides a metric that can be used in comparative analyses across the tree of life.

**Acknowledgements**

The authors are particularly grateful to G. H. Bolstad, T.F. Hansen, G.C. Packard, A. Nevill and an anonymous referee for providing valuable comments on an earlier version of this manuscript, and to Jarle Tufto for the discussion on allometry and other mathematical issues. C. Pélabon is supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 223257. Questioning methods, theories, or interpretations that are “classically accepted” is part of the scientific process. Therefore, we fully respect the strong debate stripped up by G.C. Packard on the methods used to analyse allometric relationships.

**References**

**Armbruster WS, Pélabon C, Bolstad GH, Hansen TF. 2014.** Integrated phenotypes: understanding trait covariation in plants and animals. *Phil. Trans. R. Soc.* B, **369**, 20130245.

**Armbruster WS, Pélabon C, Hansen TF, Bolstad GH. 2009**. Macroevolutionary patterns of pollination accuracy: a comparison of three genera. *New Phytologist*, **183**, 600-617.

**Armbruster WS, Schwaegerle KE. 1996.** Causes of covariation of phenotypic traits among populations. *Journal of Evolutionary Biology*, **9**, 261-276.

**Bales GS. 1996.** Heterochrony in brontothere horn evolution: allometric interpretations and the effect of life history scaling. *Paleobiology,* **22**, 481-495.

**Ballantyne F. 2013.** Evaluating model fit to determine if logarithmic transformations are necessary in allometry: a comment on the exchange between Packard (2009) and Kerkhoff and Enquist (2009). *Journal of Theoretical Biology*, 418-421.

**Bonduriansky R. 2007.** Sexual selection and allometry: A critical reappraisal of the evidence and ideas. *Evolution* **61**: 838-849.

**Cawley G, Janacek G. 2010**. On allometric equations for predicting body mass of dinosaurs. *Journal of Zoology*, **280**, 355-361.

**Charnov EL. 1993.** *Life history invariants: some explorations of symmetry in evolutionary ecology*. Oxford University Press, USA.

**De Kroon H, Plaisier A, Vangroenendael J, Caswell H. 1986.** Elasticity - the Relative Contribution of Demographic Parameters to Population-Growth Rate. *Ecology*, **67**, 1427-1431. 1.

**Eberhard WG, Huber BA, Rodriguez RL, Briceno RD, Salas I, Rodriquez V. 1998.** One size fits all? Relationships between the size and degree of variation in genitalia and other body parts in twenty species of insects and spiders. *Evolution*, **52**, 415-431.

**Gingerich PD. 2000.** Arithmetic or geometric normality of biological variation: an empirical test of theory. *Journal of Theoretical Biology* **204**: 201-221.

**Glazier DS. 2013.** Log-transformation is useful for examining proportional relationships in allometric scaling. *Journal of Theoretical Biology,* **334**, 200-203.

**Hereford J, Hansen TF, Houle D. 2004.** Comparing strengths of directional selection: How strong is strong? *Evolution*, **58**, 2133-2143.

**Houle D, Pélabon C, Wagner GP, Hansen TF. 2011.** Measurement and Meaning in Biology. *Quarterly Review of Biology,* **86**, 3-34.

**Huxley J. 1932.** *Problems of relative growth*. Johns Hopkins University Press, London.

**Kerkhoff AJ, Enquist BJ. 2009.** Multiplicative by nature: why logarithmic transformation is necessary in allometry. *Journal of Theoretical Biology*, **257**, 519-521.

**Lande R. 1979.** Quantitative genetic analysis of multivariate evolution, applied to brain - body size allometry. *Evolution*, **33**, 402-416.

**Lande R. 1985.** Genetic and evolutionary aspects of allometry. In: *Size and scaling in primate biology*. Springer, pp. 21-32.

**Lemaître J, Vanpé C, Plard F, Gaillard J M. 2014.** The allometry between secondary sexual traits and body size is nonlinear among cervids. *Biology Letters*, **10**, 20130869.

**Lemaître J, Vanpé C, Plard F, Pélabon C, Gaillard JM. 2015.** Response to Packard: make sure we do not throw out the biological baby with the statistical bath water when performing allometric analyses. *Biology Letters*, **11**, 20150144.

**Mascaro J, Litton CM, Hughes RF, Uowolo A, Schnitzer SA. 2013.** Is logarithmic transformation necessary in allometry? Ten, one-hundred, one-thousand-times yes. *Biological Journal of the Linnean Society*, **111**, 230-233.

**Packard GC. 2011.** Rotational distortion in conventional allometric analyses. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **159**, 392-400.

**Packard GC. 2011.** Unanticipated consequences of logarithmic transformation in bivariate allometry. *Journal of Comparative Physiology* B, **181**, 841-849.

**Packard GC. 2012.** "Is non-loglinear allometry a statistical artifact?" *Biological Journal of the Linnean Society* **107**: 764-773.

**Packard GC. 2015.** Allometric variation in the antlers of cervids: a comment on Lemaître et al. *Biology letters*,11**(6)**, 20140923.

**Packard GC. 2017**. Is complex allometry in field metabolic rates of mammals a statistical artifact? *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **203**, 322-327.

**Packard GC. 2018.** Evolutionary allometry of horn length in the mammalian family Bovidae reconciled by non-linear regression. *Biological Journal of Linnean Society*.

**Packard GC, Birchard GF. 2008.** Traditional allometric analysis fails to provide a valid predictive model for mammalian metabolic rates. *Journal of Experimental Biology*, **211**, 3581-3587.

**Pelabon C, Bolstad GH, Egset CK, Cheverud JM, Pavlicev M, Rosenqvist G. 2013.** On the Relationship between Ontogenetic and Static Allometry. *American Naturalist*, **181**, 195-212.

**Pelabon C, Firmat C, Bolstad GH, Voje KL, Houle D, Cassara J, Le Rouzic A, Hansen TF. 2014.** Evolution of morphological allometry. *Annals of the New York Academy of Science*, **1320**, 58-75.

**Savageau MA. 1979.** Growth of complex systems can be related to the properties of their underlying determinants. *Proceedings of the National Academy of Sciences,* **76**, 5413-5417.

**Schluter D, Grant PR. 1984.** Determinants of morphological patterns in communities of Darwin's finches. *American Naturalist*, **123**, 175-196.

**Riska, B. 1991.** Regression models in evolutionary allometry. *American Naturalist*, **138**, 283-299.

**Sartori, AF, Ball AD. 2009.** Morphology and post-larval development of the ligament of *Thracia phaseolina* (Bivalvia: Thraciidae), with a discussion of model choice in allometric studies. *Journal of Molluscan Studies*, **75**, 295-304.

**Smith RJ. 1980.** Rethinking allometry. *Journal of Theoretical Biology*, **87**, 97-111.

**Smith RJ. 1984.** Allometric scaling in comparative biology: problems of concept and method." *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, **246**, R152-R160.

**Tidière M, Lemaître JF, Pélabon C, Gimenez O, Gaillard JM. 2017.** Evolutionary allometry reveals a shift in selection pressure on male horn size. *Journal of Evolutionary Biology*. **30**: 1826-1835.

**Thompson D’AW.** **1942** *On growth and form*. Cambridge University Press.

**Van Tienderen PH. 1995.** Life cycle trade‐offs in matrix population models. *Ecology*, **76**, 2482-2489.

**Van Tienderen PH. 2000.** Elasticities and the link between demographic and evolutionary dynamics. *Ecology*, **81**, 666-679.

**Voje KL. 2016.** Scaling of morphological characters across trait type, sex, and environment: A meta-analysis of static allometries. *American Naturalist,* **187**: 89-98.

**Voje KL, Hansen TF, Egset CK, Bolstad GH, Pelabon C. 2014.** Allometric Constraints and the Evolution of Allometry. *Evolution*, **68**, 866-885.

**Xiao X, White EP, Hooten MB, Durham SL. 2011**. On the use of log-transformation vs. nonlinear regression for analyzing biological power laws. *Ecology*, **92**, 1887-1894.

Fig. 1 – Correlations between observed and predicted values for the horn length in male and female Bovids using either a linear models fitted on a log-log scale without phylogenetic correction (black dots, Tidière *et al.,* 2017), or a non-linear model on an arithmetic scale (open dots, Packard, 2018). Because Packard (2018) modelled log-normally distributed residuals, we compare the fit of the two models by comparing predicted and observed values on a log-log scale. R-square values are, for the model on a log-log scale males: r2 = 0.75; females r2 = 0.67, and for the model on the arithmetic scale males: r2 = 0.71; females r2 = 0.60. The solid line represents the 1:1 relationship.

Fig. 2 – Distribution of the residuals of the two different models. Residuals of the model fitted on a log-log scale (without correction for phylogenetic relationship among species) for males and females on the upper row, and residuals of the models fitted on the arithmetic scale with log-normal error for males and females (residuals from the model fitted on the arithmetic scale, log-transformed) on the lower row.