Paleobiology

CAMBRIDGE UNIVERSITY PRESS

Inferring flight parameters of Mesozoic avians through multivariate analyses of forelimb elements in their living relatives

Journal:	*Paleobiology*
Manuscript ID	PAB-OR-2015-0058.R3
Manuscript Type:	Article
Date Submitted by the Author:	17-Aug-2016
Complete List of Authors:	Serrano, Francisco; Natural History Museum of Los Angeles County, Dinosaur Institute; Fundación Sierra Elvira, Paleontología; Universidad de Málaga, Ecología y Geología Palmqvist, Paul; Universidad de Málaga, Ecología y Geología Chiappe, Luis; Natural History Museum of Los Ángeles County, Dinosaur Institute Sanz, José; Universidad Autónoma de Madrid, Biología
Geographic Location:	
Taxonomy:	birds
Analysis:	morphometry, multivariate statistics
Geologic Age:	Mesozoic
Topic:	prediction bias, flight inference

Inferring flight parameters of Mesozoic avians through multivariate analyses of forelimb elements in their living relatives

Francisco J. Serrano, Paul Palmqvist, Luis M. Chiappe, and José L. Sanz

RRH: FLIGHT PARAMETERS ESTIMATION IN MESOZOIC BIRDS

LRH: FRANCISCO J. SERRANO ET AL.

Abstract.—Our knowledge of the diversity, ecology, and phylogeny of Mesozoic birds has increased significantly during recent decades, yet our understanding of their flight competence remains poor. Wing loading (*WL*) and aspect ratio (*AR*) are two aerodynamically relevant parameters, as they relate to energy costs of aerial locomotion and flight maneuverability. They can be calculated in living birds (i.e., Neornithes) from body mass (*BM*), wingspan (*B*) and lift surface (S_L). However, the estimates for extinct birds can be subject to biases from statistical issues, phylogeny, locomotor adaptations, and diagenetic compaction. Here we develop a sequential approach for generating reliable multivariate models that allow estimating measurements necessary to determine *WL* and *AR* in the main clades of non-neornithine Mesozoic birds. The strength of our predictions is supported by the use of those variables that show similar scaling patterns in modern and stem taxa (i.e., non-neornithine birds), and the similarity of our predictions with measurements obtained from fossils preserving wing outlines. In addition, although our *WL* and *AR* values are based on estimates (*BM, B,* and S_L) that have an associated error, there is no cumulative error in their calculation, and both parameters show low prediction errors. Therefore we present the first taxonomically broad, error-calibrated estimation of these two important aerodynamic parameters in non-neornithine birds. Such estimates show that the *WL* and *AR* of the non-neornithine birds here analyzed fall within the range of variation of modern birds (i.e., Neornithes). Our results indicate that most modern flight modes (e.g., continuous flapping, flap and gliding, flap and bounding, thermal soaring) were possible for the wide range of non-neornithine avian taxa; we found no evidence for the presence of dynamic soaring among these early birds.

*Francisco J. Serrano. Departamento de Ecología y Geología, Facultad de Ciencias, Universidad de Málaga. Campus Universitario de Teatinos s/n, 29071 Málaga, Spain. *Present addresses: The Dinosaur Institute, Natural History Museum of Los Angeles County, 900 Exposition Boulevard, Los Angeles, CA, U.S.A. Fundación Sierra Elvira, Avenida de Andalucía 139, Atarfe, 18230 Granada, Spain. E-mail: fjsa@uma.es*

Paul Palmqvist. Departamento de Ecología y Geología, Facultad de Ciencias, Universidad de Málaga. Campus Universitario de Teatinos s/n, 29071 Málaga, Spain. E-mail: ppb@uma.es

Luis M. Chiappe. The Dinosaur Institute, Natural History Museum of Los Angeles County, 900 Exposition Boulevard, Los Angeles, CA, U.S.A. E-mail: lchiappe@nhm.org

José L. Sanz. Unidad de Paleontología, Departamento de Biología, Facultad de Ciencias, Universidad Autónoma de Madrid, Calle Darwin 2, Cantoblanco, 28049 Madrid, Spain. E-mail: dinoproyecto@gmail.com

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.63k11

Introduction

Since the 1980s, a steady flow of discoveries of Mesozoic fossils worldwide has transformed our understanding of the earliest phases of the avian evolution (e.g. Sanz et al. 1988; Chiappe and Calvo 1994; Sanz et al. 1995; 1996; Hou 1997; Clarke and Norell 2002; Zhou and Zhang 2002; Zhou et al. 2005; Chiappe et al. 2007; Zhang et al. 2008; Zhou et al. 2009; O´Connor et al. 2010; Brusatte et al. 2015). Numerous extinct lineages have been recorded—particularly from the Early Cretaceous of northeastern China (Chiappe and Meng 2016)—and their evolutionary

relationships, and ecological differentiation have been deciphered (e.g. Chiappe et al. 1999; Chiappe and Witmer 2002; You et al. 2006; Bell and Chiappe 2011; O´Connor et al. 2011; Chiappe et al. 2014). Nonetheless, important questions about the aerodynamic competence of these early relatives of modern birds have remained poorly answered. This is largely due to our inability to characterize, in these extinct birds, a series of key variables known to influence the flight of their living counterparts. Wings are the anatomical structures that allow birds to travel through the air (their aerodynamic design generates lift force necessary to support the animal's weight, allowing the bird to focus its muscles' work on flapping the wings to overcome aerodynamic drag; Pennycuick 2008). From an aerodynamic point of view, bird wings can be characterized by two basic measurements, wing span (B) and wing area (S_L), defined as the distance from one wing tip to the other and the area that supports the bird´s airborne flight, respectively.

To study the aerial behavior of the stem birds, some authors have obtained values for B and S_L from outline reconstructions based on well-preserved specimens such as *Archaeopteryx* (Yalden 1971; 1984; Elzanowski 2002; Longrich 2006; Wellnhofer 2008) and *Hongshanornis* (Chiappe et al. 2014). However, these types of studies are scarce and can only be confidently applied to nearly complete avian fossils with well-preserved flight feathers. An alternative to body reconstruction is to derive allometric equations adjusted in extant birds, to estimate B and S_L in extinct birds which preserve the variable(s) included in the equations. Scaling equations of B and S_L on body mass (BM) have been obtained from modern datasets (e.g. Greenwalt 1975; Viscor and Fuster 1987; Pennycuick 2008; Tennekes 2009), and in some cases they have been applied to fossil stem birds (e.g. Sanz et al. 2002). However, given that the BM of any extinct bird is also an estimated value, which can be subject to different errors and biases (see bellow),

the allometric equations for B and S_L should be derived from morphological traits that are preserved in the fossils studied (e.g., measurements of their bones and feathers).

As recently discussed by Serrano et al. (2015), any estimated parameter derived from regression functions can be affected by sources of error from statistical, phylogenetic, ecological, and taphonomic nature. Namely, these biases may arise from limitations imposed by: (1) sampling availability and differences in taxonomic diversity (see Mendoza et al. 2006; De Esteban-Trivigno et al. 2008); (2) multicollinearity (see Mitchell-Olds and Shaw 1987; Bowerman and O'Connell 1990; Quinn and Keough 2001); (3) statistical non-independence of species given their historical relationships (see Felsenstein 1985; Harvey and Pagel 1991); (4) uncertainty on the adequacy of assuming similar scaling patterns for crown and stem taxa (see Carrano 2001; Smith 2002; Packard et al. 2009; Campione and Evans 2012; Field et al. 2013); (5) effects of the functional adaptations on the morphology of wing bones (see Cubo and Casinos 1998; De Margerie et al. 2005; Habib and Ruff 2008; Hinic-Frlog and Motani 2010; Simmons 2010; Bell and Chiappe 2011; Simmons et al. 2011); and (6) effects of diagenetic compaction on the osteological dimensions of fossil birds. In the case of the *BM* estimates, these potential biases were avoided or reduced by using a step by step methodology for obtaining specific multiple regression models (MR) for predicting *BM* values in each Mesozoic stem taxa (Serrano et al., 2015).

In the case of wingspan values, other factors can influence the predictions obtained for extinct birds. The B value of a given bird depends on the length of the anatomical elements of its wing (i.e. humerus, ulna, carpometacarpus, major digit, and longest primary feather), the extended angles of the elbow and wrist, and the width of the body region between the wings. The absence of information on any of these factors results in increased uncertainty in the prediction

of *B* for avian stem fossils. Several studies have shown that the length proportions between wing skeletal elements of basal birds are inside the range of modern birds (e.g. Middleton and Gatesy 2000; Dyke and Nudds 2009; Benson and Choiniere 2013; Nudds et al. 2013), and hence these variables can be used for estimating their *B* values. Wang et al. (2011), however, claimed that the length of longest primary feather of some stem groups (e.g., Archaeopterygidae and Enantiornithes) was significantly shorter than those of modern birds. Yet, the scaling relationships between the length of the longest primary and the length of the skeletal wing have been little explored. On the other hand, the probability that an ancient bird fossilized with the wing completely extended is very low, which implies that elbow and wrist angles of this posture are unknown for stem birds. However, given that there are neognaths with very close angles (e.g. swifts) and others that show very open angles (e.g. many passerines) (see Videler 2006), it is reasonable to consider that the values of these angles for stem birds were inside the range of modern birds, as implicitly assumed in reconstructions of *B* for Mesozoic birds (e.g. Yalden 1971; 1984; Elzanowski 2002; Longrich 2006; Wellnhofer 2008; Chiappe et al. 2014). Finally, the length of the corporal region can be approximated by the width of the furcula (i.e. the distance between the tips of both clavicles), which scales isometrically with *B* in neognaths (Nudds and Rayner 2006). However, this scaling pattern is unexplored for most basal birds, which means that it is unknown how this variable could affect the prediction of *B*.

Wing area (S_L) depends on the area of both wings and also includes the part of the body situated between the wings. S_L correlates with *B* and with *BM* because it is the surface generating the lift that supports the weight of the airborne bird (Greenwalt 1975; Viscor and Fuster 1987; Álvarez et al. 2001; Sanz et al. 2002; Meseguer and Sanz-Andrés 2007; Pennycuick 2008). Accordingly, the S_L values of extinct birds can be accurately estimated from those osteological

measurements of the wing and dimensions of preserved flight feathers that show scaling patterns comparable to those of modern birds, in a similar way to the predictions of *BM* performed by Serrano et al. (2015).

Prediction of *B*, S_L, and *BM* values allows the calculation of two parameters crucial for the aerodynamics of any flying bird: wing loading (*WL*) and aspect ratio (*AR*). These two parameters provide an approximation to the mode in which a bird can fly more effectively. *WL* (weight supported by a unit surface of the wings) is directly proportional to the speed in the non-flapping phases of flight, and relates inversely to maneuverability (Norberg 2002; Longrich 2006; Alerstam et al. 2007; Pennycuick 2008; Tennekes 2009); *AR* (proxy of wing shape that indicates the relationship between wing length and width) is directly related to aerodynamic advantage for long flights, because high *AR* values increase the lift coefficient and reduce the mechanical power necessary to fly (by decreasing both the induced drag and the profile drag; Meseguer and Sanz 2007; Pennycuick 2008). For example, modern soaring birds use the wind to gain altitude, which is facilitated by their low-to-medium *WL* values. In addition, dynamic soarers (e.g., albatrosses) show high *AR* values, which allow them to make use of the velocity gradient of the horizontal airflow over sea, while thermal soarers (e.g., vultures) show moderate to low *AR* values for taking advantage of vertical ascending currents. In contrast, continous flapping birds usually show high *WL* values. Among these birds, the species that perform short flights (e.g., landfowls and rails) show lower *AR* values than those that develop long flights (e.g., waterfowls and diving birds) (see Viscor and Fuster 1987; Rayner 1988; Norberg 2002).

Knowledge of the flight performance of basal birds has mainly been obtained from their anatomical descriptions and from comparisons with extant neognaths (e.g. Sanz et al. 1988; 1996; Zhang and Zhou 2000; Zhou and Zhang 2002; Chiappe and Witmer 2002, Clarke and

Norell 2002; Senter 2006). The presence or absence of homologous structures that correlate with flight functions has been used to hypothesize aspects of flight modes of early birds. For example, the presence of an alula in the Early Cretaceous enantiornithine *Eoalulavis hoyasi* was interpreted as evidence of increased flight control at low speeds (Sanz et al. 1996). Several comparative studies on limb bone proportions have also contributed to a better understanding of early avian flight (e.g. Middleton and Gatesy 2000; Dyke and Nudds 2009; Bell and Chiappe 2011; Wang et al. 2011; 2012; Chan et al. 2012; Close and Rayfield 2012; Nudds et al. 2013; Benson and Choiniere 2013). However, assessments of the flight performance of stem birds based on estimates of their flight parameters are scarce. With a few recent exceptions (e.g., Chiappe et al. 2014; Evangelista et al. 2014), most aerodynamic analyses have been focused on *Archaeopteryx* (e.g. Yalden 1971; Burgers and Chiappe 1999; Chatterjee and Templin 2003; Longrich 2006; Meseguer et al. 2008; 2012).

The main objective of this study is obtaining reliable estimates of basic aerodynamic variables in Mesozoic, non-neornithine birds. Specifically, we construct multiple regression models for calculating the values (and errors) of wingspan, wing area, and aspect ratio in different lineages of early birds, and we use recent body mass estimates (Serrano et al., 2015) to calculate their wing loading values. All these estimates are then used to discuss the flight performance and flight modes of some of the earliest known birds.

Material and Methods

Data Set

A total of 41 fossils of presumed flying Mesozoic stem taxa were studied in order to estimate their B and S_L values (Supplementary Table 1). These specimens represent a broad

sample of the avian clade during the Late Jurassic and Cretaceous. The predictive models are based on a data set of modern volant birds comprised of 188 individuals from 83 flying species of Neognathae distributed among 49 families and 21 orders (Supplementary Table 2), covering a broad range of sizes (from mute swan to little woodswallow) and including species from all flight types (except hummingbirds due to their unique hovering abilities). Both datasets of stem and modern birds were characterized in Serrano et al. (2015). A total of 15 linear dimensions of the forelimb bones were measured on each specimen, as well as the width of the furcula. Corresponding data of body mass (*BM*) and wing span (*B*) were taken by Museum personnel (Burke Museum of Natural History and Culture, Seattle) while the specimens were still alive. In addition, four other variables were measured for each specimen from pictures of wings using the software tpsDIG 2.17 (available at: http://life.bio.sunysb.edu/morph/): chord at wrist level (*c*; which is a proxy of the mean chord explained in Pennycuick 2008), longest primary feather (L_{prim}), length of a wing (*Lw*), and surface of a wing (*Sw*). According to Pennycuick (2008) and Álvarez et al. (2001), total lift surface (S_L) was calculated from eq. (1), where the area of the body region situated between the wings (S_b) was obtained from eq. (2):

$$(1)\ S_L = 2\ Sw + S_b$$

$$(2)\ S_b = (B - 2\ Lw)\ c$$

The measurements used as variables in this study are defined in Table 1 and depicted in the Figure 1.

Data Proccesing and Predictive Models

Data from modern birds were used to fit linear regression functions using the ordinary least squares method (OLS), with *B* and S_L as the response variable (*y*), and the measurements

193 taken in forelimb skeleton and wing's feathers as predictor variables (x). Given that a relationship
194 of dependence for both B and S_L on the x variables can be assumed, and given that prediction is
195 the main objective of the regression analyses, the OLS method is the most adequate for deriving
196 the equations (Gingerich et al. 1982; Sokal and Rohlf 1986; MacLeod 2004; Smith 2009;
197 Campione and Evans 2012). Furthermore, given that raw data for B, S_L, and the skeletal
198 measurements show positively skewed distributions, all data were $\log_{10}$-transformed in order to
199 approximate a normal distribution and minimize the level of heteroscedasticity (Quinn and
200 Keough 2001; Cawley and Janacek 2010; however, see Packard et al. 2009). However, the
201 predicted log-values obtained from least square regression could bias the results when they are
202 retransformed to arithmetic units (Smith 1993, Hayes and Shonkwiler 2006, Clifford et al. 2013).
203 In order to avoid this potential bias, each retransformed value was multiplied by a correction
204 factor known as the ratio estimator (Snowdon 1991). In addition, plots of the residuals against
205 predicted values were also examined to detect outliers.

206 Three types of regression were derived for estimating B and S_L in order to analyze their
207 predictive reliability: 1) single regressions using the best correlated predictor variable; 2)
208 principal components regressions, selecting those principal components significantly correlated
209 to response variables (see details in Rawlings et al., 1998 and Serrano et al. 2015); and 3)
210 multiple regressions (MR) using all predictor variables (i.e. full model), as well as MR equations
211 with predictors selected by automatic stepwise procedure (selection of x-variables showing the
212 highest correlation with the y-variable in such a way that each step refits the model by
213 introducing or dropping any predictor variable based on the contribution of its partial regression
214 slope; see details in Mendoza et al. 2006). Statistical analyses were carried out using SPSS
215 statistical package version 20 (http://www.ibm.com/software/es/analytics/spss/).

Reliability of the Predictive Models

The predictive strength of the models is based on their accuracy (i.e. proximity between estimated and real values) and generalizability (i.e. predictive ability of the models on new data). The accuracy of the functions was evaluated using the following values: (1) the coefficient of determination adjusted for the number of predictor variables incorporated in the model ($R^2_{adj} = 1 - [1 - R^2][N - 1] / [N - p]$, where N is sample size and p is the number of predictor variables, which number in the MR models was set according to the general rule $N/p > 10$; see Darlington 1990), and (2) the percent prediction error (%PE = [observed – predicted]*100/predicted), which measures the degree of overestimation (%PE < 0) or underestimation (%PE > 0) for each particular observation. The sample mean of the absolute values of %PE, noted as |%MPE|, is the average percent deviation between predicted and observed values. This allowed to calculate the average prediction intervals for B and S_L as predicted value ± |%MPE| (Smith 1984; Mendoza et al. 2006; Campione and Evans 2012; Field et al. 2013).

Generalizability of the models was evaluated using two validation tests. A first validation was performed at the species level. Specimens of 25% of the species in the data set were randomly selected (the number of specimens corresponds to 20% of the whole sample) to conform a subsample for cross-validation (see detailed procedure in Hurvich and Tsai 1990, MacNally 2000, and Mendoza et al. 2006). The functions derived from the remaining 80% of specimens were used for obtaining the values (B and S_L) of the specimens of the validating subsample, and |%MPE| value was calculated for them (i.e. $|\%MPE|_{VS\text{-}25\%}$). Such a random subsample is appropriate for testing generalizability in models aimed at estimating extinct crown species. However, stem taxa are not a random assortment of crown species as they fall outside

239 the phylogenetic boundaries of the crown group. A second validation was performed to address
240 this issue. Models derived from the neoavian specimens were used to predict B and S_L values in
241 all specimens of Galloanserae (the sister-clade of Neoaves). The |%MPE| values calculated for
242 Galloanserae (i.e. $|\%MPE|_{VS\text{-}GA}$) were used to test the stemward predictive strength of the
243 models. In both approaches, the specimens of the validating subsamples were incorporated into
244 the regression functions after cross-validation, in order to include all measured individuals.
245
246 Potential Biases in the Regression Functions
247 *Taxon Weighting*.—The crown families used in this study are represented by different
248 numbers of species, which in turn are represented by different numbers of measured specimens.
249 Such uneven sampling is known to bias the regressions when the scores of over-represented and
250 under-represented families/species are not randomly distributed around the regression line. For
251 this reason, we adjusted the functions weighting the richness of specimens within a species (each
252 specimen contributed to the regressions as 1/N specimens per species) and species within
253 families (1/N specimens per family). Their %MPE values were compared with those from
254 unweighted regressions.
255 *Multicollinearity*.—Redundant information between x-variables included in a given MR
256 model (i.e. multicollinearity) is known to have negative effects on the prediction (Mitchell-Olds
257 & Shaw 1987; Bowerman & O'Connell 1990). For this reason, the predictive strength of the MR
258 equations was compared with non-collinear models, such as single regression and principal
259 component regression. Multicollinearity was measured by the variance inflation factor (VIF),
260 calculated for each variable x_j as $[1/(1 - R^2)]$ from the OLS regression of x_j against the remaining
261 x-variables (VIF values >10 suggest strong collinearity effects; see Kleinbaum et al., 1997).

Phylogenetic Non-Indepedence.—Any given sample of species does not meet the assumption of statistical independence due to their historical relationships, and hence phylogenetically corrected methods should be used (Felsenstein 1985, Harvey and Pagel 1991). A calibrated cladogram of the modern birds included in the dataset was constructed using MESQUITE software (Maddison & Maddison, 2011). Tree topology and branch length of neognath species were obtained from Prum et al. (2015). Placement in the cladogram of those species not included in Prum et al. (2015) followed the family-level topology of this study; for clades unresolved by Prum´s phylogenetic hypothesis, the internal relationships followed specific references (see details in Supplementary Figure 1). Phylogenetically independent contrasts (PICs) for each variable (Felsenstein 1985) were obtained using the PDA:PDTREE package (Midford, Garland & Maddison, 2005). Standardized PICs versus their standard deviations were used to check if branch lengths adequately fit the tip data (see Diaz-Uriarte and Garland 1998). In addition, given that the use of variables with a poor phylogenetic signal may be incorrect in statistical terms (Revell 2010), the phylogenetic signal of the predictors was tested as follows: the squared length of each trait was statistically tested against 10^4 randomly generated trees in which terminal taxa were permuted while holding tree topology and branch lengths constant. Specifically, signal is significant if the number of steps between the basal node and the terminal branches of the tree analyzed are lower than in the random trees (see details in Laurin 2004).

Least squares single regressions were derived for the PICs of each predictor variables against B and S_L to deal with the non-independence issue. PIC regressions were compared to OLS models for species means (uncorrected models) in order to check if the scaling patterns within a phylogenetic context were not significantly different from a context in which the birds

of the sample were independent (from a statistical point of view). PIC bivariate models are regressions through the origin (i.e. intercept = 0). As residuals of this type of regression usually have a nonzero mean, straightforward comparisons with OLS may not be appropiate (Eisenhauer 2003). For this reason, the difference between slopes from PICs and OLS was tested using the overlap of the 95% confidence intervals, obtained by bootstrapping (2000 replicates).

Selection of Variables: Scaling Patterns in Crown and Stem Groups.—In order to derive predictive models for estimating the wingspan (*B*) of stem birds from a modern bird database, it is necessary to determine if the portion of the body between the wings of stem birds scales similarly to their wingspan. This was tested following the calculations of Nudds and Rayner (2006) and performing a reduced major axis (RMA) analysis between the width of furcula (*WF*) and a proxy of *B* that can be measured in several fossils, Lw^* ($HL + UL + CmL + LDII + L_{prim}$). This analysis was performed on a neognath's dataset of 326 individuals from 140 species [the dataset of this study and additional data obtained from Viscor and Fuster (1987) and Bruderer et al. 2010] and 10 extinct Mesozoic species. The confidence interval for the RMA slope was obtained through bootstrapping (2000 replicates). Subsequently, we also checked if the variable L_{prim} could be used for estimating *B* in stem birds, given that Wang et al. (2011) argued that some stem taxa had L_{prim} values outside the neognath's range, hence calling into question the predictive power of this variable. This was done by plotting the values of the ratio between L_{prim} and the length of the skeletal wing (i.e. $Lw_{sk} = HL + UL + CmL + LDII$) for each stem taxa and for a dataset of neognaths composed of 256 individuals from 131 species and 67 families [additional data obtained from Viscor and Fuster (1987) and Bruderer et al. (2010)]. Finally, the length of the wing's skeletal elements (i.e. *HL*, *UL,* and *CmL*) was included in the MR functions derived for obtaining *B* values.

The MR models used for predicting the wing area –S_L– of stem birds were derived from all forelimb variables, including c and L_{prim} (see Table 1). Those variables that scale similarly in crown and stem groups were selected following the procedure of Serrano et al. (2015). This approach minimizes the potential biases that could emerge from differences in morphological scaling patterns between living and extinct taxa, and the effects of diagenetic compaction on fossil bones. The behavior of each x-variable with respect to the remaining x-variables was analyzed, and those that showed values in the fossil taxa outside the trends of Neognathae were discarded. The linear measurements were selected by plotting each measurement against the S_L predicted from a combination of the remaining x-variables. Subsequently, a regression line was fitted for modern birds in each plot, and those variables that scored within the 95% confidence intervals were selected for each stem group, allowing us to obtain particular subsets of predictors for them. Finally, the stepwise method was used in each subset for deriving the best multivariate functions. This procedure also helped to identify those skeletal measurements heavily affected by diagenetic compaction, which allows to detect individuals affected by taphonomy that would behave as an outlier.

Post-hoc Tests for Effects of Adaptive Changes.—As in Serrano et al. (2015), *post-hoc* tests were performed over the MR functions obtained for evaluating the influence of locomotion adaptations on these equations and they were also compared to the best SR models. The effects of flight mode on the prediction error (%PE) of B and S_L were statistically tested by means of one-way ANOVAs. In order to do so, modern birds were classified into four categories according to their predominant behavior during aerial locomotion, following Pennycuick (2008) and Bruderer et al. (2010): (1) continuous flapping; (2) flap and gliding; (3) flap and bounding; and (4) soaring. Given that these groupings differ in sample size, they could violate the

ANOVA's assumption of homocedasticity, which was tested with the Levene's test. In those cases where homocedasticity was rejected, the statistical differences between the locomotion groups were analyzed with the Welch's test, a non-parametric procedure.

Calculation and Error Calibration of the Flight Parameters

As mentioned above, wing loading (*WL*) refers to the body weight supported by unit lift surface (i.e. $BM\ g/S_L$, with *g* representing the acceleration of gravity). Then, using the estimates of *BM* obtained in Serrano et al. (2015) and the S_L values predicted from the MR functions developed in this study, we calculated the tentative *WL* value for each analyzed stem bird. Given that these *WL* estimates depend on two values (*BM* and S_L) that have associated errors of estimation, we expect a greater margin of error for predicting *WL*. As we know the actual values of *WL* for specimens included in the modern dataset, mean prediction errors could be estimated for this compound variable (i.e. %MPE, $MPE_{VS\text{-}25}$ and $MPE_{VS\text{-}GA}$). This calculation was performed for each combination of predictive MR functions for *BM* and S_L. The aspect ratio (*AR*) was obtained as the squared wingspan divided by the total wing surface (i.e. $AR = B^2/S_L$). As *AR* values also depend on two estimated values (i.e. *B* and S_L) with an associated error, mean prediction errors (i.e. %MPE, $MPE_{VS\text{-}25}$ and $MPE_{VS\text{-}GA}$) were also calculated for all combinations of predictive MR functions for *B* and S_L. Such procedure allows calibrating the error of the compound variables (*AR* and *WL*) from the simple variables (*B*, S_L, and *BM*). Furthermore, the obtaining of the |%MPE| values allowed us to delimit the confidence intervals for the estimates of *WL* and *AR* obtained for each basal bird.

Results

Predictive Ability from Different Regression Approaches (issues from Multicollinearity, Taxon-Weighting and Automatic Stepwise)

The best multiple regressions derived for B and S_L from the modern dataset show a greater degree of multicolinearity when compared to both (1) single regressions of the best correlated predictors (i.e. length of carpometacarpus for B, and width of distal epiphysis of humerus for S_L) and (2) regressions of the principal components (Table 2). In spite of this, multiple regressions had a predictive strength that is notably higher than the corresponding single regressions (higher values of R^2_{adj} and lower of MPEs). Multiple regressions also had slightly higher R^2_{adj} and predictive power on the whole dataset (i.e. lower |%MPE|) than principal components regressions (Table 2). The errors in predicting B and S_L values stemward were similar in multiple regression and principal components regression (Table 2).

The values of R^2_{adj}, and MPEs values of the unweighted multiple regression models were similar or better than those values from taxonomically weigthed MR models (Table 3).

As expected, the reduction of variables included in a model generated by automatic stepwise resulted in a better adjusted coefficient of determination (R^2_{adj}) when compared to the full model (Table 3). S_L prediction errors on the whole dataset (i.e. |%MPE|) were slightly lower in the models that included the full set of x-variables. However, we found that the models generated by automatic stepwise increased the predictive power (at least by 1.7%) on new specimens placed stemward from Neoaves ($|\%MPE|_{VS\text{-}GA}$ in Table 3).

Phylogenetic Non-Independence Issue

Diagnostics for suitability of PICs relative to the branch lengths of the cladogram were achieved in all variables, as shown by the lack of relationship between PICs and their

corresponding standard deviations ($P > 0.05$). The variables B, S_L, and all forelimb measurements showed a highly significant phylogenetic signal in the tree of Neognathae (permutation tests: $P < 0.001$ in all cases). This indicates that the use of phylogenetically corrected models, like PICs, derived from this data is statistically appropriate (Revell 2010). Least squares linear regressions with B as response variable were highly significant in both regression approaches, PICs and OLS on the mean species values (permutation test: $P < 0.0001$), showing an overlap in the corresponding confidence intervals for the slopes (Table 4). In all cases, phylogenetically corrected scaling relationships showed lower R^2 values than those adjusted from uncorrected data. Regressions for S_L on the unstandardized SR functions and PICs were also highly significant in all cases (permutation test: $P < 0.0001$), overlapping the confidence intervals for the slopes (Table 5). In addition, R^2 values of phylogenetically corrected functions were lower than R^2 of the OLS regressions of species means, except in the cases of *HL*, *RL,* and *CmL*.

The overlapping of the confidence intervals of the slopes observed indicates that the morphological patterns depicted by the species sampled are not strongly driven by phylogeny; the variables can be considered statistically to be 'taxon-free'. Therefore, non-independence issue does not affect the phylogenetically uncorrected models. In addition, these models also showed higher predictive power.

Predictive Models for Wing Span

The relationship between the interclavicular distance (i.e. *WF*) and the proxy for wingspan (i.e. *Lw**) in the sample of neognaths shows a slightly positive allometry (RMA slope = 1.14; 95% confidence interval = 1.10-1.18). As the slope of the stem birds analyzed (RMA

slope = 1.17; 95% confidence interval = 0.72-1.49) is included in the confidence interval of modern birds, a similar scaling pattern can be assumed (Fig. 2). Then, the prediction of *B* values for non-neornithine birds is not affected by any possible differences in the interclavicular distance (i.e., the corporal region situated between the wings) between these birds and their living counterparts.

For all analyzed stem birds, the length of the longest primary feather (L_{prim}) with respect to the sum of the lengths of the skeletal elements of the forelimb (Lw_{sk}) is included in the range of variation of neognaths (Fig. 3). This suggests that the variable L_{prim} has a potential use for generating models based on modern birds, therfore allowing the prediction of *B* in stem birds.

Three MR models for estimating *B* in extinct birds were derived (Table 6), showing very low prediction errors (|%MPE|: 2.6 to 7.4%; and $|\%MPE|_{VS\text{-}GA}$: 3.3 to 4.7%). The application of these models depends on the anatomical portions preserved in the fossils. The most complete model is MR-B1, because it includes the length of each wing element used to calculate wing span except *LDII*. The function MR-B2 was applied to those fossils that do not preserve the wing feathers (i.e., the measurement of L_{prim} was not available in them), which increases the uncertainty of the estimation. Finally, the model MR-B3 is useful to make predictions in the case of those fossils of basal birds that are less complete and cannot provide a *CmL* value; this has the poorest predictive power of the three models.

Predictive Models for Lift Surface

After selection for the predictor variables (Fig. 4, Table 7), several of them were excluded for deriving the specific functions for each stem taxa (Table 8). This results in 11 MR functions for predicting S_L values (i.e. more than one for each of the six fossil groups; Table 9). The

application of a given equation to a given fossil depends on which variables it preserves. The number of predictor variables included in the models ranged from two in the equation MR-S_LArch3 to seven in both MR-S_LArch1 and MR-S_LEnan1. Their predictive error |%MPE| was under 15% in all cases. Stemward validation suggests a good generalizability for the models selected, with 10 MR models having $|\%MPE|_{VS\text{-}GA}$ values ranging from 5.1 to 13.8%, and one model doing notably worse than the others (23.3% in MR-S_LArch3). The MR equations based on the highest skeletal completeness derived for Archaeopterygidae (MR-S_LArch1), Enantiornithes (MR-S_LEnan1), Sapeornithidae, and stem Ornithuromorpha (MR-S_LSape1 / MR-S_LOrph1) have the highest predictive strength (|%MPE| and $|\%MPE|_{VS\text{-}GA} \approx$ 5-6%). In contrast, those functions based on less variables for Archaeopterygidae (MR-S_LArch3), Enantiornithes (MR-S_LEnan4), and Sapeornithidae (MR-S_LSape2) were worse predictors (|%MPE| ≈ 11-14%; and $|\%MPE|_{VS\text{-}GA}$ 11.7-23.3%).

Prediction Errors Derived from Adaptive Differences

The *post hoc* analyses performed on the errors associated with the prediction of *B* values showed that the most predictive multivariate equation (i.e. MR-B1) was not affected by the flight mode of modern birds. Specifically, the ANOVA provided statistically non-significant results for the differences among groups of aerial locomotion (Table 10A). These differences, however, are significant in the other two MR equations (i.e. MR-B2 and MR-B3) as well as in the most predictive univariate function (i.e. SR from *UL*), which means that their *B* predictions are influenced by flight modality. However, the low mean %PE values and standard deviations obtained for each aerial group indicate that MR-B2 and MR-B3 minimize the influence of flight modality more than the best SR equation.

Table 10B shows that six of the 11 MR models derived for predicting S_L are independent of the type of aerial locomotion, as their corresponding ANOVA's are non-significant. However, the best SR equation (i.e. *deHW*) and the five MR functions result in S_L estimates that are affected by flight mode. As in the previous case, the influence of aerial locomotion is minimized in the MR functions with respect to the SR equations, as shown by the lower mean %PE values and standard deviations.

Calibrating the Aerodynamic Parameters: WL and AR

The *WL* values calculated from the estimates of *BM* and S_L have a range of prediction errors between 15.2% and 23.1% (Table 11A). Taking into account the |%MPE| range from the MR models derived for predicting *BM* (16.5% - 20.2%) and S_L (5.8% - 14.2%), the error accumulated for obtaining *WL* estimates can be considered quite low and depends mainly on the prediction error associated with the estimation of *BM*. However, it should be noted that two combinations of regression models (MR-Arch to estimate *BM* with MR-SLArch3 to estimate S_L; and MR-Orph to body mass with MR-SL Orph3 to S_L) gave values higher than 30% in the stemward validation.

On the other hand, the range of prediction errors for the *AR* values calculated from S_L and *B* estimates fluctuates between 7.0% and 17.8% (Table 11B). Given that the |%MPE| ranges associated with the prediction of *B* values (2.6% - 7.4%) and S_L values (5.8% - 14.2%), the cumulative error of the *AR* estimates can also be considered very low and, in this case, it basically depends on the prediction error of S_L. As in the previous case, combinations including MR-SL Arch3 to estimate SL resulted in the highest values for the stemward validation.

B, S_L, WL, and *AR* Values in Mesozoic Stem Taxa

The MR functions obtained (Tables 6 and 9) were used to estimate *B* and S_L values in the fossil specimens studied. These values and those obtained for *BM* (Serrano et al. 2015: Table 8) were used for calculating the *WL* and *AR* estimates for extinct birds (Table 12). The specimens of Archaeopterygidae show a range in *WL* values between 0.572 and 0.879 g/cm^2, the values obtained for the Berlin and Solnhofen specimens of *Archaeopteryx*, respectively. The *AR* values obtained for these birds range from 5.0 to 7.9 (Solnhofen and London specimens, respectively). In the case of Jeholornithidae, the largest specimen sampled (V13274) is also the most loaded, and its *AR* value is lower than in the other sampled specimen (V13353), which is considered a juvenile (Zhou and Zhang 2003). Similarly, the smallest specimen of *Sapeornis*, also considered a juvenile (DNHM 3078; Gao et al. 2012), has the lowest values for *WL* (0.376 g/cm^2) and *AR* (6.4). Larger specimens of *Sapeornis* show *WL* and *AR* ranging between 0.441 - 0.550 g/cm^2 and 7.1 - 7.8, respectively. In contrast, the smallest (and oldest) sampled specimen of Confuciusornithidae (i.e., *Eoconfuciusornis*) presents *WL* and *AR* values higher than individuals of *Confuciusornis*. The confuciusornithid specimens analyzed show *WL* and *AR* ranginig from 0.636 to 0.893 g/cm^2 and from 5.5 to 8.3, respectively. In the case of Ornithothoraces, the *WL* values of enantiornithines range between 0.240 g/cm^2 in *Longirostravis* and 0.59 g/cm^2 in *Protopteryx* and *Otogornis*, while the *AR* values vary from 4.0 in *Rapaxavis* to 8.4 in *Protopteryx*. The basal members of Ornithuromorpha have *WL* values ranging between 0.399 and 0.837 g/cm^2, which correspond to the values of *Gansus* (CAGS-04-CM003) and *Jianchangornis,* respectively. The lowest *AR* value (5.1) is estimated for *Longicrusavis*, while *Apsaravis* shows the highest value (10.0).

Table 13 show the differences between the values of *B*, S_L, *WL,* and *AR* estimated in this study and those provided by other authors using outline body reconstructions. It is evident from this comparison that our estimate of *B* for the Berlin *Archaeopteryx* (598 mm) differs more from the estimate obtained by Longrich (2006) than from the estimates provided by Yalden (1971; 1984), Elzanowski (2002) and Wellnhofer (2008). In the case of S_L, our estimate is similar to those of Wellnhofer (2008) and Longrich (2006), but differs to a greater extent from those derived using the other two previous models [e.g., Yalden (1971; 1984) and Elzanowski (2002)]. In regards to *Hongshanornis*, our estimates of *B* and S_L are close to those provided by Chiappe et al. (2014). Given that the specimen of *Hongshanornis* analyzed here (IVPP V14533) is not the same as the one used in the latter study (DNHM D2945/6), the small differences between these estimates can be interpreted as the result of intraspecific variation. A more important difference involves the *BM* values (i.e., 88 g vs. 65 g), which impact the values calculated for *WL* (i.e., 0.540 g/cm^2 vs. 0.406 g/cm^2).

Discussion

From an aerodynamic point of view, the easiest approach to characterizing the wings of a bird is to measure their wingspan (*B*) and lift surface (S_L) (Pennycuick 2008). For a given species, the values of these variables and the body mass (*BM*) allow for the calculation of two aerodynamic parameters, wing loading (*WL*) and aspect ratio (*AR*), known to influence flight speed, energy expenditure and maneuverability (Viscor and Fuster 1987; Rayner 1988; Norberg 2002; Alerstam et al. 2007; Meseguer and Sanz 2007; Pennycuick 2008). This relation highlights the need to obtain reliable estimates of *BM, B* and S_L for the purpose of characterizing the flight of extinct avian taxa. Serrano et al. (2015) provided regression models that allow estimating *BM* in

514 a variety of stem birds. Previously, the only way to obtain B and S_L estimates for stem birds was
515 to use outline reconstructions of their wings, which were only available for fossils with well-
516 preserved wing feathers. However, these reconstructions are relatively scarce and with the recent
517 exception of the basal ornithuromorph *Hongshanornis* (Chiappe et al. 2014), they have been
518 mainly focused on the Berlin *Archaeopteryx* (e.g., Yalden 1971; 1984; Elzanowski 2002;
519 Longrich 2006; Wellnhofer 2008). Estimates of B and S_L for extinct crown birds (e.g. Rando et
520 al. 2010; Ksepka 2014) have also been obtained using regression equations derived from skeletal
521 measures in crown Aves. However, these predictive equations are subject to a number of sources
522 of error and bias, especially for stem taxa, arising from statistical issues, phylogenetic history,
523 locomotor adaptations, and diagenetic compaction (Serrano et al. 2015). Developing predictive
524 models that are less influenced by these potential sources of error and bias, as those presented
525 here, improves the reliability of the estimates.

526 The statistical limitations imposed by taxonomic unevenness in the modern dataset (see
527 Smith 2002; Mendoza et al. 2006, De Esteban-Trivigno et al. 2008) and the correlation between
528 the predictor variables incorporated in the model (i.e., multicollinearity; see Mitchell-Olds and
529 Shaw 1987; Bowerman and O'Connell 1990; Quinn and Keough 2001) had no significant effects
530 on the models' ability to estimate B and S_L. Additionally, the lack of generalizability of multiple
531 regressions, especially when obtained by automatic stepwise procedure (see Smith 2002;
532 Whittingham et al. 2006), was not applicable to the MR functions for B and S_L. In contrast, the
533 stepwise procedure improved the predictive power stemward (i.e. on sister-taxa specimens that
534 were not included to derive a given function).

535 In regards to the phylogenetic dependence of the data, the fact that the confidence
536 intervals for the slopes of the PICs regressions overlapped with those derived from OLS on

537 species means (Tables 4 and 5) suggests that phylogeny did not play a major role in the scaling
538 patterns of flying birds for B and S_L, at least for the forelimb elements measured. Such results
539 indicate that the uncorrected models are not biased by phylogenetic dependence of the modern
540 dataset.

541 Given that it is not possible to directly examine the relationship between the dimensions
542 of the forelimb elements and B or S_L in the fossils, a certain degree of error must always be
543 accepted by these types of paleobiological studies (see Carrano 2001; Smith 2002; Packard et al.
544 2009; Campione and Evans 2012; Field et al. 2013). Such an error is compounded by the fact
545 that progressively stemward taxa (in our case, Mesozoic non-neornithine birds) exhibit body
546 plans that are progressively less similar to those depicted by the crown clade, which means that
547 the applicability of the regressions is likely to decrease proportionately. For these reasons, it is
548 important that the predictive equations derived from extant birds assume similar scaling patterns
549 for the extinct species under consideration. In order to obtain values of B and S_L as reliably as
550 possible for the fossil specimens included here, we have taken into account the scaling
551 relationships of the predictor variables in both modern and stem birds.

552 Wing span (B) estimations depend on the length of the anatomical elements of the wing
553 (i.e. humerus, ulna, carpometacarpus, major digit, and longest primary feather), the angles of the
554 elbow and the wrist, and the length of the body region between the wings. Given that the length
555 proportions between the main wing bones for stem birds are similar to those of modern birds
556 (e.g. Middleton and Gatesy 2000, Dyke and Nudds 2009, Benson and Choiniere 2013), the
557 length of these elements were used to derive regression models for predicting B values in stem
558 Mesozoic birds. Similarly, the longest primary feather length (L_{prim}) was used as an estimator of
559 B because our results indicate that stem taxa show L_{prim}/L_{wsk} values within the range of neognaths

(Fig. 3). These results are contrary to previous claims indicating that L_{prim} was significantly shorter in some stem taxa (e.g., Archaeopterygidae and Enantiornithes) (Wang et al. 2011). We argue that such disagreement stems in part from the fact that the limited taxonomic sample used by Wang et al. (2011)—50 species from a non-specified number of families and orders—covers a smaller range of morphological variation than the one represented here—131 species from 67 families and 22 orders—. In addition to the similarity in the proportions of the wing bones between modern and Mesozoic taxa, the angle of the elbow and wrist of the stretched wings of stem birds fall within the wide spectrum of modern birds, which ranges from very close to very open joint angles (e.g., swifts and many passerines, respectively; see Videler 2006). Our results also show that the relationship between the proxies used in stem birds for the interclavicular distance (*WF*) and wingspan (*Lw**) are within the neognath's range, because the slope obtained for the 10 fossil specimens analyzed here is included in the confidence interval of the slope for modern birds (Fig. 2). Therefore, the prediction of *B* in stem birds using equations derived from modern ones is presumably not affected by morphological differences in the width of the corporal region situated between the wings. Finally, we obtained three MR models that allow estimating *B* in fossils with a varying degree of skeletal completeness. Unsurprisingly, the models that incorporate a high number of measures from wing elements have higher predictive power than those based on a low number (Table 6).

S_L correlates with *B* because it spans the area of both wings and includes the area corresponding to the part of the body situated between the wings (Pennycuick 2008). In addition, S_L correlates also with *BM* because it supports the bird's weight in the air (Greenwalt 1975; Viscor and Fuster 1987; Rayner 1988; Álvarez et al. 2001; Norberg 2002; Sanz et al. 2002; Meseguer and Sanz-Andrés 2007; Pennycuick 2008). Hence, the value of S_L can be estimated

583 from the same variables as *B* and *BM*, but it also depends on the wing chord (here considered the
584 chord at wrist level, *c*). As in the case of the models derived for predicting *BM* by Serrano et al.
585 (2015), we have used a methodology whereby the osteological and feather measurements
586 selected for each stem group were those that showed scaling patterns in extinct birds similar to
587 those found among modern birds. This procedure generated 11 clade-specific MR functions,
588 which were used to predict S_L for non-neornithine extinct birds. Within each stem taxa, several
589 functions were derived for applying them to a given fossil, in accordance with the variables that
590 it preserved. Among these MR models, those that included the two variables related to the
591 plumage (i.e. L_{prim} and *c*) as well as some osteological measurements had the highest predictive
592 power (Table 9).

593 Given that the variables used for deriving MR models for *B* and S_L show similar scaling
594 patterns in stem and modern birds, these models are robust against the bias generated by the
595 deviations in body plans between crown and stem clades.

596 Moreover, a number of studies on the functional adaptations of modern birds have shown
597 that the skeletal morphology of the wings is highly constrained by the dynamics of locomotion
598 (e.g. Cubo and Casinos 1998; De Margerie et al. 2005; Habib and Ruff 2008; Hinic-Frlog and
599 Motani 2010; Simmons 2010; Bell and Chiappe 2011; Simmons et al. 2011; Dececchi and
600 Larsson 2013), which could result in biased predictions of *B* and S_L for new specimens, either
601 extant or extinct ones. Compared to the most predictive univariate models, ulnar length (*UL*) and
602 dorsoventral width of the distal humerus (*deHW*), the *post hoc* analyses indicate that a combined
603 use of variables in the MR functions minimizing the effects of flying adaptions on the scaling
604 between wing elements and *B* or S_L (Table 10). Similar results were obtained for the effects of
605 locomotion on the predictions of *BM* values (Serrano et al. 2015).

606 The effect of diagenetic compaction can alter the osteological dimensions of fossil birds,
607 mainly the width of the ends of the long bones. The estimation of *B* values is not influenced by
608 this potential bias, as this parameter is only estimated from longitudinal measures. In contrast,
609 the S_L values obtained from univariate models could be biased in severely compacted fossils.
610 However, the use of a step-by-step methodology to obtain a specific MR model for each
611 Mesozoic stem taxa reduces these potential biases, as in the case of the *BM* predictions (see
612 Serrano et al. 2015).

613 Additional evidence on the reliability of the predictions obtained from the MR models
614 derived here comes from the congruence of the values estimated for *Archaeopteryx* (Berlin
615 specimen) and *Hongshanornis* (two different specimens) with those obtained by the most recent
616 outline reconstructions (i.e., Wellnhofer 2008, Chiappe et al. 2014). Furthermore, the *B* value of
617 *Archaeopteryx* obtained by our MR models is also very close to those derived from reconstructed
618 models (Yalden 1971; 1984; Elzanowski, 2002) and its S_L value approaches the one estimated by
619 Longrich (2006) on the basis of direct measurements of the Berlin specimen. The largest
620 differences between previously published models are related to the *BM* values, which were
621 previously calculated by either volumetric body displacement or univariate allometric inference
622 (Table 13). The advantages of using MR models for obtaining reliable predictions of *BM* in stem
623 birds were discussed in Serrano et al. (2015).

624 Modern birds seem to display a pattern of "many-to-one mapping", in which several
625 distinct combinations of wing elements lengths may result in wings that are functionally
626 equivalent and show similar *AR* values. For this reason, combining evidence from different wing
627 components allows reliable inferences on wing shape and function in neognaths to be derived
628 (Hieronymus 2015). Therefore, we suggest that the selection of variables procedure developed in

this study allows reliable stemward inferences of *AR* (and *WL*) values for extinct birds. Furthermore, in spite of the fact that the values predicted by MR models have an associated prediction error, calculation of flight parameters from them (i.e. *WL* and *AR*) does not result in an accumulation of this error. Hence, the empirical prediction errors of *WL* and *AR* resemble those of the basic variables estimated with the highest |%MPE| values (e.g. *BM* vs. *WL* and S_L vs. *AR;* see Table 11). Because this is the first time in which *WL* and *AR* values have been calculated for an extensive dataset, and the associated prediction errors are calibrated empirically, there is no reference for comparing the accuracy of the *WL* and *AR* estimates. Nonetheless, comparisons with the prediction errors provided for *BM* in previous studies (Serrano et al. 2015: Table 5 and references therein) indicate that these errors are relatively low. Therefore, the *WL* and *AR* estimates reported here can be considered suitable. In addition, given that the models to estimate *BM*, *B*, and S_L are generalizable and robust against the sources of bias discussed above, this reliability can also be applied to the flight parameters derived from them.

Implications for the Flight of Early Aves

The estimates of *WL* and *AR* for the fossils included in this study show that these stem birds fall within the range of variation of modern birds (Fig. 5). In regards to *WL*, the range of variation of continuous flappers overlaps the ranges of the remaining modern flight modes. Diving birds (e.g. alcids, loons, and diving ducks), with high wing loads, fall in the uppermost range of the continuous flapping spectrum. Specimens of Archaeopterygidae, Jeholornithidae, Confuciusornithidae, and early Ornithuromorpha show wing loads that are higher than those of Sapeornithidae and Enantiornithes. The latter clade includes the only stem birds that overlap with modern bounding fliers; the remaining stem birds included in this study overlap with other flight

652 modes (i.e., continuous flapping, flap and glide, soaring). The *WL* values of sapeornithids
653 overlap with the bulk of the soarers. With respect to *AR*, modern birds adapted to either dynamic
654 soaring or diving have the highest values (above from 9.0). The *AR* estimates for stem birds
655 ranged between 4.0 and 10.0, clearly within the range of most of modern birds (Fig. 5B). A
656 combination of high wing loading and very high aspect ratio is not inferred in any of the fossil
657 birds analyzed. This suggests that these birds were neither dynamic soaring nor diving
658 specialists. Nonetheless, the combination of their *WL* and *AR* values indicate that these stem
659 birds had evolved flight modes comparable to those of many modern birds. Furthermore, our
660 data suggests that bounding flight, common among modern Passeriformes, could have been
661 achieved by some enantiornithines.
662
663 **Conclusions**
664 Our MR models for predicting wingspan (B) and lift surface (S_L) have a high prediction power
665 and generalizability, which in turn indicate that they can provide accurate estimations. These
666 models also minimize the potential biases that could result from statistical issues and differences
667 in phylogenetic legacy, locomotor adaptations, and degree of taphonomic alteration.
668 Furthermore, using models of variables that scale similarly in crown and stem taxa, and that
669 produce estimates similar to values obtained from direct measurements of preserved wing
670 outlines, provides indirect support for the accuracy of the values predicted for non-neornithine
671 Aves. Therefore, accuracy, generalizability, and robustness of our MR models indicate that the B
672 and S_L values here obtained for stem birds are reliable. Given that the *BM* estimates are also
673 minimally biased (Serrano et al. 2015), the values calculated for the aerodynamic parameters,
674 *WL* (i.e. $BM\ g/S_L$) and *AR* (i.e. B^2/S_L), can be considered minimally influenced by the biases

675 mentioned above. In addition, although the estimations of *BM, B,* and S_L carry an associated
676 error, there is no cumulative error in the calculation of *WL* and *AR* from them; both aerodynamic
677 parameters also show relatively low prediction errors. Our estimates of *AR* and *WL* show that
678 stem birds could have flown using most of the flight modesknown for modern birds, although we
679 have found no evidence for either dynamic soarers or flighted divers (hovering was not
680 analyzed). Evidence for bounding flight was found only among enantiornithines.
681

682 **Acknowledgements**

683 We dedicate this article to the memory of José Meseguer Ruiz, whose seminal work on
684 aeronautical engineering will remain as a legacy and source of inspiration for future generations
685 of scientists interested in animal flight. We thank N. Campione, A. Casinos, J. Cubo, J.
686 Domínguez, B. Figueirido, A. Martín-Serra, J. Marugán-Lobón, J.A. Pérez-Claros and two
687 anonymous reviewers for providing suggestions and comments on earlier versions of this paper.
688 We are also grateful to R. Faucett (UWBM, Seattle), J. Barreiro (MNCN, Madrid), and Z. Zhou
689 (IVPP, Beijing) for granting us with access to the specimens under their care, and to T. Hayden
690 for providing editorial support. This work was supported by Research Project (CGL2011-30334)
691 and by FPI PhD Research Fellowship to F. Serrano, funded by the Spanish Ministry of Economy
692 and Competitiveness, and the agreement between the Natural History Museum of Los Angeles
693 County (Dinosaur Institute, NHM) and the Sierra Elvira Foundation (Granada, Spain).
694

695 **References**

696 Alerstam, T., M. Rosén, J. Bäckman, P. G. Ericson, and O. Hellgren. 2007. Flight speeds among
697 bird species: allometric and phylogenetic effects. PLoS Biology, 5(8):e197.

698 Allen, V., K. Bates, Z. Li, and J. Hutchinson. 2013. Linking the evolution of body shape and
699 locomotor biomechanics in bird-line Archosaurs. Nature 497:104-107.
700 Alvarez, J. C., J. Meseguer, E. Meseguer, and A. Perez. 2001. On the role of the alula in the
701 steady flight of birds. Ardeola 48(2):161-173.
702 Bell, A., and L. M. Chiappe. 2011. Statistical approach for inferring ecology of Mesozoic birds.
703 Journal of Systematic Palaeontology 9:119-133.
704 Benson, R. B., and J. N. Choiniere. 2013. Rates of dinosaur limb evolution provide evidence for
705 exceptional radiation in Mesozoic birds. Proceedings of the Royal Society B: Biological
706 Sciences 280(1768):20131780.
707 Bowerman, B. L., and R. T. O'Connell. 1990. Linear Statistical Models: An Applied Approach.
708 Duxbury Press, California.
709 Brockelhurst, N., P. Upchurch, P. Mannion, and J. O´Connor. 2012. The completeness of the
710 Fossil Record of Mesozoic Birds: Implications for Early Avian Evolution. PLoS One
711 7:e39056.
712 Bruderer, B., D. Peter, A. Boldt, and F. Liechti. 2010. Wing-beat characteristics of birds
713 recorded with tracking radar and cine camera. Ibis 152:272-291.
714 Brusatte, S. L., J. K. O'Connor, and E. D. Jarvis. 2015. The origin and diversification of birds. *Curr.*
715 *Biol*. 25, 888-898.
716 Burgers, P., and L. M. Chiappe. 1999. The wing of *Archaeopteryx* as a primary thrust generator.
717 Nature 399(6731):60-62.
718 Campbell, K. E., and L. Marcus. 1992. The relationships of hindlimb bone dimensions to body
719 weight in birds. Natural History Museum of Los Angeles County Science Series 36:395-412.
720 Campione, N., and D. Evans. 2012. A universal scaling relationship between body mass and
721 proximal limb bone dimensions in quadrupedal terrestrial tetrapods. BMC Biology 10:60.

722 Carrano M. T. 2001. Implications of limb bone scaling, curvature and eccentricity in mammals
723 and non-avian dinosaurs. Journal of Zoology 254:41-55.
724 Cawley G. C., and G. J. Janacek. 2010. On allometric equation s for predicting body mass of
725 dinosaurs. Journal of Zoology 280:355-361.
726 Chan, N. R., G. J. Dyke, and M. J. Benton. 2013. Primary feather lengths may not be important
727 for inferring the flight styles of Mesozoic birds. Lethaia 46(2):146-153.
728 Chatterjee, S., and R. J. Templin. 2003. The flight of *Archaeopteryx*. Naturwissenschaften
729 90(1):27-32.
730 Chiappe, L. M. 2007. Glorified dinosaurs: the origin and early evolution of birds. John Wiley &
731 Sons. Inc., Hoboken, New Jersey
732 Chiappe, L. M., and J. O. Calvo. 1994. *Neuquenornis volans*, a new Late Cretaceous bird
733 (Enantiornithes: avisauridae) from Patagonia, Argentina. Journal of Vertebrate Paleontology
734 14(2):230-246.
735 Chiappe, L. M., and L. Witmer. 2002. Mesozoic birds: above the heads of dinosaurs. University
736 of California Press, Berkeley.
737 Chiappe, L.M. and Q. Meng. 2016. Birds of Stone. Johns Hopkins University Press, Baltimore.
738 Chiappe, L. M., S. A. Ji, Q. Ji, and M. A. Norell. 1999. Anatomy and systematics of the
739 Confuciusornithidae (Theropoda, Aves) from the late Mesozoic of northeastern China.
740 Bulletin of the American Museum of Natural History 242.
741 Chiappe, L. M., B. Zhao, J. K. O'Connor, G. Chunling, X. Wang, M. Habib, and X. Cheng.
742 2014. A new specimen of the Early Cretaceous bird *Hongshanornis longicresta*: insights into
743 the aerodynamics and diet of a basal ornithuromorph. PeerJ 2:e234.

744 Clarke, J., and M. A. Norell. 2002. The Morphology and Phylogenetic Position of *Apsaravis*
745 *ukhaana* from the Late Cretaceous of Mongolia. American Museum Novitates 3387.
746 Clarke, A., P. Rothery, and N. J. Isaac. 2010. Scaling of basal metabolic rate with body mass and
747 temperature in mammals. Journal of Animal Ecology 79:610-619.
748 Clifford, D., N. Cressie, J. R. England, S. H. Roxburgh and K. I. Paul. 2013. Correction factors
749 for unbiased, efficient estimation and prediction of biomass from log–log allometric models.
750 Forest Ecology and Management 310:375-381.
751 Close, R. A., and E. J. Rayfield. 2012. Functional morphometric analysis of the furcula in
752 Mesozoic birds. PLoS One 7(5):e36664.
753 Cubo J., and A. Casinos. 1998. Biomechanical significance of cross-sectional geometry of avian
754 long bones. European Journal of Morphology 36:19-28.
755 De Esteban-Trivigno, S., M. Mendoza, and M. De Renzi. 2008. Body mass estimation in
756 Xenarthra: a predictive equation suitable for all quadrupedal terrestrial placentals? Journal of
757 Morphology 269:1276-1293.
758 De Margerie, E., S. Sanchez, J. Cubo, J. Castanet. 2005. Torsional resistance as a principal
759 component of the structural design of long bones: Comparative multivariate evidence in birds.
760 Anatomical Record Part A 282A:49-66.
761 Dececchi, T. A., and H. C. Larsson. 2013. Body and limb size dissociation at the origin of birds:
762 uncoupling allometric constraints across a macroevolutionary transition. Evolution 67:2741-
763 2752.
764 Díaz-Uriarte, R., and T. Garland. 1996. Testing hypotheses of correlated evolution using
765 phylogenetically independent contrasts: Sensitivity to deviations from Brownian motion.
766 Systematic Biology 45:27-47.

767 Dyke, G. J., and R. L. Nudds. 2009. The fossil record and limb disparity of enantiornithines, the
768 dominant flying birds of the Cretaceous. Lethaia 42(2):248-254.
769 Eisenhauer, J. G. 2003. Regression through the origin. Teaching Statistics 25(3):76-80.
770 Elzanowski, A. 2002. Archaeopterygidae (Upper Jurassic of Germany). Pp. 129-159 *in* L. M.
771 Chiappe and L. M. Witmer, eds. Mesozoic Birds: Above the Heads of Dinosaurs. University
772 of California Press, Berkeley.
773 Evangelista, D., G. Cardona, E. Guenther-Gleason, T. Huynh, A. Kwong, D. Marks, N. Ray, A.
774 Tisbe, K. Tse, and M. Koehl. 2014. Aerodynamic characteristics of a feathered dinosaur
775 measured using physical models: effects of form on static stability and control effectiveness.
776 PLoS One 9:e85203.
777 Felsenstein, J. 1985. Phylogenies and the Comparative Method. American Naturalist 125:1-15.
778 Field, D. J., C. Lynner, C. Brown, and S. A. Darroch. 2013 Skeletal correlates for body mass
779 estimation in modern and fossil flying birds. PLoS One 8:e82000.
780 Figueirido, B., F. J. Serrano-Alarcón, G. J. Slater, and P. Palmqvist. 2010. Shape at the cross-
781 roads: homoplasy and history in the evolution of the carnivoran skull towards herbivory.
782 Journal of Evolutionary Biology 23:2579-2594.
783 Gao, C., L. M. Chiappe, F. Zhang, D. Pomeroy, C. Shen, A. Chinsamy, and M. Walsh. 2012. A
784 subadult specimen of the Early Cretaceous bird *Sapeornis chaoyangensis* and a taxonomic
785 reassessment of sapeornithids. Journal of Vertebrate Paleontology 32: 1103-1112.
786 Gingerich, P., B. Smith, and K. Rosenberg. 1982. Allometric scaling in the dentition of primates
787 and prediction of body weight from tooth size in fossils. American Journal of Physical
788 Anthropology 58:81-100.

789 Greenwalt, C. H. 1975. The flight of birds. Transactions of the American Philosophical Society
790 (NS) 65:1-65.
791 Habib, M. B., and C. Ruff. 2008. The effects of locomotion on the structural characteristics of
792 avian limb bones. Zoological Journal of the Linnean Society 153:601-624.
793 Hackett, S., R. Kimball, S. Reddy, R. Bowie, E. Braun, M. Braun, J. Chojnowski, W. Cox, K.
794 Han, J. Harshman, C. Huddleston, B. Marks, K. Miglia, W. Moore, F. Sheldon, D. Steadman,
795 C. Witt, and T. Yuri. 2008. A Phylogenomic study of birds reveals their evolutionary history.
796 Science 320:1763-67.
797 Hansen, T. F., and S. H. Orzack. 2005. Assessing current adaptation and phylogenetic inertia as
798 explanations of trait evolution: the need for controlled comparisons. Evolution 59(10):2063-
799 2072.
800 Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford
801 University Press, Oxford.
802 Hayes, J. P. and J. S. Shonkwiler. 2006. Allometry, antilog transformations, and the perils of
803 prediction on the original scale. Physiological and Biochemical Zoology 79(3): 665-674.
804 Hinic-Frlog, S., and R. Motani. 2010. Relationship between osteology and aquatic locomotion in
805 birds: determining modes of locomotion in extinct Ornithurae. Journal of Evolutionary
806 Biology 23:372-385.
807 Hou, L. H. 1997. Mesozoic Birds of China. Feng-huang-ku Bird Park, Taiwan, (in Chinese).
808 Hurvich, C. M., and C. L. Tsai. 1990. The impact of model selection on inference in linear
809 regression. American Statistician 44:214-217.
810 Ksepka, D. T. 2014. Flight performance of the largest volant bird. Proceedings of the National
811 Academy of Sciences 111(29):10624-10629.

812 Laurin, M. 2004. The evolution of body size, Cope's rule and the origin of amniotes. Systematic
813 Biology 53:594-622.
814 Longrich, N. 2006. Structure and function of hindlimb feathers in *Archaeopteryx lithographica*.
815 Paleobiology 32:417-431.
816 MacLeod, N. 2004. Palaeo-math 101: Regression 2. The Palaeontological Association,
817 Newsletter 56:60-71.
818 MacNally, R. 2000. Regression and model-building in conservation biology, biogeography and
819 ecology: the distinction between –and reconciliation of– 'predictive'and 'explanatory'models.
820 Biodiversity & Conservation 9:655-671.
821 Maddison, W. P., and D. R. Maddison. 2011. Mesquite: a modular system for evolutionary
822 analysis. Version 2.75. http://mesquiteproject.org
823 Mendoza, M, C. M. Janis, and P. Palmqvist. 2006. Estimating the body mass of extinct
824 ungulates: a study on the use of multiple regression. Journal of Zoology 270:90-101.
825 Meseguer, J., and A. Sanz-Andrés. 2007. Aerodinámica del vuelo: aves y aeronaves. Cuadernos
826 Aena, 9. Aena, Aeropuertos Españoles y Navegación Aérea, España.
827 Meseguer, J., A. Sanz-Andrés, A. Pedro, M. I. Pérez Grande, S. N. Franchini, J. L. Sanz García,
828 and L. M. Chiappe. 2008. Control de capa limite en el vuelo a bajos números de Reynolds.
829 IAA. Ingeniería Aeronáutica y Astronáutica 387:15-24.
830 Meseguer, J., L. M. Chiappe, J. L. Sanz, F. Ortega, A. Sanz-Andrés, M. I. Pérez-Grande, and S.
831 Franchini. 2012. Lift devices in the flight of *Archaeopteryx*. Revista Española de
832 Paleontología 27(2):125-130.
833 Middleton, K. M., and S. M. Gatesy. 2000. Theropod forelimb design and evolution. Zoological
834 Journal of the Linnean Society 128(2):149-187.

Mitchell-Olds, T., and R. G. Shaw. 1987. Regression Analysis of Natural Selection: Statistical Inference and Biological Interpretation. Evolution 41:1149-1161.

Norberg, U. M. 2002. Structure, form, and function of flight in engineering and the living world. Journal of Morphology 252(1):52-81.

Nudds, R. L. 2007. Wing-bone length allometry in birds. Journal of Avian Biology 38:515-519.

Nudds, R. L., and J. M. V. Rayner. 2006. Scaling of body frontal area and body width in birds. Journal of Morphology 267(3):341-346.

Nudds, R. L., J. Atterholt, X. Wang, H. L. You, and G. J. Dyke. 2013. Locomotory abilities and habitat of the Cretaceous bird *Gansus yumenensis* inferred from limb length proportions. Journal of Evolutionary Biology 26(1):150-154.

O'Connor, J., K. Gao, and L. M. Chiappe. 2010. A new Ornithuromorph (Aves: Ornithothoraces) bird from the Jehol Group indicative of higher-level diversity. Journal of Vertebrate Paleontology 30:311-321.

O'Connor, J. K., L. M. Chiappe, and A. Bell. 2011. Pre-modern Birds: Avian Divergences in the Mesozoic. Pp. 39-114 *in* G. Dyke and G. Kaiser, eds. Living Dinosaurs. Wiley-Blackwell, Oxford.

Packard, G. C., T. J. Boardman, and G. F. Birchard. 2009. Allometric equations for predicting body mass of dinosaurs. Journal of Zoology 279:102-110.

Pennycuick, C. J. 2008. Modelling the Flying Bird. AP Theoretical Ecology Series Elsevier. Academic Press, Oxford.

Peters, W. S., D. S. Peters. 2009. Life history, sexual dimorphism and 'ornamental' feathers in the Mesozoic bird *Confuciusornis sanctus*. Biology Letters 5:817-882.

857 Price, T. 1997. Correlated evolution and independent contrasts. Philosophical Transactions of the
858 Royal Society of London Ser. B Biological Sciences 352:519-529.
859 Prum, R. O., J. S. Berv, A. Dornburg, D. J. Field, J. P. Townsend, E. M. Lemmon and A. R.
860 Lemmon. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation
861 DNA sequencing. Nature 526:569-577.
862 Quinn, G. P., and M. J. Keough. 2001. Experimental design and data analysis for biologists.
863 Cambridge University Press.
864 Rando, J. C., J. A. Alcover, and J. C. Illera. 2010. Disentangling ancient interactions: a new
865 extinct passerine provides insights on character displacement among extinct and extant island
866 finches. PLoS One 5(9):e12956-e12956.
867 Rayner, J. M. 1988. Form and function in avian flight. Current Ornithology 5:1-66.
868 Revell, L. J. 2010. Phylogenetic signal and linear regression on species data. Methods in Ecology
869 and Evolution 1:319-329.
870 Ricklefs, R. E., and J. M. Starck. 1996. Applications of phylogenetically independent contrasts: a
871 mixed progress report. Oikos 77:167-172.
872 Sanz, J. L., J. F. Bonaparte, and A. Lacasa. 1988. Unusual early Cretaceous birds from Spain.
873 Nature 331(6155):433-435.
874 Sanz, J. L., L. M. Chiappe, and A. D. Buscalioni. 1995. The osteology of *Concornis lacustris*
875 (Aves, Enantiornithes) from the Lower Cretaceous of Spain and a reexamination of its
876 phylogenetic relationships. American Museum Novitates 3133.
877 Sanz, J. L., L. M. Chiappe, B. P. Pérez-Moreno, A. D. Buscalioni, J. J. Moratalla, F. Ortega, and
878 F. J. Poyato-Ariza. 1996. An Early Cretaceous bird from Spain and its implications for the
879 evolution of avian flight. Nature 382(6590):442-444.

Sanz J. L., J. C. Álvarez, C. Soriano, F. Hernández-Carrasquilla, B. Pérez-Moreno, and J. Meseguer 2002. Wing loading in Primitive Birds. Pp. 253-258 *in* Z. Zhou and F. Zhang, eds. Proceedings of the 5th Symposium of the Society of Avian Paleontology and Evolution. Science Press, Beijing.

Senter, P. 2006. Scapular orientation in theropods and basal birds, and the origin of flapping flight. Acta Palaeontologica Polonica 51(2):305.

Serrano, F. J., P. Palmqvist, and J. L. Sanz. 2015. Multivariate analysis of neognath skeletal measurements: implications for body mass estimation in Mesozoic birds. Zoological Journal of the Linnean Society 173(4):929-955.

Simmons, E. L. 2010. Forelimb skeletal morphology and flight mode evolution in pelecaniform birds. Zoology 113:39-46.

Simmons, E. L., T. L. Hieronymus, P. M. O´Connor. 2011. Cross sectional geometry of the forelimb skeleton and flight mode in pelecaniform birds. Journal of Morphology 3:958-971.

Smith, R. J. 1984. Allometric scaling in comparative biology: problems of concept and method. American Journal of Physiology Regulatory Integrative and Comparative Physiology 246:152-160.

Smith, R. J. 1993. Logarithmic transformation bias in allometry. American Journal of Physiological Anthropology 90:215-228.

Smith, R. J. 2002. Lead review: estimation of body mass in paleontology. Journal of Human Evolution 43:271-287.

Smith, R. J. 2009. Use and Misuse of the Reduced Major Axis for Line-Fitting. American Journal of Physiological Anthropology 140:476-486.

902 Snowdon P. 1991. A ratio estimator for bias correction in logarithmic regressions. Canadian
903 Journal of Forest Research 21: 720-724.
904 Sokal, R. R., and J. F. Rohlf. 1986. Introduction to Biostatistics. W.H. Freeman and Company,
905 San Francisco.
906 Tennekes, H. 2009. The simple science of flight: from insects to jumbo jets. MIT press.
907 Turner, A., and H. J. O'Regan. 2002. The assessment of size in fossil Felidae. Estudios
908 Geológicos 58:45-54.
909 Vanhooydonck, B., R. S. James, J. Tallis, P. Aerts, Z. Tadic, K. Tolley, and A. Herrel. 2014. Is
910 the whole more than the sum of its parts? Evolutionary trade-offs between burst and sustained
911 locomotion in lacertid lizards. Proceedings of the Royal Society B: Biological Sciences
912 281:20132677.
913 Videler, J. J. 2005. Avian Flight. Oxford Ornithology Series. Oxford University Press, Oxford.
914 Viscor, G., and J. F. Fuster. 1987. Relationships between morphological parameters in birds with
915 different flying habits. Comparative Biochemistry and Physiology Part A: Physiology
916 87(2):231-249.
917 Wang, X., R. L. Nudds, and G. J. Dyke. 2011. The primary feather lengths of early birds with
918 respect to avian wing shape evolution. Journal of Evolutionary Biology 24(6):1226-1231.
919 Wang, X., R. L. Nudds, C. Palmer, and G. J. Dyke. 2012. Size scaling and stiffness of avian
920 primary feathers: implications for the flight of Mesozoic birds. Journal of Evolutionary
921 Biology 25:547-555.
922 Wellnhofer, P. 2008. *Archaeopteryx*, der urvogel von Solnhofen. Verlag Dr. Friedrich Pfeil,
923 München.
924 Yalden, D. W. 1971. The flying ability of *Archaeopteryx*. Ibis 113(3):349-356.

Yalden, D. W. 1984. What size was *Archaeopteryx*? Zoological Journal of the Linnean Society 82:177-188.

You, H. L., M. C. Lamanna, J. D. Harris, L. M. Chiappe, J. M. O'Connor, S. Ji, L. Jun-chang, Y. Chong-xi, L. Da-qing, Z. Xing, K.J. Lacovara, P. Dodson, and Q. Ji. 2006. A nearly modern amphibious bird from the Early Cretaceous of northwestern China. Science 312(5780):1640-1643.

Zhang, F., and Z. Zhou. 2000. A primitive enantiornithine bird and the origin of feathers. Science 290(5498): 1955-1959.

Zhang, F., Z. Zhou, and M. J. Benton. 2008. A primitive confuciusornithid bird from China and its implications for early avian flight. Science in China Series D: Earth Sciences 51(5):625-639.

Zhou, Z., and F. Zhang. 2002. A long-tailed, seed-eating bird from the Early Cretaceous of China. Nature 418(6896):405-409.

Zhou, Z., and F. Zhang. 2003. Anatomy of the primitive bird *Sapeornis chaoyangensis* from the Early Cretaceous of Liaoning, China. Canadian Journal of Earth Sciences, 40(5):731-747.

Zhou, Z., L. M. Chiappe, and F. Zhang. 2005. Anatomy of the Early Cretaceous bird *Eoenantiornis buhleri* (Aves: Enantiornithes) from China. Canadian Journal of Earth Sciences 42(7):1331-1338.

Zhou, Z., F. Zhang, and Z. H. Li. 2009. A new basal ornithurine bird (*Jianchangornis microdonta* gen. et sp. nov.) from the Lower Cretaceous of China. Vertebrata PalAsiatica 47(4):299-310.

948 FIGURE CAPTIONS

949

950 **Figure 1.** Measurements of wing elements used in this study (defined in Table 1). A, skeletal
951 forelimb elements: humerus (cranial aspect), ulna (dorsal aspect), radius (dorsal aspect). B, bird
952 silhoutte for illustrating the furcula and feathered wing measurements.

953

954 **Figure 2.** Relationship between interclavicle distance (*WF*) and a proxy of wingspan, *Lw** (taken
955 as the cumulative length of *HL*, *UL*, *CmL*, *LDII* and L_{prim}) in modern flying birds and those stem
956 birds in which these variables could be unequivocally measured. The grey line corresponds to the
957 RMA slope for Neognathae (log *WF* = 1.142 log *Lw** – 1.555; N= 326). The black line
958 represents the RMA slope for stem birds (log *WF* = 1.178 log *Lw** – 1.607; N= 10). R^2 and the *p*-
959 values of the regressions (obtained with permutation test with 10000 replicates) are displayed in
960 the graph.

961

962 **Figure 3.** Proportion between the longest primary feather (L_{prim}) and the length of the wing
963 skeleton (Lw_{sk}) in the main avian taxa. Abbreviations: Arch – Archaeopterygidae, Jeh –
964 Jeholornithidae, Sape – Sapeornithidae, Conf – Confuciusornithidae, Enan – Enantiornithes,
965 Orph – basal members of Ornithuromorpha, Neo – Neognathae. Box length in neognaths
966 indicates the interquartile range (25th and 75th percentiles) and horizontal lines show the 5–95%
967 confidence limits. Grey points are outliers.

968

969 **Figure 4.** Scaling of the anatomical variables against estimated wing surface (S_L) in modern and
970 stem birds. The biplots are used in the procedure of selection of variables for each basal group of

971 Mesozoic birds (identified with different symbols). Each of 16 anatomical variable was plotted
972 against a combination of the remaining variables used for estimating S_L (Table 7). Each biplot
973 shows the OLS regression line fitted for extant birds (grey circles) with the 95% confidence
974 intervals for S_L predictions (dotted lines), the R^2 value and the *p*-value of the regression (obtained
975 with permutation test with 10000 replicates).
976
977 **Figure 5.** Range of values for the aerodynamic variables calculated in the present study for stem
978 avian taxa and flying neognaths. Modern birds are distributed according their predominant flight
979 mode. Each box plot represents one stem taxa or modern flight mode. Box length shows the
980 interquartile range (25th and 75th percentiles). Whiskers length indicate the 5–95% confidence
981 limits. Unfilled circles represents individuals out of this range, and asterisks indicate outliers. **A.**
982 Wing loading values. **B.** Aspect ratio values.

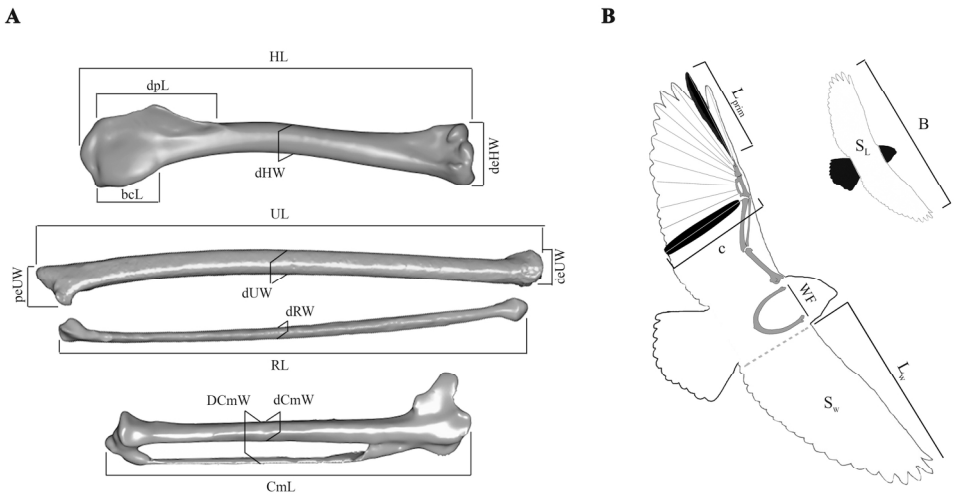

Figure 1. Measurements of wing elements used in this study (defined in Table 1). A, skeletal forelimb elements: humerus (cranial aspect), ulna (dorsal aspect), radius (dorsal aspect). B, bird silhoutte for illustrating the furcula and feathered wing measurements.

85x48mm (600 x 600 DPI)

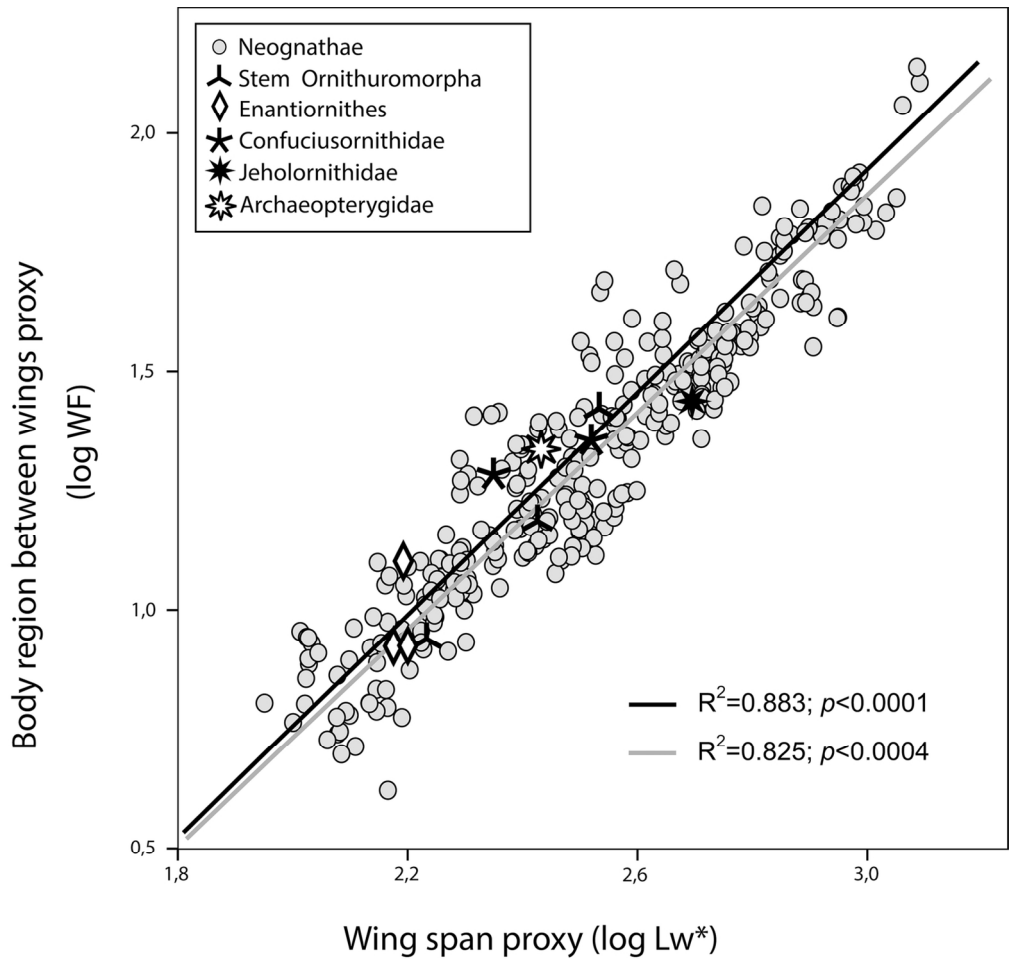

Figure 2. Relationship between interclavicle distance (WF) and a proxy of wingspan, Lw* (taken as the cumulative length of HL, UL, CmL, LDII and Lprim) in modern flying birds and those stem birds in which these variables could be unequivocally measured. The grey line corresponds to the RMA slope for Neognathae (log WF = 1.142 log Lw* – 1.555; N= 326). The black line represents the RMA slope for stem birds (log WF = 1.178 log Lw* – 1.607; N= 10). R2 and the p-values of the regressions (obtained with permutation test with 10000 replicates) are displayed in the graph.

66x63mm (600 x 600 DPI)

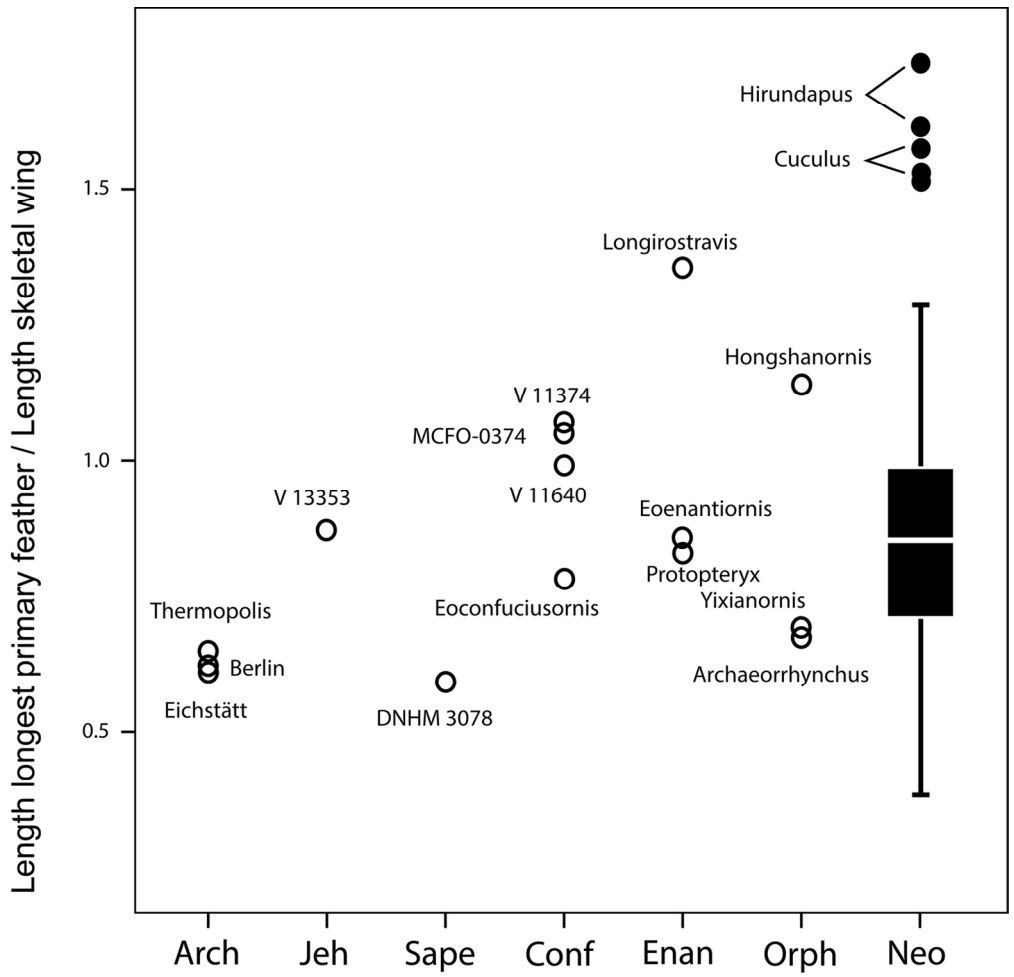

Figure 3. Proportion between the longest primary feather (Lprim) and the length of the wing skeleton (Lwsk) in the main avian taxa. Abbreviations: Arch – Archaeopterygidae, Jeh – Jeholornithidae, Sape – Sapeornithidae, Conf – Confuciusornithidae, Enan – Enantiornithes, Orph – basal members of Ornithuromorpha, Neo – Neognathae. Box length in neognaths indicates the interquartile range (25th and 75th percentiles) and horizontal lines show the 5–95% confidence limits. Grey points are outliers.

67x65mm (600 x 600 DPI)

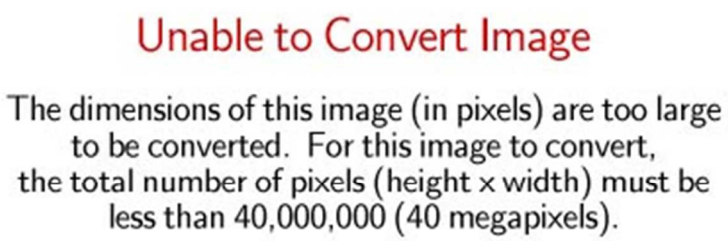

Figure 4. Scaling of the anatomical variables against estimated wing surface (SL) in modern and stem birds. The biplots are used in the procedure of selection of variables for each basal group of Mesozoic birds (identified with different symbols). Each of 16 anatomical variable was plotted against a combination of the remaining variables used for estimating SL (Table 7). Each biplot shows the OLS regression line fitted for extant birds (grey circles) with the 95% confidence intervals for SL predictions (dotted lines), the R2 value and the p-value of the regression (obtained with permutation test with 10000 replicates).

Unable to Convert Image

The dimensions of this image (in pixels) are too large to be converted. For this image to convert, the total number of pixels (height x width) must be less than 40,000,000 (40 megapixels).

Unable to Convert Image

The dimensions of this image (in pixels) are too large to be converted. For this image to convert, the total number of pixels (height x width) must be less than 40,000,000 (40 megapixels).

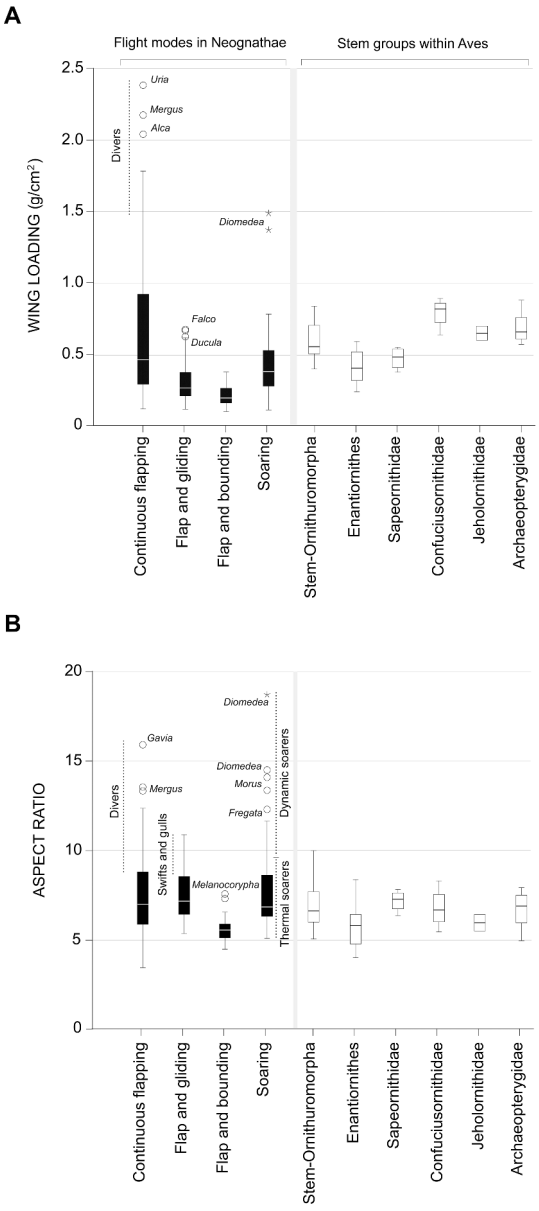

Figure 5. Range of values for the aerodynamic variables calculated in the present study for stem avian taxa and flying neognaths. Modern birds are distributed according their predominant flight mode. Each box plot represents one stem taxa or modern flight mode. Box length shows the interquartile range (25th and 75th percentiles). Whiskers length indicate the 5–95% confidence limits. Unfilled circles represents individuals out of this range, and asterisks indicate outliers. A. Wing loading values. B. Aspect ratio values.

158x356mm (600 x 600 DPI)

Table 1. Definition of measurements of the forelimb elements used in this study, which are illustrated in Fig.1.

HL	Length of humerus
dpL	Length of delto-pectoral crest
bcL	Length of bicipital crest
deHW	Dorsoventral width of distal humerus
dHW	Dorsoventral width at midshaft of the humerus
UL	Length of ulna
peUW	Dorsoventral width of proximal ulna
deUW	Dorsoventral width of distal ulna
dUW	Craniocaudal width at midshaft of the ulna
RL	Length of radius
dRW	Craniocaudal width at midshaft of the radius
CmL	Length of carpometacarpus from the semilunate to the joint with major digit
dCmW	Craniocaudal width at midshaft of major metacarpal
DCmW	Midshaft width from the cranial edge of major metacarpal to the caudal edge of minor metacarpal
LDII	Length of the two phalanx of the major digit
Lw_{sk}	Skeletal wing length, calculated as summation of *HL*, *UL*, *CmL*, and *LDII*
L_{prim}	Length of the longest primary feather
Lw	Wing length from the shoulder to the wingtip
*Lw**	Wingspan proxy, calculated as summation of *HL*, *UL*, *CmL*, *LDII* and L_{prim}
c	Wing chord at wrist level (proxy of the mean chord)
Sw	One-wing area
B	Wing span, from one to the other wingtip
S_L	Full wing area or lift surface
WF	Interclavicular distance, from the tip of one furcular ramus to the other

Table 2. Ordinary least-squares single regressions (SR) with the best correlated *X*-variables (*; i.e. higher Pearson coefficient), the best multiple regressions (MR), and the principal components regressions (PCR), derived to predict *B* and S_L. |%MPE|, mean percentage prediction error calculated over the whole data set; $|\%MPE|_{VS\text{-}25}$, mean percentage prediction error calculated over the validating subsample of 25% species from the dataset; $|\%MPE|_{VS\text{-}GA}$, mean percentage prediction error calculated using Galloansarae as validating subsample; VIF, factor of variance inflation, $1/(1 - R^2)$.

Y-variable	Reg. Model	N	R^2_{adj}	|%MPE|	$\|\%MPE\|_{VS\text{-}25\%}$	$\|\%MPE\|_{VS\text{-}GA}$	intercept *a*	X-variables	slope *b*	VIF
B	Simple (SR)	182	0.948	8.6	7.8	6.7	1.464	logLCMC*	0.9	1.0
	Multiple (MR)	141	0.995	2.6	3.4	3.3	0.953			
								logLH	0.244	33.3
								logLU	0.082	29.5
								logLCMC	0.208	20.5
								logfprim	0.455	5.0
	Principal components (PCR)	143	0.992	2.9	3.4	3.4	2.825			
								PC1	0.211	1.0
								PC2	0.017	1.0
								PC4	0.006	1.0
SL	Simple (SR)	188	0.965	14.2	15.7	17.1	-3.073	logedH*	1.901	1.0
	Multiple (MR)	140	0.992	6.4	8.5	5.1	-4.694			
								logLH	0.227	27.7
								logedHDV	0.420	33.1
								logLU	-0.796	592.3
								logLR	0.893	535.4
								logfprim	0.364	7.2
								Cwr mm	0.855	15.5
	Principal components (PCR)	136	0.991	6.6	7.6	5.0	-1.137			
								PC1	0.395	1.0
								PC3	0.058	1.0
								PC2	0.033	1.0
								PC5	-0.018	1.0
								PC6	0.015	1.0
								PC8	0.012	1.0

Table 3. Effects of the taxonomical weighting and the stepwise procedure on the prediction errors of the best multiple regressions derived from the modern birds' dataset. |%MPE|, mean percentage prediction error calculated over the whole data set; $|\%MPE|_{VS\text{-}25}$, mean percentage prediction error calculated over the validating subsample of 25% species from the dataset; $|\%MPE|_{VS\text{-}GA}$, mean percentage prediction error calculated using Galloansarae as validating subsample. (*) indicates that stepwise procedure extracted the same function than the full model.

Model	Weighting taxa	Full model					Automatic stepwise				
		N	R^2_{adj}	|%MPE|	$\|\%MPE\|_{VS\text{-}25\%}$	$\|\%MPE\|_{VS\text{-}GA}$	N	R^2_{adj}	|%MPE|	$\|\%MPE\|_{VS\text{-}25\%}$	$\|\%MPE\|_{VS\text{-}GA}$
RM-B	Unweighed*	141	0.995	2.6	3.4	3.3	141	0.995	2.6	3.4	3.3
	Weighed by families	35	0.993	2.7	3.7	2.7	36	0.994	2.8	4.0	2.7
	Weighed by species*	64	0.995	2.6	3.6	2.7	64	0.995	2.6	3.6	2.7
MR-SL	Unweighed	136	0.991	5.9	7.2	6.8	140	0.992	6.4	8.5	5.1
	Weighed by families	34	0.978	6.0	8.2	11.9	36	0.990	6.4	7.3	6.5
	Weighed by species	61	0.989	5.9	7.1	7.8	67	0.991	6.3	8.1	6.1

Table 4. Ordinary least-squares single regressions obtained for each predictor variable against wing span (*B*), using the means of species calculated from raw data and the appropriate phylogenetically independent contrasts (PIC's). PIC's assume a regression through the origin ($a = 0$). Statistical significance of the regressions was checked by permutation tests with 10000 replicates. 95% bootstrapped confidence intervals of the slopes are also shown. Slope difference significance between OLS-raw data and OLS-PICs were tested using t-Student ($\alpha = 0.05$). For abbreviations of variables, see Table 1.

X-variable	Raw data (log *B* = log *a* + *b* log *X*)							PICs (log *B* = *b* log *X*)						Slopes comparison (*p*)
	N	log *a*	*b*		interv. conf. *b*		R^2	N	*b*		interv. conf. B		R^2	
log *LH*	71	1.487	0.770	***	0.725	0.820	0.930	70	0.811	***	0.752	0.871	0.916	ns
log *LU*	71	1.320	0.841	***	0.800	0.887	0.946	70	0.842	***	0.790	0.892	0.938	ns
log *LCMC*	69	1.458	0.903	***	0.872	0.939	0.946	68	0.916	***	0.865	0.970	0.935	ns
log L_{prim}	67	0.560	1.047	***	0.963	1.135	0.893	66	1.023	***	0.935	1.107	0.888	ns

*** $p < 0.0001$

ns - non significant ($p > 0.05$)

Table 5. Ordinary least-squares single regressions obtained for each predictor variable against wing area (S_L), using the means of species calculated from raw data and the appropriate phylogenetically independent contrasts (PIC's). PIC's assume a regression through the origin ($a = 0$). Statistical significance of the regressions was checked by permutation tests with 10000 replicates. 95% bootstrapped confidence intervals of the slopes are also shown. Slope difference significance between OLS-raw data and OLS-PICs were tested using t-Student ($\alpha = 0.05$). For abbreviations of variables, see Table 1.

X-variable	Raw data (log S_L = log a + b log X)							PICs (log S_L = b log X)						slopes comparison (p)
	N	log a	b		interv. conf. b		R^2	N	b		interv. conf. B		R^2	
log *HL*	71	-3.558	1.374	***	1.251	1.494	0.877	70	1.563	***	1.453	1.688	0.889	*
log *dpL*	71	-2.661	1.262	***	1.161	1.376	0.895	70	1.293	***	1.117	1.547	0.828	ns
log *bcL*	71	-2.492	1.458	***	1.336	1.578	0.885	70	1.560	***	1.440	1.723	0.851	ns
log *deHW*	71	-3.056	1.889	***	1.805	1.959	0.966	70	1.840	***	1.761	1.911	0.948	ns
log *dHW*	71	-2.328	1.858	***	1.758	1.944	0.946	70	1.831	***	1.727	1.930	0.933	ns
log *UL*	71	-3.883	1.516	***	1.420	1.619	0.910	70	1.622	***	1.521	1.727	0.910	ns
log *peUW*	71	-2.733	1.931	***	1.772	2.090	0.886	70	1.768	***	1.572	2.035	0.839	ns
log *deUW*	71	-2.697	1.892	***	1.757	2.027	0.923	70	1.871	***	1.740	2.018	0.903	ns
log *dUW*	71	-2.193	1.882	***	1.751	1.751	0.908	70	1.776	***	1.551	2.099	0.846	ns
log *RL*	69	-3.723	1.453	***	1.357	1.562	0.894	68	1.590	***	1.480	1.706	0.909	ns
log *dRW*	69	-1.502	1.292	***	1.145	1.422	0.841	68	1.339	***	1.107	1.622	0.739	ns
log *CmL*	69	-3.598	1.604	***	1.511	1.718	0.884	68	1.745	***	1.620	1.873	0.897	ns
log *dCmW*	70	-1.921	1.981	***	1.825	2.116	0.872	69	1.855	***	1.669	2.062	0.836	ns
log *DCmW*	68	-2.687	2.056	***	1.922	2.187	0.923	67	2.048	***	1.891	2.225	0.898	ns
log L_{prim}	66	-5.346	1.929	***	1.756	2.085	0.880	65	1.952	***	1.739	2.148	0.836	ns
log *c*	71	-5.446	2.057	***	1.950	2.167	0.952	70	1.900	***	1.792	2.028	0.937	*

*** $p < 0.0001$

* $0.01 < p < 0.05$

ns - non significant ($p > 0.05$)

Table 6. Ordinary least-squares multiple regressions derived for predicting wing span (*B*) in avian taxa. For abbreviations of variables, see Table 1.

Model	$\log B = \log a + b_1 \log X_1 + b_2 \log X_2 \ldots + b_p \log X_p$	N	R^2_{adj}	|%MPE|	$\|\%MPE\|_{VS\text{-}25\%}$	$\|\%MPE\|_{VS\text{-}GA}$
MR-B1	$log\ B = 0{,}953 + 0{,}244\ log\ HL + 0{,}082\ log\ UL + 0{,}208\ log\ CmL + 0{,}455\ log\ L_{prim}$	141	0.995	2.6	3.4	3.3
MR-B2	$log\ B = 1{,}318 - 0{,}064\ log\ HL + 0{,}528\ log\ UL + 0{,}443\ log\ CmL$	164	0.977	5.3	7.1	3.8
MR-B3	$log\ B = 1{,}322 - 0{,}267\ log\ HL + 0{,}580\ log\ UL$	172	0.96	7.4	7.6	4.7

N = number of individuals.

R^2_{adj} = coefficient of determination adjusted for the number of predictors included in the model

|%MPE| = mean percent prediction error calculated over the whole dataset.

$|\%MPE|_{VS\text{-}25}$ = mean percent prediction error calculated over the validating subsample (25% species).

$|\%MPE|_{VS\text{-}GA}$ = mean percentage prediction error calculated using Galloanserae as validating subsample.

Table 7. Ordinary least-squares multiple regressions used in the procedure for selection of variables. Three functions to predict S_L were obtained from different combinations of the anatomical predictors (MR Sel.1-3). The S_L values obtained in these functions were plotted against each predictor variable to test and select those for specific stem taxa (Fig. 4 and Table 8).

MR model	log S_L = log a + b1 log X1 + b2 log X2…+ bp log Xp	N	R^2	Tested variables
Sel.1	log S_L = -3,815 + 0,798 log *dHW* + 0,5 log *RL* + 0,584 log *Lprim*	142	0,975	*HL, dpL, bcL, UL, peUW, deUW, dUW, dRW, CmL, dCmW, DCmW, c*
Sel.2	log S_L = -3,982 + 0,827 log *UL* + 0,469 log *dUW* - 0,277 log *CmL* + 0,355 log *DCmW* + 0,570 log *Lprim*	141	0,973	*dHW, deHW, RL*
Sel.3	log S_L = -3,517 + 0,637 log *dHW* + 0,824 log *UL* - 0,26 log *dRW* - 0,330 log *peUW* + 0,162 log *dpL*	159	0,972	*Lprim*

Table 8. Forelimb variables discarded after the selection procedure for predicting S_L on each avian taxa.

Taxa	Excluded variables
Archaeopterygidae	*HL, dpL, peUW, RL*
Jeholornithidae	*dpL, dHW, dRW, DCmW*
Sapeornithidae	*dpL, dRW*
Confuciusornithidae	*dpL, cbL, dCmW, UL*
Enantiornithes	*dpL, dHW, dUW, dRW*
Stem Ornithuromorpha	*dpL, dHW, dUW, dRW*

Table 9. Ordinary least-squares multiple regressions obtained for predicting wing area (S_L) in each avian taxa. For abbreviations of variables, see Table 1. For abbreviations of fossil taxa, see legend of Fig. 3.

Model	$\log S_L = \log a + b_1 \log X_1 + b_2 \log X_2 \ldots + b_p \log X_p$	N	R^2_{adj}	|%MPE|	$\|\%MPE\|_{VS\text{-}25\%}$	$\|\%MPE\|_{VS\text{-}GA}$
MR-S_L Arch1	log SL = – 4,879 + 0,335 log LU + 0,258 log edU + 0,193 log LCMC – 0,198 log dCMC + 0,201 log DCMC + 0,285 log Lprim + 0,923 log c	139	0.993	6.1	6.4	5.7
MR-S_L Arch2	log SL = – 3,905 + 0,54 log dH + 0,6 log LU + 0,251 log DCMC + 0,525 log Lprim	136	0.978	11	11.2	9.3
MR-S_L Arch3	log SL = – 3,2 + 0,968 log dH + 0,798 log LU	173	0.965	14.2	16.2	23.3
MR-S_L Jeho1 = Conf = Ornph2	log SL = – 3,883 + 1,096 log edH + 0,357 log LR + 0,458 log Lprim	144	0.985	8.8	14.5	6.8
MR-S_L Jeho2 = Orph4	log SL = – 3,284 – 0,695 log LH + 1,545 log LU + 1,032 log dU	172	0.963	14.1	13.7	6.3
MR-S_L Sape1 = Ornph1	log SL = – 4,771 + 0,259 log LH + 0,342 log edH – 0,63 log LU + 0,709 log LR + 0,391 log Lprim + 0,888 log c	135	0.994	5.8	7.2	5.1
MR-S_L Sape2	log SL = – 3,436 + 1,275 log edH + 0,556 log LU	176	0.977	11.2	13.6	13.8
MR-S_L Enan1	log SL = – 4,787 + 0,285 log edH + 0,271 log LU + 0,275 log LCMC + 0,207 log DCMC – 0,171 log dCMC + 0,283 log Lprim + 0,832 log c	141	0.992	6.2	7	6.0
MR-S_L Enan2	log SL = – 3,963 + 1,044 log edH + 0,397 log LU + 0,48 log Lprim	145	0.985	8.8	10	7.0
MR-S_L Enan3 = Ornph3	log SL = – 3,459 + 1,254 log edH - 0,382 log epU + 0,598 log LU + 0,375 log DCMC	167	0.98	10.5	14.4	11.6
MR-S_L Enan4	log SL = – 3,507 + 0,889 log LU + 1,015 log DCMC	167	0.968	13.1	15.1	11.7

N = number of individuals.

R^2_{adj} = coefficient of determination adjusted for the number of predictors included in the model

|%MPE| = mean percent prediction error calculated over the whole dataset.

$|\%MPE|_{VS\text{-}25}$ = mean percent prediction error calculated over the validating subsample (25% species).

$|\%MPE|_{VS\text{-}GA}$ = mean percentage prediction error calculated using Galloanserae as validating subsample.

Table 10. Effect of different aerial locomotion on the mean prediction error (%PE) values of the best SR equation and the MR models obtained. Flying mode abbreviations: CF, continuous flapping; FG, flap and gliding; FB, flap and bounding; S, soaring. For abbreviations of variables, see Table 1. When Levene's test (Lev.) was non-significant, the difference between groups was tested by a one-way ANOVA. In other case, Welch's test was performed. **A.** Predictive models for wingspan (*B*). **B.** Predictive models for wing area (S_L). For abbreviations of fossil taxa, see legend of Fig. 3.

A.

Model	Lev.	ANOVA/Welch		CF			FG			FB			S			Total		
	p	F	*p*	N	%PE	Sd	N	%PE	Sd	N	%PE	Sd	N	%PE	Sd	N	%PE	Sd
SR-*UL*	**	3,195	*	68	1,1	13,4	41	4	14,8	40	-3,5	6,4	30	0,6	9,4	179	0,7	12,1
RM-B1	ns	1,128	ns	58	-0,5	3,6	32	0,9	3,9	26	-0,4	3,0	25	-0,3	3,2	141	-0,1	3,5
RM-B2	ns	3,991	*	63	-2,0	8,0	37	3,1	7,1	34	0,8	6,2	30	0,6	7,4	164	0,2	7,5
RM-B3	*	3,262	*	67	-1,3	11,8	39	3,7	9,0	36	-1,5	6,2	30	1,3	9,6	172	0,2	10,0

B.

Model	Lev.	ANOVA/Welch		CF			FG			FB			S			Total		
	p	F	*p*	N	%PE	Sd	N	%PE	Sd	N	%PE	Sd	N	%PE	Sd	N	%PE	Sd
SR-deHW	*	6,431	**	72	2,9	20,5	42	5	20,4	43	-5,4	10,2	31	3,4	18,1	188	1,6	18,5
RM-S_L Arch1	ns	0,470	ns	56	0,9	7,6	30	-0,4	7,7	27	1,0	8,0	26	-0,8	8,1	139	0,3	7,7
RM-S_L Arch2	ns	8,344	**	56	2,0	14,5	29	-5,8	9,7	25	10,7	12,3	26	-3,1	13,0	136	0,9	13,9
RM-S_L Arch3	**	3,447	*	65	-0,2	21,4	41	-1,4	11,5	37	8,7	16,3	30	-0,2	17,2	173	1,4	17,9
RM-S_L Jeho1 = RM-S_L Conf = RM-S_L Ornph2	ns	0,571	ns	55	0,3	7,8	30	1,2	7,2	24	0,6	5,7	26	-1,3	6,9	135	0,3	7,1
RM-S_L Jeho2 = RM-S_L Orph4	*	,049	ns	65	1,4	17,4	42	0,8	12,1	40	0,5	10,4	29	0,3	13,8	176	0,9	14,1
RM-S_L Sape1 = RM-S_L Ornph1	*	3,434	*	58	3,4	13,5	32	-2,7	9,1	29	2,4	7,7	25	-2,7	11,0	144	0,8	11,4
RM-S_L Sape2	ns	1,542	ns	56	1,9	8,0	30	-0,7	8,0	28	0,4	8,5	27	-1,9	7,8	141	0,3	8,1

RM-S_L Enan1	*	3,863	*	57	3,7	13,5	33	-3,0	9,1	29	2,1	7,7	26	-3,0	11,3	145	0,6	11,5
RM-S_L Enan2	ns	1,802	ns	64	0,5	15,1	36	-2,4	10,6	38	4,5	10,8	29	1,5	13,2	167	0,9	13,1
RM-S_L Enan3 = RM-S_L Ornph3	**	9,851	**	64	-0,2	16,4	36	-5,9	9,5	38	11,6	17,6	29	1,3	16,9	167	1,5	16,6
RM-S_L Enan4	*	1,495	ns	66	3,6	22,1	41	-1,4	13,5	36	4,2	16,0	29	-2,0	17,1	172	1,6	18,3

** $p < 0.01$; * $0.05 > p > 0.01$; ns = $p > 0.05$.

Sd indicates the standard deviations for the mean values.

Table 11. Calibration of the cumulative prediction error on flight parameters. **A.** Prediction error in the calculation of *WL*, obtained from values estimated with MR models derived for predicting *BM* (in Serrano et al. 2015) and S_L in the present work (Table 9). **B.** Prediction error in the calculation of *AR*, obtained from values estimated with MR models derived for predicting *B* (Table 6) and S_L (Table 9).

A.

Models to estimate body mass (*BM*)		Models to estimate lift surface (S_L)		Wing loading (*WL*)			
Model	|%MPE|	Model	|%MPE|	N	|%MPE|	$\|\%MPE\|_{VS\text{-}25\%}$	$\|\%MPE\|_{VS\text{-}GA}$
MR-Arch	18.8	MR-SL Arch1	6.1	122	17.3	15.4	16.4
		MR-SL Arch2	11.0	122	19.1	18.2	18.0
		MR-SL Arch3	14.2	152	21.3	24.9	38.5
MR-Jeho	17.3	MR-SL Jeh1	8.8	127	19.2	21.4	17.4
		MR-SL Jeh2	14.1	147	20.0	26.9	15.6
MR-Sape	16.5	MR-SL Sap1	5.8	118	15.2	17.5	16.0
		MR-SL Sap2	11.2	142	18.5	25.0	24.9
MR-Conf	16.9	MR-SL Conf1	8.8	123	19.4	19.2	22.6
MR-Enan	19.4	MR-SL Enan1	6.2	128	18.5	22.1	13.1
		MR-SL Enan2	8.8	132	19.3	24.5	16.6
		MR-SL Enan3	10.5	151	20.7	30.8	25.1
		MR-SL Enan4	13.1	151	21.5	35.0	17.9
MR-Orph	19.4	MR-SL Orph1	5.8	129	18.0	17.6	23.5
		MR-SL Orph2	8.8	153	21.2	19.8	23.8
		MR-SL Orph3	10.5	155	21.0	25.7	30.1
		MR-SL Orph4	14.1	165	23.1	25.8	21.5
MR-Ornl	20.2	MR-SL Enan3 - Orph3	10.5	159	20.2	30.5	23.3
		MR-SL Orph4	14.1	159	22.5	25.8	21.5

B.

Models to estimate wing span (B)		Models to estimate lift surface (S_L)		Aspect ratio (AR) prediction errors			
Model	|%MPE|	Model	|%MPE|	N	|%MPE|	$\|\%MPE\|_{VS\text{-}25\%}$	$\|\%MPE\|_{VS\text{-}GA}$
MR-B1	2.6	MR-SL Arch1	6.1	133	7.5	7.0	7.9
		MR-SL Arch2	11.0	133	12.6	10.9	13.2
		MR-SL Jeh1, Conf, Orph2	8.8	137	10.6	10.8	11.6
		MR-SL Sap1, Orph1	5.8	133	7.0	8.6	9.5
		MR-SL Enan1	6.2	135	7.7	7.2	7.5
		MR-SL Enan2	8.8	139	10.5	12.0	11.0
		MR-SL Enan3, Orph3	10.5	134	12.4	17.5	21.9
MR-B2	5.3	MR-SL Arch3	14.2	163	13.7	16.8	25.3
		MR-SL Sap2	11.2	162	11.5	16.5	13.6
		MR-SL Enan3, Orph3	10.5	157	12.1	17.0	13.8
		MR-SL Enan4	13.1	157	14.7	14.1	14.6
MR-B3	7.4	MR-SL Arch3	14.2	170	16.8	18.7	24.4
		MR-SL Jeh2, Orph4	14.1	170	17.8	14.4	12.2

Table 12. Estimates of the basic and compound variables for the specimens of Mesozoic stem taxa. The multiple regression model from which wingspan (*B*) and wing area (S_L) were obtained, are also indicated (MR-B and MR-S_L, from Tables 6 and 9). The body mass values (*BM*) obtained in Serrano et al. (2015) are also included for a summary. The prediction errors (% MPEs) of wing loading (WL) and aspect ratio (*AR*) calculated on the modern dataset, are also shown for different combinations of *B*, S_L, and *BM* estimates. Prediction intervals were calculated as predicted values ± |%MPE| values.

Taxa	Specie	Museum code	BM (g)	B (mm)	B ± |%MPE|		MR-B	SL (cm2)	SL ± |%MPE|		MR-SL	WL (g/cm2)	WL ± |%MPE|		AR	AR ± |%MPE|	
Arch	Archaeopteryx lithographica	BMNH 37001	450	766	823	709	B3	740	845	635	Arch3	0,609	0,738	0,479	7,9	9,3	6,6
	A. lithographica	HMN 1880/1881	343	598	614	583	B1	601	637	564	Arch1	0,572	0,671	0,473	6,0	6,4	5,5
	A. lithographica	JM 2257	176	400	410	390	B1	232	257	206	Arch2	0,759	0,903	0,614	6,9	7,8	6,0
	A. lithographica	BMMS 500	1088	784	826	743	B2	1238	1414	1063	Arch3	0,879	1,066	0,692	5,0	5,6	4,3
	A. lithographica	WDC-CSG-100	264	549	563	535	B1	401	446	357	Arch2	0,657	0,782	0,532	7,5	8,5	6,6
Jeho	Jeholornis prima	V 13274	1516	1094	1175	1013	B3	2172	2477	1866	Jeh2	0,698	0,838	0,559	5,5	6,5	4,5
	J. prima	V 13353	983	1027	1054	1000	B1	1641	1784	1497	Jeh1	0,599	0,714	0,484	6,4	7,1	5,7
Sape	Sapeornis chaoyangensis	V 12698	900	1264	1331	1197	B2	2041	2269	1813	Sap2	0,441	0,523	0,359	7,8	8,8	6,8
	S. chaoyangensis	V 13275	1041	1186	1248	1123	B2	1893	2104	1682	Sap2	0,550	0,652	0,448	7,4	8,4	6,5
	S. chaoyangensis	V 13276	1038	1188	1250	1125	B2	1979	2200	1758	Sap2	0,524	0,621	0,427	7,1	8,0	6,2
	S. chaoyangensis	DNHM 3078	491	912	935	888	B1	1305	1381	1229	Sap1	0,376	0,434	0,319	6,4	6,8	5,9
Conf	Eoconfuciusornis zhenghi	V 11977	217	450	462	438	B1	244	265	222	Conf1	0,893	1,065	0,720	8,3	9,2	7,4
	Confuciusornis sanctus	V 11374	520	668	685	651	B1	818	890	746	Conf1	0,636	0,759	0,513	5,5	6,0	4,9
	C. sanctus	V 11640	762	786	806	766	B1	938	1020	856	Conf1	0,813	0,970	0,656	6,6	7,3	5,9
	C. sanctus	MCFO-0374	550	674	692	656	B1	669	727	610	Conf1	0,823	0,982	0,664	6,8	7,5	6,1
Enan	Pengornis houi	V 15336	437	695	731	658	B2	1008	1114	902	Enan3	0,433	0,523	0,343	4,8	5,4	4,2
	Cathayornis yandica	V 9169	63	309	326	293	B2	199	220	178	Enan3	0,318	0,384	0,252	4,8	5,4	4,2
	Cuspirostrisornis houi	V 10897	70	330	347	312	B2	169	187	151	Enan3	0,412	0,497	0,326	6,4	7,2	5,6
	Eocathayornis walkeri	V 10916	56	286	301	271	B2	193	213	173	Enan3	0,290	0,349	0,231	4,2	4,7	3,7
	Eoenantiornis buhleri	V 11537	80	317	325	309	B1	148	161	135	Enan2	0,541	0,645	0,437	6,8	7,5	6,1
	Longchengornis sanyanensis	V 10530	86	335	353	318	B2	167	185	150	Enan3	0,518	0,625	0,410	6,7	7,5	5,9
	Longirostravis hani	V 11309	39	318	326	310	B1	164	174	154	Enan1	0,240	0,284	0,195	6,2	6,6	5,7
	Otogornis genghisi	V 9607	171	432	455	409	B2	291	322	261	Enan3	0,588	0,707	0,469	6,4	7,2	5,6
	Longipteryx chaoyangensis	V 12325	193	463	487	438	B2	373	412	334	Enan3	0,517	0,624	0,410	5,7	6,4	5,0
	Rapaxavis pani	DNHM D2522	47	238	251	225	B2	141	156	126	Enan3	0,334	0,403	0,265	4,0	4,5	3,5
	Protopteryx fengningensis	V 11665	70	316	324	308	B1	119	130	109	Enan2	0,589	0,702	0,475	8,4	9,3	7,5
	Vescornis hebeiensis	NIGPAS 130722	50	272	287	258	B2	126	140	113	Enan3	0,394	0,475	0,312	5,9	6,6	5,2
	Concornis lacustris	LH-2814	70	342	360	324	B2	202	228	176	Enan4	0,347	0,421	0,272	5,8	6,6	4,9
	Eoalulavis hoyasi	LH-13500	45	262	276	248	B2	160	177	143	Enan3	0,279	0,336	0,223	4,3	4,8	3,8
Orph	Archaeorhynchus spathula	V 14287	272	550	564	536	B1	479	521	437	Orph2	0,568	0,689	0,448	6,3	7,0	5,6
	Yixianornis grabaui	V 12631	321	522	536	509	B1	440	479	402	Orph2	0,729	0,884	0,575	6,2	6,9	5,5

Zhongjianornis yangi	V 15900	570	724	763	686	B2	825	912	738	Orph3	0,691	0,835	0,546	6,4	7,1	5,6
Yanornis martini	V 12558	772	773	814	732	B2	1079	1192	966	Orph3	0,715	0,865	0,565	5,5	6,2	4,9
Y. martini	V 10996	618	820	863	776	B2	1154	1276	1033	Orph3	0,535	0,648	0,423	5,8	6,5	5,1
Jianchangornis microdonta	V 16708	820	822	844	801	B1	980	1083	877	Orph3	0,837	1,013	0,661	6,9	7,7	6,1
Hongshanornis longicresta	V 14533	88	350	359	341	B1	163	173	154	Orph1	0,540	0,637	0,443	7,5	8,0	7,0
Longicrusavis houi	PKUP V1069	89	299	315	283	B2	176	195	158	Orph3	0,504	0,611	0,397	5,1	5,7	4,5
Gansus yumenensis	CAGS 04-CM003	156	537	565	508	B2	392	433	351	Orph3	0,399	0,482	0,315	7,3	8,2	6,5
G. yumenensis	CAGS 04-CM004	162	515	543	488	B2	322	355	288	Orph3	0,505	0,607	0,403	8,3	9,3	7,3
Ichthyornis dispar	YPM 1450	284	682	732	631	B3	587	669	587	Orph4	0,423	0,520	0,325	7,9	9,3	6,5
Apsaravis ukhaana	IGM 100/1017	184	546	586	505	B3	298	339	298	Orph4	0,618	0,758	0,479	10,0	11,8	8,2

Table 13. Comparison between the estimates for Berlin's *Archaeopteryx* and for *Hongshanornis* (two specimens) obtained using MR models (first line of each taxa) and outline body reconstructions. Superscript numbers indicate the method for obtaining *BM* values used in each study. The percent variation between values from published reconstruction and the values obtained at present work is also shown.

Reconstruction´s reference	*BM* (g)	Δ*BM* (%)	*B* (mm)	Δ*B* (%)	S_L (cm^2)	ΔS_L (%)	*WL* (g/cm^2)	Δ*WL* (%)	*AR*	Δ*AR* (%)
Archaeopteryx HMN 1880	343[1]	-	598	-	601	-	0,572	-	6,0	-
Yalden (1971)	-	-	588	1,7	479	20,2	-	-	7,2	-21,1
Yalden (1984)	271[2]	21,0	545	8,9	479	20,2	0,566	1,0	6,2	-4,1
Elzanowski (2002)	276[3]	19,5	550	8,0	500	16,7	0,552	3,4	6,1	-1,5
Longrich (2006)	250[4]	27,1	523	12,5	546	9,1	0,458	19,9	5,0	16,0
Wellnhofer (2008)	250[4]	27,1	580	3,0	611	-1,7	0,409	28,4	5,5	7,6
Hongshanornis (V 14533)	88[5]	-	350	-	163	-	0,540	-	7,5	-
Chiappe et al. (2014) (D2945/6)	65[6]	26,3	320	8,6	160	2,1	0,406	24,7	6,4	7,5

[1] Multivariate allometric equation using 6 skeletal variables (ARCH), from Serrano et al. (2015)
[2] Volumetric displacement
[3] Univariate allometric equation using femoral width at midshaft (*dFW*), from Campbell and Marcus (1992)
[4] Univariate allometric equation using *dFW*, from Yalden (1984)
[5] Multivariate allometric equation using 7 skeletal variables (ORPH), from Serrano et al. (2015)
[6] Univariate allometric equation using *FL*, from Peters and Peters (2009)