RESEARCH ARTICLE

Aerobic power and flight capacity in birds: a phylogenetic test of the heart-size hypothesis

Roberto F. Nespolo[1,2,*], César González-Lagos[2,3], Jaiber J. Solano-Iguaran[1], Magnus Elfwing[4], Alvaro Garitano-Zavala[5], Santiago Mañosa[6], Juan Carlos Alonso[7] and Jordi Altimiras[4,*]

ABSTRACT

Flight capacity is one of the most important innovations in animal evolution; it only evolved in insects, birds, mammals and the extinct pterodactyls. Given that powered flight represents a demanding aerobic activity, an efficient cardiovascular system is essential for the continuous delivery of oxygen to the pectoral muscles during flight. It is well known that the limiting step in the circulation is stroke volume (the volume of blood pumped from the ventricle to the body during each beat), which is determined by the size of the ventricle. Thus, the fresh mass of the heart represents a simple and repeatable anatomical measure of the aerobic power of an animal. Although several authors have compared heart masses across bird species, a phylogenetic comparative analysis is still lacking. By compiling heart sizes for 915 species and applying several statistical procedures controlling for body size and/or testing for adaptive trends in the dataset (e.g. model selection approaches, phylogenetic generalized linear models), we found that (residuals of) heart size is consistently associated with four categories of flight capacity. In general, our results indicate that species exhibiting continuous hovering flight (i.e. hummingbirds) have substantially larger hearts than other groups, species that use flapping flight and gliding show intermediate values, and that species categorized as poor flyers show the smallest values. Our study reveals that on a broad scale, routine flight modes seem to have shaped the energetic requirements of birds sufficiently to be anatomically detected at the comparative level.

KEY WORDS: Comparative phylogenetics, Cardiovascular system, Stroke volume, Aves, Ornstein–Uhlenbeck models

INTRODUCTION

Aerobic power (i.e. the capacity to endure intense and sustained aerobic activity) is supported by a cascade of processes nested in several organizational levels, which are ultimately constrained by circulatory adjustments (Bernheim et al., 2013; Bishop, 1997; Hillman and Hedrick, 2015; La Gerche et al., 2014). Active lifestyles (e.g. flying) require a comparatively efficient circulatory system that, in endotherms, is characterized by four-chambered hearts, high systolic pressure and high resting metabolism. Compared with other vertebrates, birds have high aerobic capacity, which is frequently interpreted as an adaptation to the energetic burden of flight (Bishop, 2005; Hedenströem, 2008). In this sense, among the many factors that limit aerobic capacity in animals, stroke volume (the volume of blood pumped from the ventricle to the body during each beat) seems to be central (Bishop, 1997; Bishop and Butler, 1995; Hillman and Hedrick, 2015).

Birds are a specialized lineage of theropod dinosaurs that experienced a long evolutionary period as a single clade (ca. 160 million years for Paraves; see Lee et al., 2014; Puttick et al., 2014). During this period, the lineage experienced reductions in body size and diversified into at least 30 orders, subsequently giving rise to great variation in flight modes (Gower, 2001; Hackett et al., 2008; Lee et al., 2014; Puttick et al., 2014). Bird flight ability ranges from non-volant sedentary, species such as rheas and ostriches, to sophisticated fliers, like hummingbirds and swifts. Additionally, there are imperfect flyers that perform short flights, but spend most of their time on the ground, such as tinamous and many galliforms (Viscor and Fuster, 1987). According to some authors, sustained flight capacity is correlated with aerobic power, which in turn seems to be correlated with the size of the heart (at a given body size) (Bishop, 1997, 1999). For example, the heart of a hummingbird is about 3% of body mass, whereas in a pelican this proportion is smaller, about 0.8%. By making formal comparisons of a comprehensive set of species and by taking into account scaling relationships, some authors suggested that hummingbirds have extremely large hearts, whereas poorly flying birds (e.g. galliforms) have smaller hearts (see Bishop, 1997; Hartman, 1955). These patterns suggest that heart size is a strong constraint on flight capacity, but are they the result of an underlying evolutionary process?

A number of phenomena could explain a given comparative pattern. For instance, specialization to a given habitat or lifestyle (e. g. migratory behavior, running or diving capacity) other than flight could entail compensation in aerobic power and/or heart size (e.g. Vágási et al., 2016). Alternatively, adaptive compensation for reducing flight energy costs is commonplace; examples of such compensations include aerodynamic adaptations and other morpho-physiological or behavioral adjustments that increase the efficiency of flight and reduce energy loads, and hence energetic costs (e.g. Alerstam et al., 2007; Hedenströem, 2008).

If heart size is such a strong requisite for the evolution of some energetically demanding flight modes, we should be able to detect this signature above other factors. This can be attained using comparative methods combined with a model selection approach that takes into account phylogenetic relationships (see Materials and methods). Using these methods, we performed the following

[1]Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile. [2]Center of Applied Ecology and Sustainability (CAPES), Facultad de Ciencias Biológicas, Universidad Católica de Chile, Santiago 6513677, Chile. [3]Centro de Investigación en Recursos Naturales y Sustentabilidad (CIRENYS), Universidad Bernardo O'Higgins, Santiago, Chile. [4]AVIAN Behavioural Genomics and Physiology Group, Division of Biology, Department of Physics, Chemistry and Biology, Linköping University, Linköping, Sweden. [5]Instituto de Ecología, Carrera de Biología, Universidad Mayor de San Andrés, La Paz, Bolivia. [6]Departament de Biologia Evolutiva, Ecologia i Ciéncies Ambientals, Institut de Recerca de la Biodiversitat (IRBio), Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain. [7]Departamento de Ecología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC), Madrid, Spain.

*Authors for correspondence (robertonespolorossi@gmail.com, jordi.altimiras@liu.se)

R.F.N., 0000-0003-0825-9618; J.J.S.-I., 0000-0002-1319-8568

analysis: according to what is observed in nature and using predefined criteria for flight classification, an independent observer classified flight into categories ranging from worst flight capacity to best flight capacity. Central to our reasoning is the fact that flight categories should be independent of any physiological factor underlying aerobic power. Assuming (i) that aerobic power is inextricably linked with heart size, and (ii) that the energetic requirements of different flight modes constitute important selective constraints, we then predicted a distribution of evolutionary optima from worst flight capacity to best flight capacity. Otherwise, the data would be explained by random white noise (our null hypothesis).

We applied a family of analyses based on model selection and information theory specifically designed to contrast evolutionary hypotheses including phylogenetic, non-phylogenetic or purely random trait distributions models (see Materials and methods). We compiled anatomical data (heart mass, body mass; discussed in Materials and methods) for several species and codified flight mode according to a previously defined criterion (Viscor and Fuster, 1987). These compilations (915 species; the complete dataset will be available online; see Materials and methods) and the analyses provided in this study support the hypothesis that the size of the heart (adjusted to body size and phylogeny) evolved toward optimal values that coincide with the preferred flight mode of birds (Bishop, 1997; Hartman, 1955).

MATERIALS AND METHODS
Data
We compiled data from 915 species, by conducting exhaustive literature searches in numerous databases (Scopus, Google scholar, Web of Science, Zoological Record). Studies were only considered if heart and body mass data were available for adult birds. For some species, we also included our own unpublished data (for example, we had previously collected data for two species of bustards, four species of tinamous, and red junglefowl). Heart masses included the fresh mass in grams of both ventricles and atria after dissection of the outflow tract and removal of blood clots (to the nearest 0.01 g). Mass was obtained from fresh or frozen carcasses. Average values per species and sex were obtained either from the published article or by calculation after requesting the original data from the authors. When values from different conditions were provided in the original publication (season, altitude or experimental acclimatization), an average value was calculated. The full database with individualized entries from the different studies is available from Nespolo et al. (2017b).

Flight mode classification
A complete description of flight mode classification can be found in Nespolo et al. (2017b). Initially, all species were grouped into five flight modes: no flight, short flight, flapping flight, gliding and soaring, and hovering flight using the criteria of Viscor and Fuster (1987) and Videler (2005). Because only five species were classified as 'no flight' and two of them, the ostrich (*Struthio camelus*) and the emu (*Dromaius novaehollandiae*), were removed as outliers according to Cook's *D* distances, the no flight mode was removed from the analysis. In any event, including these species gave similar results. In order to check whether changing the selection criteria for character state classification generated different results (i.e. a 'sensitivity' analysis), we generated three additional datasets: one where conflicting species were 'upgraded' (i.e. shifted to a higher character state, 132 species), another where conflicting species were 'downgraded' (i.e. shifted to a lower character state, 87 species), and a third that included both cases (219 species). In no case were the results different to what is presented here.

Phylogenetic comparative analyses
We used a calibrated phylogenetic tree of birds that includes over 9000 species (Jetz et al., 2012); this tree was generated by an automated provider of 100 phylogenetic trees (http://birdtree.org/). We initially transformed body mass and heart mass to $\log_{10}$, then we controlled for variation in body mass by: (1) using residuals from ordinary least-squares regressions of heart mass and body mass; (2) using residuals from phylogenetic linear regressions (i.e. generalized least squares, assuming a covariance structure where internal branch lengths in the variance–covariance phylogenetic matrix are multiplied by a constant (lambda) (corPagel option, for the gls command in nlme and ape) (Martins and Hansen, 1997); or (3) using the ratio of heart mass to body mass. Given that these approaches gave similar final results, here we only present the residuals from phylogenetic linear regressions. To account for potential effects of multiple measurements per species, we repeated all analyses using the median by species instead of the mean (the results did not change).

We performed two types of phylogenetic comparative analyses. In order to explore and visualize whether the patterns of trait diversification adjust to different models, we used both Brownian motion (BM) and Hansen's (1997) multiple optimum Ornstein–Uhlenbeck (OU) models, which are described in detail elsewhere (Butler and King, 2004; Hansen, 1997). Briefly, the BM model, as first described by Felsenstein (1973), is:

$$\mathrm{d}X(t) = \sigma \mathrm{d}B(t), \tag{1}$$

where $\mathrm{d}X(t)$ represents the change in mean trait value of a given lineage, σ represents the noise parameter (i.e. the rate of increase in the variance of trait values over time) and $\mathrm{d}B(t)$ represents a sample of the Brownian process. This process predicts a monotonic increase in trait variance over time. Selection was incorporated into this model by adding the term $\alpha[\theta - X(t)]\mathrm{d}t$, according to the OU model (Butler and King, 2004; Hansen, 1997), for which BM is a particular case (when α=0):

$$\mathrm{d}X(t) = \alpha[\theta - X(t)]\mathrm{d}t + \sigma \mathrm{d}B(t). \tag{2}$$

This model, a multiple-optimum OU process describing the evolution of a continuous trait subject to selection and Brownian motion, has two additional parameters: θ, which represents an evolutionary optimum that acts as an attractor of trait values, and α, representing the strength of selection 'pulling' to the optimum (Butler and King, 2004). Importantly, the OU model allows the optimal trait value (θ) to vary along the branches of the phylogenetic tree to represent changes in selective regimes ('adaptive zone'; *sensu* Simpson, 1953) of the lineages. We worked with these models to propose a few *a priori* hypotheses, representing alternative explanations for the observed pattern of trait evolution. It is important to note that these evolutionary hypotheses must be specified *a priori* to be statistically valid and test the importance of particular evolutionary factors, and that they have more power if the alternatives are fewer in number.

Previous to any analysis, we compared a phylogenetic model with a non-phylogenetic model using the fitContinuous command in geiger, and evaluated whether a phylogenetic model is actually a better description of the data. This was performed by comparing a 'white-noise' model (this is equivalent to a 'star' phylogeny; e.g. Dlugosz et al., 2013; Spoor et al., 2007) with a series of phylogenetic models including BM and OU (see Nespolo et al., 2017a, for details of the models). For the specific question of whether flight modes are associated with different evolutionary

optima, we used the OUwie package (Beaulieu et al., 2012). Using this procedure, we fitted a BM model (a Brownian motion model), an OU1 model (a model assuming one optimum), a BMS model (a model that assumes Brownian motion with different rates according to the selective regime meaning that flight mode influences the rate of heart mass evolution but not to an optimum) and an OUM model (a Brownian motion model assuming selection towards different optima according to each flight mode).

The selection of the best model was performed using Akaike information criteria (AICc and AIC weights; Burnham and Anderson, 2002). All statistical procedures were performed using the R platform (http://www.R-project.org/). In order to visualize how the different flight modes were distributed in our working phylogenetic tree, we used stochastic character mapping (Huelsenbeck et al., 2003). From this, we generated 1000 maps according to the procedure detailed in Price and Hopkins (2015). We checked model performance, reliability of the parameter estimates and model likelihood by evaluating the model eigenvalues, which should be positive. Because difficulties in estimating the parameters of some models can lead to problematic inference, inflated standard errors around mean parameter estimates and negative Hessian eigenvalues (Beaulieu et al., 2012), models were not considered if parameters could not be estimated.

Finally, we compared the results of the OU models with the output of phylogenetic generalized linear models (caper package). This is a more classic approach that includes scaling effects to body size explicitly in the model, and makes use of phylogenetic information by branch-length transformation according to the phylogenetic signal (lambda) of the data (Freckleton et al., 2002). Given that there was no interaction between log body mass (M_b) and levels of the factor, we compared the intercepts, assuming a common slope (as in an ANCOVA). These results were interpreted as in an ordinary glm (Crawley, 2007), where log heart mass (M_h) was entered as a function of $\log M_b$ and flight mode. We considered the short flight level as the intercept, and all the other levels are expressed as the distance from this value (see Results).

RESULTS

Our compilation covered 28 orders and 103 families of birds. According to the distribution of flight modes provided by the stochastic map, flapping flight is the most common flight mode (Fig. 1), followed by gliding and soaring, short flight and, finally, hovering flight (Fig. 1). However, exceptions were common. For instance, there were a few species that did not exhibit the common flight mode of the group: quails, *Coturnix coturnix*, were coded as flapping flight though the Galliformes are predominantly classified as short flight (hence the blue spot within the orange in Fig. 1). These exceptions represent independent trait acquisitions. Comparing a phylogenetic model with a white-noise model (star phylogeny) indicated that the former better explains the data than the latter (OU model, AICc=−1553.43, weight=1; white-noise model, AICc=−1073.277, weight=0). By comparing the models assuming single or multiple optima, we also found that the model employing multiple optima ranked the highest (OUM model, AICc weight=1.0; Table 1). This is recognized in the phenogram, which shows how residuals of heart size diverged in hovering birds, compared with flapping flight birds and short flight species (Fig. 2). The OUM model assumed one different optimum for each flight mode (Table 2). Visualizing the distribution of residuals and estimated optima in a kernel density plot, it becomes evident that the estimated optima (dotted lines in Fig. 3, from Table 2) coincide with the observed trait distribution (peaks in Fig. 3). However, hovering

flight (the flight mode of Trochilidae; see Fig. 1) was one exception to this trend as the optimum was considerably higher than the mean trait values (Fig. 3). A phylogenetic generalized linear model showed qualitatively similar results (Table 3). That is, short flight (here denoted as the intercept) ranked the lowest, followed by gliding and soaring and flapping flight (which are indistinguishable, given the standard errors) and hovering flight, which shows the highest value for heart size (Table 3).

DISCUSSION

At the resolution level of our analysis, the results support the idea that the energetic burden of flight is pervasive enough to be reflected in the heart size of a broad sample of species, including phylogenetic relationships. We hasten to indicate that the flight modes considered here should be viewed as *a priori* hypotheses, which were associated with competing statistical models that were, in turn, contrasted with the available data. Therefore, the conclusions are restricted by the limits of the dataset and the statistical power of the analysis (see Cressler et al., 2015). Nevertheless, three facts support our conclusions. First, a white-noise model for trait evolution was not selected as the best description of the data, thus suggesting that the phylogeny should be included. Second, results were robust to different combinations of flight categorization, to different statistical control of body size effects and to the removal of conflicting bird groups (e.g. Tinamidae). Finally, in all cases, the model that best described the data is the one supporting the idea that heart size is a good proxy of aerobic power (Bishop, 1997, 2005; Hillman and Hedrick, 2015).

Although it has been known for decades that trained birds and mammals display a functional enlargement of the ventricle mass (i.e. 'athletic heart') (Krautwald-Junghanns et al., 2002; Saltin and Rowell, 1980), this evidence came from experimental studies in single species. Bishop (1997) showed that this morpho-physiological adjustment for high aerobic work is also observed at the interspecific level, and particularly for flying animals. In this study, we complement this information by providing a phylogenetic analysis that includes a graphic mapping of flight modes across the avian clade (the stochastic map; Fig. 1), together with a calibrated phenogram showing how flight mode and residual heart mass diverged about 77 million years ago (Fig. 2) (see also Lee et al., 2014; Puttick et al., 2014). Our analyses suggest the following ranking in the aerobic requirements of flight: poor fliers (i.e. non-flying species and species that mostly take short flights) rank lowest, followed by gliding and soaring birds, then flapping birds (these two, however, showed large standard errors) and finally by hovering birds (the generalized flight mode of hummingbirds, Trochilidae, 32 species in our dataset; see Figs 1 and 2). Our results confirm not only that hummingbirds are strongly constrained by their flight style (Chai and Dudley, 1996; Clark and Dudley, 2010; Fernández et al., 2011) but also that species that are poor fliers also have relatively small hearts, which suggests that the energy burden of flight is relaxed in them (e.g. Wright et al., 2016). In the following paragraphs, we discuss some of these conclusions.

The highly specialized flight mode of hummingbirds (Krebs and Harvey, 1986) involves not only a powerful heart to support the high metabolic needs of such flight but also miniaturization, a compact body and other strategies to minimize energy consumption while resting (e.g. torpor; Carpenter, 1974). Hummingbirds have powerful flight muscles, the capacity for large cardiac output, short circulatory turnover, blood with high oxygen carrying capacity, high capillary surface area, and highly refined pulmonary structural components (Bishop, 1997; Johansen et al., 1987; Maina, 2000; Suarez et al.,
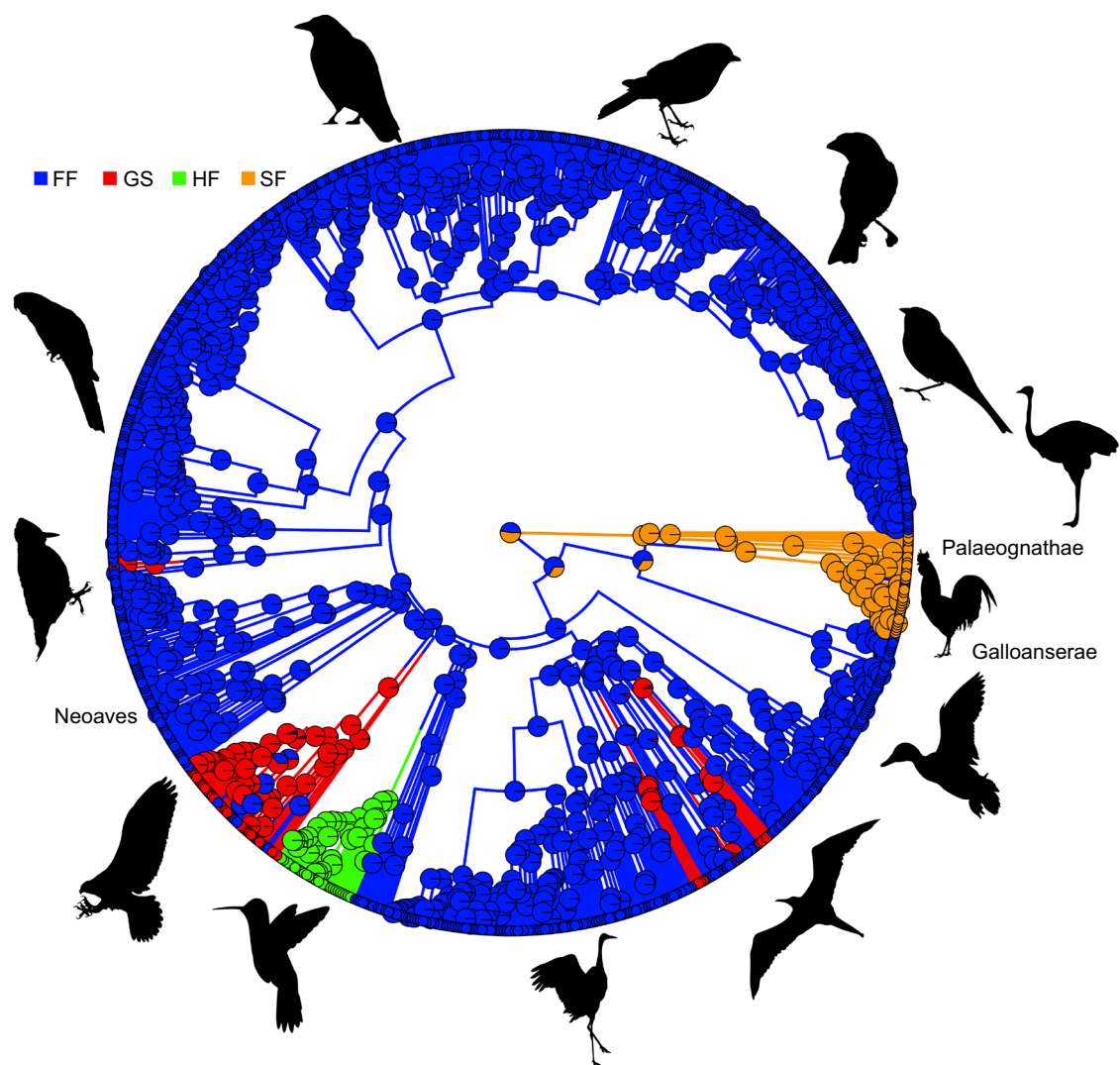
Fig. 1. Summary of the flight mode classification used. Classification is depicted as a stochastic character map, obtained from 1000 simulated trees using the fitER function in phytools (this is one of the 1000 trees obtained). Each circle at each node represents the most likely flight mode of the ancestor, given the data (for instance, the common ancestor of all birds had a ¾ probability of having been a short flyer). Bird silhouettes are from PhyloPic (www.phylopic.org/). *Struthio camelus* by Matt Martyniuk (vectorized by T. Michael Keesey); *Gallus gallus domesticus* (rooster) by Steven Traver; *Anas plathyrynchos* (duck) by Sharon Wegner-Larsen; *Fregata* sp. (frigates) by Thea Boodhoo (photograph) and T. Michael Keesey (vectorization); *Grus canadensis* (sandhill crane) by Sharon Wegner-Larsen; *Pandion haliaetus haliaetus* (osprey) by Steven Traver; *Sphyrapicus varius* (woodpeckers) by Nancy Wyman (photo), John E. McCormack, Michael G. Harvey, Brant C. Faircloth, Nicholas G. Crawford, Travis C. Glenn, Robb T. Brumfield and T. Michael Keesey; *Alisterus scapularis* (parrot) by Michael Scroggie; *Corvus brachyrhynchos* by Peileppe; *Catharus* genus (Turidae) by Sharon Wegner-Larsen; *Serinus* genus (Fringillidae) by Francesco Veronesi (vectorized by T. Michael Keesey); *Emberiza citrinella* (Emberizidae) by L. Shyamal. FF, flapping flight; GS, gliding and soaring; HF, hovering flight; SF, short flight.

Table 1. The model selection approach

	lnL	AICc	dAICc	AICw
BM1	737.52	−1471.03	152.34	0
OU1	769.67	−1523.18	100.20	0
BMS	779.73	−1553.43	69.95	0
OUM	817.73	−1623.38	0.00	1

A comparison of the goodness of fit (based on information theory) of several evolutionary models for residuals of heart mass in 915 birds. The models assume Brownian motion (BM1), a single optimum (OU1) or several optima (OUM) according to the different flight modes considered in this study (Fig. 1; see Materials and methods). The best model (in bold) had the highest Akaike weight (or smallest AICc). lnL, log-likelihood; AICc, Akaike 'small sample' statistic; dAICc, difference between the actual AICc and the smallest (best) AICc; AICw, AIC weight.

1991). According to the OU model, the optimum (residual) heart size for hovering flight was 2.43 times higher than it was for flapping flight (Table 2), whereas this relationship is 2.5 times if we use the intercepts of the pgls analysis (Table 3). By measuring oxygen consumption (a proxy of aerobic power), experimental biologists have shown that the energy expenditure of hovering flight is about 2.5 times the cost of flapping flight (Bartholomew and Lighton, 1986; Lasiewski, 1963; Maina, 2000; Wells, 1993). Therefore, it seems that our estimations are comparable with the known energetic costs of different modes of flight (when ratios are considered). The fact that different approaches yield comparable results is in itself interesting as these data were obtained by very different approaches (i.e. multispecific versus monospecific; comparative analyses versus experimental studies).
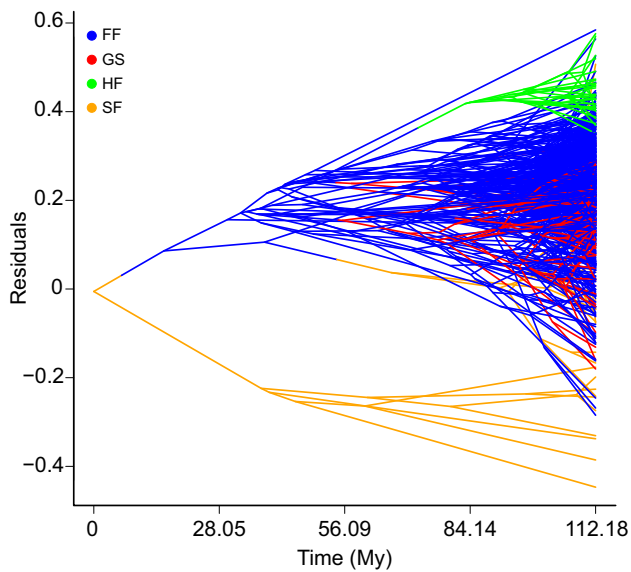
Fig. 2. Phenogram showing trait diversification over time, according to the different flight modes. A phenogram is a combined plot of phylogenetic relationship trait values; each line represents a lineage, and each tip a present-day species. Time calibration was obtained from the original phylogeny (Jetz et al., 2012). FF, flapping flight; GS, gliding and soaring; HF, hovering flight; SF, short flight.

Maina (2000) has pointed out that an important energetic barrier separates flying from non-flying vertebrates. The maximum energy expenditure of non-flying endotherms is about 4−15 times their resting metabolic rate (Hinds et al., 1993; Maina, 2000; Nespolo et al., 2017a), but the metabolic rates of flying endotherms (i.e. birds and bats) when flying is about 10−20 times their resting metabolic rate (Butler and Woakes, 1990; Maina, 2000). Hence, the selective pressures for increasing aerobic power at the transition between SF and FF should have been important. Interestingly, in our analysis, birds that were poor at flying showed the smallest heart sizes and this flight category roughly coincides with the most basal group of birds (Tinamiformes and Galliformes) (Jetz et al., 2012). According to our ancestral trait reconstruction at the node of the bird phylogeny, short flight seems to be the most likely ancestral mode of flight of birds (see Fig. 1, yellow area in the pie chart at the center). Thus, it would reasonable to conclude that there was an important selective pressure for heart enlargement during the short flight to flapping flight transition, which, according to the phenogram (Fig. 2), would have occurred when the avian clade was about ∼25 million years old (see also Jetz et al., 2012; Lee et al., 2014).

In order to analyze the relationship between physiological capacity and flight performance, several authors have used

Table 2. Estimated optima (θ) and computed standard errors for heart mass residuals associated with the different modes of flight

	θ	±s.e.
SF	−0.204	0.039
GS	0.162	0.039
FF	0.205	0.013
HF	0.498	0.076

θ (see Eqn 2) for heart mass residuals and computed standard errors were obtained with the OUwie package (Table 1), and are presented for the different modes of flight assigned to the 915 species of birds considered in this study (see Fig. 2). Different categorizations of flight mode did not produce different results (see Materials and methods for details). SF, short flight; GS, gliding and soaring; FF, flapping flight; HF, hovering flight.

Table 3. Phylogenetic generalized linear model with $\log M_h$ as a function of $\log M_b$ and flight mode

Predictors	Estimate	s.e.	t-value	Significance
Intercept	−2.17	0.09	−25.61	***
$\log M_b$	0.92	0.01	85.03	***
FF	0.23	0.07	3.37	***
GS	0.24	0.07	3.16	**
HF	0.50	0.14	3.72	***

$n=915$, Pagel's $\lambda=0.894$ (0.849–0.926), $r^2=0.89$.
Here, the term 'intercept' represents the short flight mode, and each estimate represents the relative distance from this value. See Materials and methods for details. M_h, heart mass; M_b, body mass; FF, flapping flight; GS, gliding and soaring; HF, hovering flight. **$P<0.01$; ***$P<0.001$.

composite indexes based on linear measurements (e.g. wing loading, wingspan, pectoral muscle mass) combined with multivariate statistics (Alerstam et al., 2007; Vágási et al., 2016; Wright et al., 2014). This approach has the advantage of considering continuous traits of evident biological meaning (e.g. pectoral muscle mass, respiratory pigments, wingspan) as explanatory variables. Instead, we used categorical predictors of flight mode, which we believe has advantages for the question being addressed. First, it avoids the problem of multiple autocorrelations as the two variables (flight and heart size) were obtained from different sources and different observers. Second, it simplifies the problem of comparing a wide range of species, which is especially important when OU models are involved (see discussion of OU models in the context of bioenergetics in Nespolo et al., 2017a).

In summary, this study provides support for the idea that the subtle differences in routine flight mode that we commonly see in birds represent important constraints for shaping the anatomical underpinnings of aerobic power. Given the caveats discussed before, overall our results suggest three main conclusions. First, the
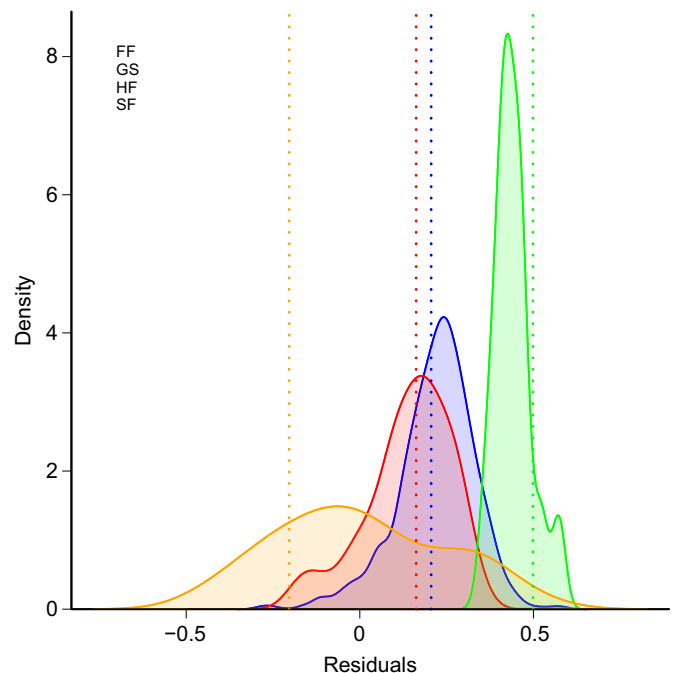

Fig. 3. Kernel density plots showing the actual distribution of trait values for heart mass residuals for different flight modes in bird species. See Fig. 1. The different evolutionary optima obtained by the OUwie procedure (see means±s.e. in Table 2) are indicated by the dotted lines. FF, flapping flight; GS, gliding and soaring; HF, hovering flight; SF, short flight.

flight mode of Trochilidae imposes important selective pressure for increasing heart size. Second, there would be a selective pressure to increase heart size at the short flight to flapping flight transition (short flight being probably the ancestral flight mode). Third, the flapping flight to gliding and soaring transition seems not to have involved a reduction in heart size. This would be either due to the fact that gliding/soaring birds also use flapping flight frequently (i.e. the classification is arbitrary at this boundary) or because the energetic cost of maintaining large hearts in these species does not represent a fitness cost. These conclusions and interpretations are open to debate, which, together with this dataset, may hopefully be extended to improve our understanding of the origin and evolution of flight in birds.

Data availability
Data and extended classification criteria are available from the Dryad digital repository (Nespolo et al., 2017b): doi:10.5061/dryad.1th6k.

References
Alerstam, T., Rosén, M., Bäckman, J., Ericson, P. G. P. and Hellgren, O. (2007). Flight speeds among bird species: allometric and phylogenetic effects. *PLoS Biol.* **5**, 1656-1662.

Bartholomew, G. A. and Lighton, J. R. B. (1986). Oxygen-consumption during hover-feeding in free-ranging anna hummingbirds. *J. Exp. Biol.* **123**, 191-199.

Beaulieu, J. M., Jhwueng, D.-C., Boettiger, C. and O'Meara, B. C. (2012). Modeling stabilizing selection: Expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution* **66**, 2369-2383.

Bernheim, A. M., Jost, C. H. A., Zuber, M., Pfyffer, M., Seifert, B., De Pasquale, G., Linka, A., Faeh-Gunz, A., Medeiros-Domingo, A. and Knechtle, B. (2013). Right ventricle best predicts the race performance in amateur ironman athletes. *Med. Sci. Sports Exerc.* **45**, 1593-1599.

Bishop, C. M. (1997). Heart mass and the maximum cardiac output of birds and mammals: Implications for estimating the maximum aerobic power input of flying animals. *Philos. Trans. R. Soc. Lond. Series B Biol. Sci.* **352**, 447-456.

Bishop, C. M. (1999). The maximum oxygen consumption and aerobic scope of birds and mammals: getting to the heart of the matter. *Proc. R. Soc. B Biol. Sci.* **266**, 2275-2281.

Bishop, C. M. (2005). Circulatory variables and the flight performance of birds. *J. Exp. Biol.* **208**, 1695-1708.

Bishop, C. M. and Butler, P. J. (1995). Physiological modeling of oxygen-consumption in birds during flight. *J. Exp. Biol.* **198**, 2153-2163.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Inference: a Practical Information-Theoretical Approach.* New York: Springer-Verlag.

Butler, M. A. and King, A. A. (2004). Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *Am. Nat.* **164**, 683-695.

Butler, P. and Woakes, A. (1990). The physiology of bird flight. In *Bird Migration* (ed. E. Gwinner), pp. 300-318. Berlin: Springer.

Carpenter, F. L. (1974). Torpor in an andean hummingbird: its ecological significance. *Science* **183**, 544-547.

Chai, P. and Dudley, R. (1996). Limits to flight energetics of hummingbirds hovering in hypodense and hypoxic gas mixtures. *J. Exp. Biol.* **199**, 2285-2295.

Clark, C. J. and Dudley, R. (2010). Hovering and forward flight energetics in Anna's and Allen's hummingbirds. *Physiol. Biochem. Zool.* **83**, 654-662.

Crawley, M. J. (2007). *The R Book.* Chichester: Wiley.

Cressler, C. E., Butler, M. A. and King, A. A. (2015). Detecting adaptive evolution in phylogenetic comparative analysis using the ornstein-uhlenbeck model. *Syst. Biol.* **64**, 953-968.

Dlugosz, E. M., Chappell, M. A., Meek, T. H., Szafranska, P. A., Zub, K., Konarzewski, M., Jones, J. H., Bicudo, J. E. P. W., Nespolo, R. F., Careau, V. et al. (2013). Phylogenetic analysis of mammalian maximal oxygen consumption during exercise. *J. Exp. Biol.* **216**, 4712-4721.

Felsenstein, J. (1973). Maximum likelihood estimation of evolutionary trees from continuous characters. *Am. J. Human Genet.* **25**, 471-492.

Fernández, M. J., Dudley, R. Bozinovic, F. (2011). Comparative energetics of the giant hummingbird (Patagona gigas). *Physiol. Biochem. Zool.* **84**, 333-340.

Freckleton, R. P., Harvey, P. H. and Pagel, M. (2002). Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.* **160**, 712-726.

Gower, D. J. (2001). Possible postcranial pneumaticity in the last common ancestor of birds and crocodilians: evidence from Erythrosuchus and other Mesozoic archosaurs. *Naturwissenschaften* **88**, 119-122.

Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C. K., Braun, E. L., Braun, M. J., Chojnowski, J. L., Cox, W. A., Han, K.-L., Harshman, J. et al. (2008). A phylogenomic study of birds reveals their evolutionary history. *Science* **320**, 1763-1768.

Hansen, T. F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evolution* **51**, 1341-1351.

Hartman, F. A. (1955). Heart weight in birds. *The Condor* **57**, 221-238.

Hedenströem, A. (2008). Power and metabolic scope of bird flight: a phylogenetic analysis of biomechanical predictions. *J. Comp. Physiol. a-Neuroethol. Sens. Neural Behav. Physiol.* **194**, 685-691.

Hillman, S. S. and Hedrick, M. S. (2015). A meta-analysis of in vivo vertebrate cardiac performance: implications for cardiovascular support in the evolution of endothermy. *J. Exp. Biol.* **218**, 1143-1150.

Hinds, D. S., Baudinette, R. V., Macmillen, R. E. and Halpern, E. A. (1993). Maximum metabolism and the aerobic factorial scope of endotherms. *J. Exp. Biol.* **182**, 41-56.

Huelsenbeck, J. P., Nielsen, R. and Bollback, J. P. (2003). Stochastic mapping of morphological characters. *Syst. Biol.* **52**, 131-158.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. and Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature* **491**, 444-448.

Johansen, K., Berger, M., Bicudo, J. E. P. W., Ruschi, A. and de Almeida, P. J. (1987). Respiratory properties of blood and myoglobin in hummingbirds. *Physiol. Zool.* **60**, 269-278.

Krautwald-Junghanns, M. E., Pees, M. and Schutterle, N. (2002). Echocardiographic examinations in unsedated racing pigeons (Columbia livia forma domestica) under special consideration of the physical training. *Berl Munch Tierarztl Wochenschr* **115**, 221-224.

Krebs, J. R. and Harvey, P. H. (1986). Avian physiology: busy doing nothing-efficiently. *Nature* **320**, 18-19.

La Gerche, A., Roberts, T. and Claessen, G. (2014). The response of the pulmonary circulation and right ventricle to exercise: exercise-induced right ventricular dysfunction and structural remodeling in endurance athletes (2013 Grover Conference series). *Pulm. Circ.* **4**, 407-416.

Lasiewski, R. C. (1963). Oxygen consumption of torpid, resting, active and flying hummingbirds. *Physiol. Zool.* **36**, 122-140.

Lee, M. S. Y., Cau, A., Naish, D. and Dyke, G. J. (2014). Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science* **345**, 562-566.

Maina, J. N. (2000). What it takes to fly: the structural and functional respiratory refinements in birds and bats. *J. Exp. Biol.* **203**, 3045-3064.

Martins, E. P. and Hansen, T. F. (1997). Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* **149**, 646-667.

Nespolo, R. F., Solano-Iguaran, J. J. and Bozinovic, F. (2017a). Phylogenetic analysis supports the aerobic-capacity model for the evolution of endothermy. *Am. Nat.* **189**, 13-27.

Nespolo, R. F., Gonzalez-Lagos, C., Solano-Iguaran, J. J., Elfwing, M., Garitano-Zavala, A., Manosa, S., Alonso, J. C. and Altamiras J. (2017a). Data from: Aerobic power and flight capacity in birds: a phylogenetic test of the heart-size hypothesis. Dryad Digital Repository. https://doi.org/10.5061/dryad.1th6k

Price, S. A. and Hopkins, S. S. B. (2015). The macroevolutionary relationship between diet and body mass across mammals. *Biol. J. Linn. Soc.* **115**, 173-184.

Puttick, M. N., Thomas, G. H. and Benton, M. J. (2014). High rates of evolution preceded the origin of birds. *Evolution* **68**, 1497-1510.

Saltin, B. and Rowell, L. B. (1980). Functional adaptations to physical activity and inactivity. *FASEB* **39**, 1506-1512.

Simpson, G. G. (1953). *The Major Features of Evolution*. New York: Columbia University Press.

Spoor, F., Garland, T., Krovitz, G., Ryan, T. M., Silcox, M. T. and Walker, A. (2007). The primate semicircular canal system and locomotion. *Proc. Natl. Acad. Sci. USA* **104**, 10808-10812.

Suarez, R. K., Lighton, J. R., Brown, G. S. and Mathieucostello, O. (1991). Mitochondrial respiration in hummingbird flight muscles. *Proc. Natl. Acad. Sci. USA* **88**, 4870-4873.

Vágási, C. I., Pap, P. L., Vincze, O., Osváth, G., Erritzøe, J. and Møller, A. P. (2016). Morphological adaptations to migration in birds. *Evol. Biol.* **43**, 48-59.

Videler, J. J. (2005). *Avian Flight*. Oxford: Oxford University Press.

Viscor, G. and Fuster, J. F. (1987). Relationships between morphological parameters in birds with different flying habits. *Comp. Biochem. Physiol.* **87A**, 231-249.

Wells, D. J. (1993). Muscle performance in hovering hummingbirds. *J. Exp. Biol.* **178**, 39-57.

Wright, N. A., Gregory, T. R. and Witt, C. C. (2014). Metabolic 'engines' of flight drive genome size reduction in birds. *Proc. R. Soc. B Biol. Sci.* **281**, 20132780.

Wright, N. A., Steadman, D. W. and Witt, C. C. (2016). Predictable evolution toward flightlessness in volant island birds.. *Proc. Natl. Acad. Sci. USA* **113**, 4765-4770.