A Phylogenetic Analysis of the Allometry of Diving

ARTICLE in THE AMERICAN NATURALIST · FEBRUARY 2006

Impact Factor: 4.45 · DOI: 10.1086/499439 · Source: PubMed

CITATIONS
81

3 AUTHORS, INCLUDING:

Lewis Halsey
University of Roehampton
93 PUBLICATIONS 1,297 CITATIONS

Tim M Blackburn
University College London
274 PUBLICATIONS 13,659 CITATIONS

Available from: Lewis Halsey
Retrieved on: 25 August 2015
A Phylogenetic Analysis of the Allometry of Diving

Lewis G. Halsey,* Patrick J. Butler,† and Tim M. Blackburn‡

Centre for Ornithology, School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom

Submitted April 28, 2005; Accepted September 7, 2005; Electronically published January 9, 2006

Online enhancements: appendixes, zip file.

Abstract: The oxygen store/usage hypothesis suggests that larger animals are able to dive for longer and hence deeper because oxygen storage scales isometrically with body mass, whereas oxygen usage scales allometrically with an exponent $b \leq 1$ (typically 0.67–0.75). Previous tests of the allometry of diving tend to reject this hypothesis, but they are based on restricted data sets or invalid statistical analyses (which assume that every species provides independent information). Here we apply information-theoretic statistical methods that are phylogenetically informed to a large data set on diving variables for birds and mammals to describe the allometry of diving. Body mass is strongly related to all dive variables except dive : pause ratio. We demonstrate that many diving variables covary strongly with body mass and that they have allometric exponents close to 0.33. Thus, our results fail to falsify the oxygen store/usage hypothesis. The allometric relationships for most diving variables are statistically indistinguishable for birds and mammals, but birds tend to dive deeper than mammals of equivalent mass. The allometric relationships for all diving variables except mean dive duration are also statistically indistinguishable for all major taxonomic groups of divers within birds and mammals, with the exception of the procellariiforms, which, strictly speaking, are not true divers.

Keywords: allometry, birds, diving behavior, mammals, model, phylogenetics.

Large body mass is usually assumed to confer a variety of benefits on diving animals. Oxygen storage capacity scales isometrically with body mass (e.g., Lasiewski and Calder 1971; Hudson and Jones 1986), while rate of oxygen metabolism scales allometrically with body mass (exponent 0.67–0.75), so that larger animals should be able to dive for longer (the oxygen store/usage hypothesis; see Butler and Jones 1982). Whether they do may depend on any of a number of other factors that could also influence dive duration, including prey quantity, density, distribution, and vertical location in the water column (Butler 2004). Nonetheless, larger animals do have this potential, and longer dives allow extended transit durations and thus the possibility of deeper dives (e.g., Georges et al. 2000; Costa and Gales 2003). Deeper dives in turn often correlate with extended bottom durations to enhance foraging efficiency (the proportion of time spent at the foraging site; Wilson 1991), which further increases overall dive duration in larger divers. However, increased dive depths and durations are likely to require increased surface durations because of the greater oxygen deficits to be repaid (e.g., Kooyman et al. 1980; Costa and Gales 2000). Furthermore, assuming that a diver finishes a dive with the same level of oxygen stores regardless of the depth and duration of the dive (Kramer 1988; Houston and Carbone 1992), oxygen for longer dives will be acquired at a lower average rate because of the diminishing returns of acquisition over time at the surface (Kramer 1988; Halsey et al. 2003a, 2003b). From this it follows that the percentage of time spent at the surface will increase, while the percentage of time underwater will decrease, as dive duration increases. If so, the ratio of dive duration to surface duration (dive : pause ratio; coined by Dewar [1924] as a simple measure of diving efficiency) should decrease with increasing body mass.

Such assumptions about the physiological and behavioral mechanisms behind the apparent allometric scaling of diving behavior (i.e., variation in diving behavior with body mass) are prevalent in the literature on comparative diving behavior, both within and between species (e.g., Stonehouse 1967; Piatt and Nettleship 1985; Gentry et al. 1986; Prince and Harris 1988; Burger 1991; Wilson 1991; Watanuki et al. 1996; Croll et al. 2001; Schreer et al. 2001). Furthermore, a number of such studies have specifically investigated the allometry of diving behavior for a range of taxonomic groups (e.g., table B1 in the online edition of the American Naturalist). For example, Hochachka and Mottishaw (1998) concluded that there is a significant al-
lometric relationship for maximum dive duration in pho-
cid seals but not in otariid seals. Watanuki and Burger
(1999) decided that allometric equations for maximum
dive duration in alcids and penguins did not support a
simple oxygen store/usage hypothesis. Boyd and Croxall
(1996) investigated allometric relationships for pinnipeds
and seabirds and argued that dive duration in pinnipeds
continued to field metabolic rate, while in seabirds, it scaled
to wing surface area, thus suggesting a limitation on diving
for avian species resulting from the conduction of heat to
water. To date, the largest study on the allometry of diving
is that of Schreer and Kovacs (1997), who looked at max-
imum diving depth and duration in relation to body mass
for 132 species of birds, marine mammals, and marine
turtles. Strong allometric correlations were found for max-
imum dive duration and depth for a number of taxonomic
groups, but there were notable exceptions, such as most
groups of diving, volant birds (e.g., cormorants).

Unfortunately, the reliability of the conclusions from
most studies of the allometry of diving may be compro-
mised by a failure to account for the nonindependence of
species due to their interrelatedness. The hierarchical na-
ture of evolutionary relationships means that closely re-
lated species tend to resemble each other because they
share many characters through common descent rather
than through convergent or parallel evolution (Harvey and
Pagel 1991). This has several important consequences for
analyses exploring allometric relationships. Notably, a fail-
ure to incorporate information on the relatedness of spe-
cies can lead to estimates of allometric exponents that are
biased and to increased rates of Type I error (see app. A
in the online edition of the American Naturalist for further
discussion). The conclusions of analyses that fail to in-
corporate phylogenetic information are thus open to
doubt.

The main aim of this study is to present a comparative
analysis of the allometry of diving behavior (dive duration
and depth, surface duration, and dive : pause ratio) for
mammals and birds that incorporates phylogenetic infor-
mation on the species concerned. We use an information-
theoretic approach (Burnham and Anderson 2002) to
identify the most likely model for allometric relationships
of the different dive variables from a candidate set of mod-
els that make different assumptions about how those vari-
ables have evolved. We use this approach to produce best
estimates of the form of allometric relationship for each
dive variable, given current data. This allows us to assess
whether the relationship differs in either elevation or slope
between birds and mammals, whether any families of birds
and mammals deviate significantly from the overall allom-
etry for the group, and what the data for dive variables
can tell us about the likely pattern of evolution of these
traits. We finish by discussing the implications of our re-
sults for the allometry of diving.

**Methods**

**Data**

A database was compiled on behavioral diving variables
for as many diving avian and mammalian species as could
be found in the literature, using as many published sources
as possible, some unpublished sources, and correspondence
with authors. The database includes data for 195
species, taken from 286 studies (see table in zip file avail-
able for download in the online edition of the American
Naturalist), making full use of the increase in diving lit-
iture since the study by Schreer and Kovacs (1997).

Body masses of the species were often present in the
literature, but where not, they were provided by the au-
ths on request or taken from alternative published
sources (e.g., Dunning 1992) or from reputable Web sites.
Usually, mean values for dive duration, dive depth, and
surface duration were taken directly from the literature.
Sometimes they were calculated from tabulated data or as
accurately as possible from histograms (e.g., Croxall et al.
1988; Le Boeuf et al. 1988; Kooyman et al. 1992; Boyd
and Croxall 1996). The analyses in this study necessitated
only one value for each diving variable per species, so mean
values were calculated in every case. While the data for
diving behavior of some species are not normally distrib-
uted or even unimodal (e.g., dive depth of macaroni pen-
guins; Green et al. 2003), the error variance that this may
introduce to the mean estimate for any given species will
be small compared with the variation in body mass and
dive variables across the species analyzed. Mean maximum
values of dive duration and dive depth were derived from
the means of the longest and deepest dives, respectively,
of each individual animal. The mean dive : pause ratio was
sometimes stated in the source text; however, if this was
not the case but mean dive duration and mean surface
duration were stated, then mean dive : pause ratio was
calculated from these values. While some animals under-
take some partially anaerobic dives, the two studies to date
that have measured postdive blood lactate levels suggest
that arterial lactate concentrations exceed resting levels
after only a very few dives (2.7% in Weddell seals [Koo-
man et al. 1980] and only after dives exceeding around 7
min in emperor penguins [Ponganis et al. 1997], about
6% of all dives [Kooyman and Kooyman 1995]). Thus,
while the diving variables calculated for some species in
this study may be derived in part from dives incorporating
an anaerobic element, such dives are at most a small pro-
portion of all dives performed.

The majority of data found in the literature were added
to the database; however, values that were deemed to be insufficiently valid or robust were not included. This was the case, for example, with studies where water depths for the particular population observed were deemed to be unusually shallow for that species (e.g., Cooper 1986). No laboratory-based data were used (e.g., Carbone et al. 1996) because of the prominent confounds that experimental work often introduces into the behavior of species (Halsey et al., forthcoming). Where multiple field observations were recorded for the same species at different locations (e.g., Cooper 1986), an average was calculated for each diving variable. Behavioral data are available for some species from multiple sources, and in such cases, the mean for each diving variable for all the studies was used in the database.

**Phylogenies**

Phylogenetically informed analyses require an estimate of the phylogenetic relationships among the taxa concerned. We assumed that the relationship among bird families followed Sibley and Ahlquist (1990). We used the following references for the relationships between species within families with more than two species: Alcidae, Friesen et al. (1996); Anatidae, Livezey (1995a, 1995b, 1996, 1997); Gaviidae, Boertmann (1990); Phalacrocoracidae, Kennedy et al. (2000); Podicipedidae, del Hoyo et al. (1992); Procellariidae, Kennedy and Page (2002); Spheniscidae, Giannini and Bertelli (2004); Sulidae, Friesen and Anderson (1997). In addition, we assumed that *Eudyptes chrysocephalus filholi* forms a polytomy with *Eudyptes chrysocephalus* and *Eudyptes chrysocephalus moseleyi* and that *Pachyptila belcheri* would cluster with other *Pachyptila* species. The resulting bird phylogeny is presented in figure C1 in the online edition of the *American Naturalist*.

We assumed that the relationship among mammal orders followed Murphy et al. (2001a, 2001b) and Arnason et al. (2002), who agree in their topology for the orders in our data set. For the relationships between species within orders with more than two species, we assumed that Carnivora followed Bininda-Emonds et al. (1999) and that Cetacea followed Bininda-Emonds et al. (2005) and that the phylogeny of Messenger and McGuire (1998) lodged on Treebase (www.treebase.org), with *Hyperoodon ampullatus* as the sister taxon of *Hyperoodon planifrons*. The resulting phylogeny for mammals is presented by Sibley and Ahlquist (1990) in their tapestry phylogeny, with the heights calculated using our proportionality assumption. These spanned the range from the oldest node in the bird phylogeny down to nodes at the genus level. The correlation between node heights calculated these two ways was 0.93 (N = 20, P < .001). Thus, the proportionality assumption produces a realistic set of branch lengths, at least for the bird taxa in our data set.

**Analyses**

We used the method of phylogenetic generalized least squares (PGLS) to control for phylogenetic independence (Grafen 1989; Martins and Hansen 1997; Garland and Ives 2000). The PGLS approach was implemented in R (Ihaka and Gentleman 1996) using the APE (Analysis of Phylogenetics and Evolution) package (Paradis et al. 2004) and code written by R. P. Duncan. Because the PGLS method may be unfamiliar to some readers, we present a detailed description of it in appendix A. Here, we present only a brief outline.

PGLS controls for phylogenetic nonindependence by explicitly incorporating the expected covariance among species into a statistical model fitted by generalized least squares. The correlation between error terms is thus altered to reflect the degree of phylogenetic relatedness among the species to which they relate. Traits are assumed to evolve by a “Brownian motion” model of evolution (Rohlf 2001), under which the expected trait covariance between any two species is directly proportional to the amount of shared evolutionary history. However, the covariance matrix can be modified in PGLS to accommodate the degree to which trait evolution deviates from Brownian
motion, using a measure of phylogenetic correlation $\lambda$, derived by Pagel (1999; see also Freckleton et al. 2002); $\lambda$ is a multiplier of the off-diagonal elements of the covariance matrix (i.e., those quantifying the degree of relatedness between species), normally varying between 0 and 1. If the covariance matrix is constructed assuming a Brownian motion model of evolution, then $\lambda = 1$ retains that model, while $\lambda = 0$ specifies phylogenetic independence. Intermediate values of $\lambda$ specify models in which trait evolution is phylogenetically correlated but to a lesser extent than expected under the Brownian motion model. The maximum-likelihood value of $\lambda$ can be estimated by fitting PGLS models with different values of $\lambda$ and finding the value that maximizes the log likelihood. This best-fitting model can be used as a basis for inference (e.g., of the form of an allometric relationship), while the value of $\lambda$ associated with it can be used as a metric of the degree of phylogenetic correlation in the data (Freckleton et al. 2002).

We also modeled evolution using an Ornstein-Uhlenbeck (O-U) process, which constrains the Brownian motion “random walk” toward some central point (Martins 1994; Butler et al. 2000; Butler and King 2004). Evolution according to O-U processes results in greater similarity between related species than does the Brownian motion model but a more rapid decay in similarity, with increasing evolutionary separation between species, as expected, for example, under convergent evolution or stabilizing selection. The strength of the constraining force is given by a parameter $\alpha$, the maximum-likelihood value of which can also be estimated by fitting PGLS models with different values of $\alpha$ and finding the value that maximizes the log likelihood.

Thus, we specified seven different models for the evolution of the allometry of the various diving variables. These are first, a nonphylogenetic model using simple ordinary least squares regression, equivalent to a phylogenetic model with $\lambda = 0$ or to the assumption that all species are equally related (a “star-shaped” phylogeny); second, Brownian motion evolution ($\lambda = 1$) with proportional branch lengths; third, Brownian motion evolution ($\lambda = 1$) with equal branch lengths; fourth, a modification of the Brownian motion model using the maximum-likelihood-optimized value of $\lambda$, with proportional branch lengths; fifth, a modification of the Brownian motion model using the maximum-likelihood-optimized value of $\lambda$, with equal branch lengths; sixth, an O-U model of evolution under stabilizing selection, with proportional branch lengths; and finally, an O-U model of evolution under stabilizing selection, with equal branch lengths. Note, however, that the O-U model does not converge for all combinations of branch lengths and data and is therefore missing from some comparisons.

We applied each of these seven evolutionary models to three different statistical models for the allometry of each diving variable. First, we modeled the relationship between each diving variable and body mass for birds and mammals combined. Second, we modeled the relationship between each diving variable and body mass, including class (birds or mammals) as a covariate. This tests for different intercepts for birds and mammals in the relationship between each diving variable and mass but assumes a common regression slope (allometric exponent). Finally, we repeated the second model with the addition of the interaction term between mass and class. This tests for different intercepts and allometric slopes for birds and mammals. This gives a maximum of 21 different models for each allometric relationship.

Burnham and Anderson’s (2001, 2002) framework for model comparison was used to identify the most plausible of the model(s) of trait evolution based on Akaike’s Information Criterion (AIC) as a measure of model fit. The best out of all of the (evolutionary + statistical) models tested to explain each dive variable was the model with the lowest AIC. The probability that any given model is actually the best fit out of those tested was measured by its Akaike weight, the relative penalized likelihood of the model compared to all others (the penalized likelihood of the model divided by the sum of the penalized likelihoods of all other models). Finally, for each diving variable, we explored the extent to which the data were better fitted by a model replacing taxonomic class with the following nine taxonomic groups: auks, cetaceans, seals, sea lions (including fur seals), procellariiforms, grebes, cormorants, ducks, and penguins. These are groups with at least 10 species represented in our data. These groups span a range of taxonomic levels from subfamily (e.g., auks) to order (e.g., cetaceans) but can legitimately be compared because they are all monophyletic taxa that represent different evolutionary responses to the challenges posed by diving. We excluded 10 groups for which we had too few species for meaningful conclusions about the allometry of any of the dive variables to be drawn: anhingas ($n = 1$), boobies (4), tropicbirds (1), loons (4), platypuses (1), dugongs (1), insectivores (2), rodents (1), primates (1), and otters (3). For some variables, some of the nine groups with more than 10 species had to be excluded because of low sample sizes. These “family” models cannot be directly compared by AIC with those for the full data set because the models are conditional on the data to which they are fitted. Therefore, we compared the fit of each family-level model to the best-fit model from the full data set recalculated on the reduced set for each parameter.
Results

Table 1 summarizes the best single models for the six dive variables. Because the statistical process to explore the allometry of each dive variable is relatively involved, we provide in appendix A an extended example using the allometry of mean dive duration. Here, we present the results for the diving variables, calculated as described in appendix A, in abbreviated form.

The relationship between mass and mean dive duration is presented in figure 1. Mean dive duration is best fitted by a model with mass as the sole predictor, with evolution along equal branch lengths and a maximum-likelihood value of $\lambda = 0.91$ (table 1). However, the Akaike weight for this model is only 0.279, and four other models receive reasonable support ($\Delta \text{AIC} < 2$). These other models are those with mass alone and equal branch lengths (either $\lambda = 1$ or maximum likelihood [ML] $\lambda = 0.93$) and those with mass + class and ML $\lambda$ (either $\lambda = 0.91$ and equal branch lengths or ML $\lambda = 0.93$ and proportional branch lengths). A model with mass alone as a predictor is the most likely for mean dive duration (sum of Akaike weights $\Sigma w = 0.62$), although the model with mass + class also receives some support ($\Sigma w = 0.27$).

In no case did a model distinguishing between the nine lower taxonomic groups with $>10$ species provide a better fit to the data for mean dive duration than the best-fit model in table 1 applied to the reduced data set (taxa of $<10$ species excluded). However, most of the models suggested that the slope and intercept for procellariiforms differed from those of other groups (see fig. 2). Repeating the analyses in table 1 (i.e., on the full data set) and distinguishing between procellariiforms and other species resulted in a best model that included separate terms for both slope and intercept for procellariiforms relative to all other species (proportional branch lengths and ML $\lambda = 0.88$) and is a substantial improvement in fit over a model with mass alone ($\text{AIC} = 55.36$ vs. 68.24). The allometric slopes are $-0.419 \pm 0.240$ for procellariiforms and 0.368 ± 0.045 for all other species. The allometric relationship for nonprocellariiforms is mean dive duration = 21.2M$^{0.561}$. This slope is plotted in figure 1 (dotted line).

Maximum dive duration is best fitted by a model for mass alone, with Brownian motion evolution along equal branch lengths. The only other model receiving reasonable support was the same model with class added as a factor ($\Delta \text{AIC} = 1.98$, Akaike weight = 0.216). Thus, we base the allometric equation for mean maximum dive duration on the clearly best-fitting model: mean maximum dive duration = 35.5M$^{0.325}$.

Mean surface duration is best fitted by a model for mass alone with Brownian motion evolution along equal branch lengths. However, five other models receive reasonable support ($\Delta \text{AIC} < 2$). These six best models are the models with equal branch lengths (either $\lambda = 1$ or ML $\lambda = 0.91$, and with mass alone, mass + class, or mass + class + [class × mass]). Thus, while a model with mass alone as a predictor is the most likely for surface duration ($\Sigma w = 0.47$), those for mass + class ($\Sigma w = 0.27$) and mass × class ($\Sigma w = 0.26$) also receive reasonable support. We base the allometric equation for mean surface duration on the weighted average across all models: mean surface duration = 18.8M$^{0.331}$.

The dive : pause ratio is best fitted by a model for mass + class with Brownian motion evolution along equal branch lengths (table 1). However, the equivalent evolutionary models for mass alone and mass + class + (mass × class) fit almost equally well ($\text{AIC} = 78.77$ and 78.76, $\Delta \text{AIC} = 0.19$ and 0.18, Akaike weights = 0.264 and 0.266, respectively). Neither class nor its interaction term with mass is significant in these models ($P > .14$ in all cases), suggesting that the most parsimonious model is one with mass alone. Because no single model for dive : pause ratio is clearly the best, we base the allometric equation on the weighted average across all models: mean dive : pause ratio = 1.8M$^{0.35}$. Mean dive depth is best fitted by a model for mass + class, with trait evolution modeled by the Ornstein-Uhlenbeck process along proportional branch lengths (table 1). The next most likely model is the equivalent evolutionary model for mass alone (Akaike weight = 0.352). The class term indicates that mammals tend to make shallower dives for a given body mass, although class is not formally statistically significant in this model ($P = .12$). Nevertheless, models with a class term are better supported overall than those without one (Akaike weights of 0.452 and 0.352, respectively), and we suggest that separate models for birds and mammals may be more appropriate in this case: mean dive depth = 10.5M$^{0.389}$ for birds and 3.8M$^{0.389}$ for mammals. Note that the weighted-average allometric exponent across all models is similar to that for the best model, at 0.361. Mean maximum dive depth is best fitted by a model for mass + class with Brownian motion trait evolution along equal branch lengths modified by a maximum-likelihood $\lambda = 0.91$ (table 1). However, three other models are reasonably well supported ($\Delta \text{AIC} < 2$), including the equivalent models for mass alone and with mass × class added and the model for mass alone with stabilizing selection ($\alpha = 19.78$) along proportional branch lengths. Overall, models with a class term are only slightly more probable fits to the data than models with mass alone ($\Sigma w = 0.399$ and 0.377, respectively), while the class term is never formally statistically significant, suggesting that the most parsimonious model is one with mass alone.
Table 1: Most likely single model for diving variables, given the data

<table>
<thead>
<tr>
<th>Statistical model</th>
<th>N</th>
<th>Phylogeny</th>
<th>Evolutionary model</th>
<th>Allometric exponent (± SE)</th>
<th>Intercept (± SE)</th>
<th>AIC</th>
<th>Akaike weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean dive duration:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mass</td>
<td>152</td>
<td>Proportional</td>
<td>(\lambda = .91)</td>
<td>.349 ± .048</td>
<td>1.322 ± .376</td>
<td>68.24</td>
<td>.279</td>
</tr>
<tr>
<td>Weighted-average parameters</td>
<td></td>
<td></td>
<td></td>
<td>.337</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean maximum dive duration:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mass</td>
<td>90</td>
<td>Equal</td>
<td>(\lambda = 1)</td>
<td>.326 ± .043</td>
<td>1.550 ± .175</td>
<td>31.82</td>
<td>.582</td>
</tr>
<tr>
<td>Mean surface duration:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mass</td>
<td>91</td>
<td>Equal</td>
<td>(\lambda = 1)</td>
<td>.245 ± .071</td>
<td>.993 ± .177</td>
<td>72.12</td>
<td>.242</td>
</tr>
<tr>
<td>Weighted-average parameters</td>
<td></td>
<td></td>
<td></td>
<td>.331</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean dive : pause ratio:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mass</td>
<td>86</td>
<td>Equal</td>
<td>(\lambda = 1)</td>
<td>.103 ± .078</td>
<td>.208 ± .283</td>
<td>78.58</td>
<td>.291</td>
</tr>
<tr>
<td>Class</td>
<td></td>
<td></td>
<td></td>
<td>.035</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted-average parameters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean dive depth:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mass</td>
<td>87</td>
<td>Proportional</td>
<td>O-U, (\alpha = 20)</td>
<td>.398 ± .091</td>
<td>1.023 ± .140</td>
<td>118.28</td>
<td>.449</td>
</tr>
<tr>
<td>Class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−.422 ± .268</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean maximum dive depth:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mass</td>
<td>100</td>
<td>Equal</td>
<td>(\lambda = .91)</td>
<td>.326 ± .071</td>
<td>1.029 ± .305</td>
<td>116.42</td>
<td>.303</td>
</tr>
<tr>
<td>Class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−.892 ± .472</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted-average parameters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Values for class in the intercept column represent the difference between bird and mammal slopes. AIC = Akaike’s Information Criterion.
Figure 1: Relationship between log mean body mass and log mean dive duration, distinguishing between birds (filled circles, thin solid line) and mammals (open circles, dashed line; lines for each class separately are based on model Q in table A1). The thicker regression lines indicate the best-fit model for all birds and mammals (thick solid line; model in table 1) and the best-fit model for all nonprocellariiforms (dotted line; model in table 2). Thus, in all cases the regression lines are derived from models that incorporate phylogenetic information.

Discussion

Interspecific comparative analyses must consider the possibility that the traits being compared exhibit phylogenetic autocorrelation because the species analyzed share many characteristics through common descent. Phylogenetic autocorrelation can result in biased estimates of regression coefficients, such as allometric exponents, and in underestimate of standard errors. In allometric studies, the latter may lead to erroneous conclusions about the difference between observed and theoretical exponents (see also White and Seymour 2004).

Our analyses strongly suggest a high degree of phylogenetic autocorrelation in dive variables for birds and mammals. For each variable, the most likely model from those tested included a hierarchical model for evolutionary relationships among species (rather than the star phylogeny representing a nonphylogenetic model; table 1). Moreover, the most likely models in most cases included trait evolution that closely matched Brownian motion along the phylogenetic topology, as indicated by $\lambda$ equal to or close to 1. The one exception was the best model for mean dive depth, which was consistent with this trait evolving according to stabilizing selection. Nevertheless, stabilizing selection produces a pattern of variation whereby closely related species are more similar to each other in the trait than expected under Brownian motion. Thus, in no case is the assumption of independence across species upheld. Indeed, in most cases, the Akaike weights of the nonphylogenetic models suggested that the probability that they were the best models for any given parameter was negligible.

That said, our results also highlight the dangers of fitting
loses significance; mass is a highly significant component that the relationship between mass and diving variables the confidence intervals do not increase to such an extent (e.g., one-quarter vs. one-third power scaling). However, components or about their difference from theoretical values conclusions will be drawn about the values of such exponents. This in turn reduces the likelihood that false increases the likelihood that these intervals embrace the true association with the allometric exponent and hence also in-
tion does, however, increase the confidence intervals as-

In other words, there is no evidence that different forces are driving the allometry of diving variables within groups and between them. Incorporating phylogenetic information does, however, increase the confidence intervals associated with the allometric exponent and hence also increases the likelihood that these intervals embrace the true exponent. This in turn reduces the likelihood that false conclusions will be drawn about the values of such exponents or about their difference from theoretical values (e.g., one-quarter vs. one-third power scaling). However, the confidence intervals do not increase to such an extent that the relationship between mass and diving variables loses significance; mass is a highly significant component of the best models for all the various dive variables, with the exception of dive : pause ratio. Thus, the allometry of dive variables is not simply a consequence of phylogenetic nonindependence.

For all diving variables except dive : pause ratio, the equations produced by the analyses of this study have a mass exponent close to 0.33 (table 2). While most of the mass exponents for the diving variables in our study are not significantly different from 0.25, they are consistently closer to 0.33 and so perhaps offer evidence for a metabolic rate during diving that scales with body mass to the two-thirds power. More importantly for this study, these values are quite different from most of the exponents found in previous allometric studies of diving behavior (table B1). This often leads to quite different conclusions being drawn. For example, despite the many other factors that can affect dive duration (see the introduction to this article), the data from this study do not reject the oxygen store/usage hypothesis. Indeed, even mean dive duration conforms to this hypothesis, suggesting that diving animals may in some way monitor rate of oxygen uptake during submersion. These conclusions contrast with those of previous diving allometry studies, whose results have tended to reject the oxygen store/usage hypothesis, for example, for penguins and alcids (e.g., Watanuki and Burger 1999). It is interesting to note, however, that the oxygen store/usage hypothesis is based on the notion that the rate of oxygen usage underwater scales with body mass to the power of 0.67–0.75, that is, similarly to basal metabolic rate. However, there is growing evidence that based on their estimated usable oxygen stores and measured dive durations, some species must be using oxygen during their longer dives at a lower rate than when they are resting at the surface (Bevan et al. 2002; Butler 2004). In other words, assuming that the estimates of the usable oxygen stores are reasonable and that there is little or no utilization of anaerobic metabolic pathways, including the use of the phosphocreatine and ATP stores, these species must become hypometabolic during their longer dives. This may potentially affect predictions of the oxygen store/usage hypothesis if only species at certain body masses become

### Table 2: Summary of allometric equations for dive variables suggested by the analyses

<table>
<thead>
<tr>
<th>Diving variable</th>
<th>Allometric equation</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean dive duration</td>
<td>21.2M$^{0.558}$</td>
<td>Excluding procellariiforms</td>
</tr>
<tr>
<td>Mean maximum dive duration</td>
<td>35.5M$^{0.326}$</td>
<td>Single best model</td>
</tr>
<tr>
<td>Mean surface duration</td>
<td>18.8M$^{0.531}$</td>
<td>Weighted average over all models</td>
</tr>
<tr>
<td>Mean dive : pause ratio</td>
<td>1.8M$^{0.055}$</td>
<td>Weighted average over all models</td>
</tr>
<tr>
<td>Mean dive depth:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birds</td>
<td>10.5M$^{0.589}$</td>
<td>Single best model, indicating a difference between classes</td>
</tr>
<tr>
<td>Mammals</td>
<td>3.8M$^{0.589}$</td>
<td>Single best model, indicating a difference between classes</td>
</tr>
<tr>
<td>Mean maximum dive depth</td>
<td>9.4M$^{0.327}$</td>
<td>Weighted average over all models</td>
</tr>
</tbody>
</table>

unjustified phylogenetic models. For example, a model for mean dive duration in terms of body mass alone assuming proportional branch lengths and Brownian motion trait evolution ($\lambda = 1$) is a poorer fit to the data than the equivalent nonphylogenetic model (app. $A$; note also that if the equivalent analysis is performed in CAIC, ver. 2.6.9 [Purvis and Rambaut 1995], warning messages are produced notifying the user of violations of the evolutionary assumptions). Nevertheless, a greatly improved model fit is achieved by the apparently small modification of scaling the branch lengths by $\lambda = 0.93$ instead of $\lambda = 1$.

Despite the considerably better fit to the data normally obtained by incorporating information on phylogenetic autocorrelation, the allometric exponents derived from these models do not differ dramatically from those obtained by standard regression. For example, allometric exponents ($\pm$ SE) from the best-fit model for mean surface duration and the nonphylogenetic model were 0.245 $\pm$ 0.071 and 0.252 $\pm$ 0.051, respectively, while the equivalent exponents for mean maximum dive duration were 0.326 $\pm$ 0.043 and 0.358 $\pm$ 0.025. Other analyses have found that incorporating phylogenetic information can markedly alter regression exponents (e.g., Gaston and Blackburn 2000, p. 257), as may happen, for example, if the slope changes with taxonomic level (e.g., “grade shifts”; Pagel and Harvey 1988). That our analyses do not show such differences means, in effect, that the allometric relationship is the same within and across taxonomic groups. In other words, there is no evidence that different forces are driving the allometry of diving variables within groups and between them. Incorporating phylogenetic information does, however, increase the confidence intervals associated with the allometric exponent and hence also increases the likelihood that these intervals embrace the true exponent. This in turn reduces the likelihood that false conclusions will be drawn about the values of such exponents or about their difference from theoretical values (e.g., one-quarter vs. one-third power scaling). However, the confidence intervals do not increase to such an extent that the relationship between mass and diving variables loses significance; mass is a highly significant component of the best models for all the various dive variables, with the exception of dive : pause ratio. Thus, the allometry of dive variables is not simply a consequence of phylogenetic nonindependence.

For all diving variables except dive : pause ratio, the equations produced by the analyses of this study have a mass exponent close to 0.33 (table 2). While most of the mass exponents for the diving variables in our study are not significantly different from 0.25, they are consistently closer to 0.33 and so perhaps offer evidence for a metabolic rate during diving that scales with body mass to the two-thirds power. More importantly for this study, these values are quite different from most of the exponents found in previous allometric studies of diving behavior (table B1). This often leads to quite different conclusions being drawn. For example, despite the many other factors that can affect dive duration (see the introduction to this article), the data from this study do not reject the oxygen store/usage hypothesis. Indeed, even mean dive duration conforms to this hypothesis, suggesting that diving animals may in some way monitor rate of oxygen uptake during submersion. These conclusions contrast with those of previous diving allometry studies, whose results have tended to reject the oxygen store/usage hypothesis, for example, for penguins and alcids (e.g., Watanuki and Burger 1999). It is interesting to note, however, that the oxygen store/usage hypothesis is based on the notion that the rate of oxygen usage underwater scales with body mass to the power of 0.67–0.75, that is, similarly to basal metabolic rate. However, there is growing evidence that based on their estimated usable oxygen stores and measured dive durations, some species must be using oxygen during their longer dives at a lower rate than when they are resting at the surface (Bevan et al. 2002; Butler 2004). In other words, assuming that the estimates of the usable oxygen stores are reasonable and that there is little or no utilization of anaerobic metabolic pathways, including the use of the phosphocreatine and ATP stores, these species must become hypometabolic during their longer dives. This may potentially affect predictions of the oxygen store/usage hypothesis if only species at certain body masses become
hypo-metabolic, producing a different allometry for hypo-
metabolic rate (Fahlman et al. 2005). The analyses in this study suggest that dive : pause ratio is invariant with respect to body mass and hence that there is no tendency for diving birds and mammals of different sizes to spend different proportions of time diving and recovering. As already indicated, a mass exponent close to one-third for dive duration is consistent with the oxygen store/usage hypothesis. However, a similar mass exponent for surface duration is not so easily explained. It is not clear that the replacement of oxygen after a dive should scale with metabolic rate in the same way as oxygen usage during a dive. Furthermore, surface duration may be more related to the removal of carbon dioxide than to the replacement of oxygen, and the rate of the former may be markedly slower than that of the latter, depending on the location and activity of carbonic anhydrase (Boutilier et al. 2001). Moreover, some species of birds, at least, remove CO$_2$ in excess of the rate at which it is produced just before submerging, probably to ensure that high levels of CO$_2$ during submersion do not produce an urge to surface and breathe while there are still sufficient oxygen stores remaining (Halsey et al. 2003a). A similar phenomenon has been proposed for Weddell seals (Stephenson 2005). In addition, birds in particular may spend varying periods of time at the surface engaged in other activities, such as preening. All of these factors may also contribute to the fact that dive : pause ratio varies considerably at a given mass, albeit that it does not vary systematically with mass.

In this study, distinguishing between birds and mam-
mals generally gave very little improvement in terms of model fit. For three of the dive variables (table 1), models with taxonomic class were unequivocally inferior to those with mass alone. Thus, there is little evidence from these data that for a given body mass, birds and mammals differ in mean dive duration (see fig. 1), mean maximum dive duration, or mean surface duration. For dive : pause ratio and maximum dive depth, adding class produced marginal improvements in model fit, although class was not formally significant (at $\alpha = 0.05$) in the best-fit models. In these cases, a model with mass alone was considered the most parsimonious (note also that models selected by AIC are often argued to be overfitted). Additional data might elucidate whether birds and mammals truly tend to differ in these traits. Although there was very little overlap in body mass between birds and mammals, our analysis sug-
gests that diving behavior varies little between these two classes. Only for mean dive depth did birds and mammals clearly differ. For a given body mass, our data indicate that on average, birds dive to greater depths. Overall, how-
ever, our analyses suggest that the allometry of diving dif-
ers little between birds and mammals.

Possible differences between taxa were also explored by testing for variation at lower taxonomic levels. This al-
lowed us to detect a difference between procellariiforms and other species for mean dive duration (fig. 2). The allometric slope for procellariiforms was negative, reflect-
ing the reduced dive durations of large-bodied albatrosses compared to those of smaller-bodied petrels. Reasons for this are not obvious, but a speculative possibility is that the larger procellariiforms are more buoyant. They cer-
tainly have a less densely packed abdominal cavity than penguins, ducks, geese, gannets, and cormorants (P. J. But-
ler, personal observation). Albatrosses spend only a small proportion of their time underwater (Prince et al. 1994) and are better evolved to glide and soar than to dive. Indeed, they are often considered not to be true divers. For most other diving variables, the sample sizes were too small to explore differences among any more than the six best-represented groups (cetaceans, seals, sea lions, auks, cormorants, penguins). Thus, it was not possible to assess the extent to which procellariiforms differed in other traits. Distinguishing between the “big six” groups did not im-
prove the fit of any models, although the models in this study are contingent on the data used, and more data might have told a different story.

Dive depth and dive duration show strong covariation with body mass, for the most part in a consistent manner across taxonomic groups. Nevertheless, body mass is unlikely to determine diving ability directly; rather, it probably correlates with something that does. The predictions of the oxygen store/usage hypothesis that allometric ex-
ponents for diving variables should be close to 0.33 are not falsified by our data, suggesting that one strong pos-
sibility is that body mass is a surrogate for size of oxygen stores. Nevertheless, there is clearly considerable variation in the diving ability of individual species that is not ex-
plained by body mass alone, as illustrated in figures 1 and 2. For example, the sperm whale Physeter catodon and the gray whale Eschrichtus robustus have similar body masses, yet the mean dive duration of the former is more than an order of magnitude greater than that of the latter. Similarly, the harp seal Pagophilus (Phoca) groenlandica typically dives for three times as long as the similarly sized southern sea lion Otaria flavescens (see table in zip file available for download). While the amount of variation in diving be-
havior accounted for by body mass is impressive when considering the diversity of organisms included within the data set, nevertheless there is value in considering the causes of the unexplained variation. We can think of two possible reasons for this variation.

First, body mass is only an approximate surrogate for oxygen stores and usage. The predictions of the oxygen store/usage hypothesis derive from the assumptions that oxygen stores scale isometrically with mass, whereas ox-
xygen use scales to the two-thirds (or three-quarters) power.
However, there is scatter around both of these allometric relationships, which will in turn lead to scatter around the allometric relationships for diving. While this is surely true, we doubt that it is the only explanation.

Second, species vary in their ecology, their behavior, and aspects of their physiology other than their ability to store oxygen. Variation in any or all of these factors may lead to additional variation in diving ability. For example, seals (Phocidae) may be able to dive deeper and for longer, on average, than sea lions (Otariidae) by dint of slower dive speeds and the relatively higher oxygen carrying capacity provided by larger spleens (Hochachka and Mottishaw 1998; our data do not corroborate other suggestions by Hochachka and Mottishaw [1998] that these differences relate to differences in body mass or diving allometry between these groups; fig. 2 shows that the mass ranges of these groups are very similar and that the ranges of mean dive durations overlap considerably). Sperm whales may make longer dives than gray whales because the former have greater oxygen stores and/or lower rates of oxygen usage during diving than the latter. Alternatively, the sperm whale feeds at great depths (mean dive depth = 571 m) that require extended dive durations, whereas the gray whale feeds in shallower coastal waters that do not (mean dive depth = 12.5 m; see table in zip file). The gray whale may have the capacity to make longer and deeper dives, but it rarely has either the need or the opportunity to do so (cf. Ridgway et al. 1984; Martin and Smith 1992 with regard to beluga whales Delphinapterus leucas). While further exploration of the residual variation around these allometric relationships is important, nevertheless, these differences in ecology and physiology do not impinge on the general allometry of diving. This study’s analyses of the relationship between diving ability and body mass suggest that whatever underlies these allometric relationships, be it the interaction between oxygen stores and usage or an alternative explanation, it does so in a consistent manner across a wide range of avian and mammalian groups.

Acknowledgments

We thank R. P. Duncan for allowing us to use his R code for PGLS, for extensive illuminating discussions about this methodology, and for commenting on the manuscript. We also thank A. Fahlman and T. Clark for their help obtaining papers, B. M. Bezerra for assistance collating the extensive reference list, and P. Stephens for commenting on an earlier draft of this manuscript. We received support from the Natural Environment Research Council (grant NER/ A/S/2000/01074).

Literature Cited


Pagel, M. D., and P. H. Harvey. 1988. The taxon-level problem in