High flight costs, but low dive costs, in auks support the biomechanical hypothesis for flightlessness in penguins

Kyle H. Elliotta,b,1, Robert E. Ricklefsb,1, Anthony J. Gastona, Scott A. Hatchd, John R. Speakmane,f, and Gail K. Davorenb

*Department of Zoology, University of Manitoba, Winnipeg, MB, Canada R3T 2N2; bDepartment of Biology, University of Missouri, St. Louis, MO 63121-4499; cCanadian Wildlife Service, National Wildlife Research Centre, Carleton University, Ottawa, ON, Canada K1A 0H3; dAlaska Science Center, USGS, Anchorage, AK 99506-4626; eInstitute of Biological Sciences, University of Aberdeen, Aberdeen AB24 2TZ, Scotland, United Kingdom; and fState Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, People’s Republic of China

Contributed by Robert E. Ricklefs, April 7, 2013 (sent for review November 19, 2012)

Flight is a key adaptive trait. Despite its advantages, flight has been lost in several groups of birds, notably among seabirds, where flightlessness has evolved independently in at least five lineages. One hypothesis for the loss of flight among seabirds is that animals moving between different media face tradeoffs between maximizing function in one medium relative to the other. In particular, biomechanical models of energy costs during flying and diving suggest that a wing designed for optimal diving performance should lead to enormous energy costs when flying in air. Costs of flying and diving have been measured in free-living animals that use their wings to fly or to propel their dives, but not both. Animals that both fly and dive might approach the functional boundary between flight and nonflight. We show that flight costs for thick-billed murres (Uria lomvia), which are wing-propelled divers, and pelagic cormorants (Phalacrocorax pelagicus) (foot-propelled divers), are the highest recorded for vertebrates. Dive costs are high for cormorants and low for murres, but the latter are still higher than for flightless wing-propelled diving birds (penguins). For murres, flight costs were higher than predicted from biomechanical modeling, and the oxygen consumption rate during dives decreased with depth at a faster rate than estimated biomechanical costs. These results strongly support the hypothesis that function constrains form in diving birds, and that optimizing wing shape and form for wing-propelled diving leads to such high flight costs that flying ceases to be an option in larger wing-propelled diving seabirds, including penguins.

adaptive landscape | energetics | flight performance | morphology

Flight is a key adaptation that has evolved independently on many occasions (1). Despite the apparent advantages of flying, the ability to fly has been secondarily lost in several groups. Because a major advantage of flight is reduced extrinsic mortality (2), one hypothesis for the evolution of flightlessness posits that gains in efficiency in other locomotory modalities, such as diving, offset mortality risks in relatively safe environments. The high energy demands of flight also may be disadvantageous, particularly in habitats with low productivity (3, 4). The restriction of some terrestrial flightless birds to remote, predator-free islands with low productivity supports this hypothesis (3, 4). The reasoning seems less tenable for flightless diving seabirds that often exploit highly productive waters but are vulnerable to predation by seals, whales, and sharks. Moreover, many species of penguin travel long distances between their breeding and feeding grounds on a journey that could be made far more quickly by flying than by walking and swimming (5). An alternative biomechanical hypothesis suggests that flightlessness evolved in these birds because of a tradeoff in the optimization of wing-propelled locomotion in different media. In short, as wings become more efficient for swimming they become less efficient for flying, and vice versa. At some point, adaptations to increase swimming efficiency lead to the evolution of wings that would require physiologically unsustainable flight costs.

Species cannot cross this adaptive fitness valley without leaving flight behind.

Animals moving between different media face tradeoffs between maximizing function in one medium relative to the other (1, 6–9). Seabirds that dive and fly are an excellent example of animals facing tradeoffs in the optimization of locomotion for different media because they move between air, land, and water. Unlike diving beetles (family Dytiscidae), dippers (Cinclus spp.), and foot-propelled diving seabirds—the only other animals to routinely occupy all three media as adults—wing design in auks and other wing-propelled diving seabirds functions for both underwater and aerial locomotion (10–13). Extant flightless seabirds (penguins) evolved enhancements for underwater locomotion by reducing wingspan, enlarging wing bones, increasing body mass, optimizing muscle contraction rate for low-wingbeat frequencies, and augmenting myoglobin stores to increase dive endurance (1, 14, 15). In contrast, birds that both fly and dive, such as auks, are restricted by aerial flight demands for opposing adaptations (1, 12, 14).

We tested the biomechanical hypothesis for the evolution of flightlessness in seabirds by measuring the energy costs of flight and diving in two species of free-living, diving seabirds that are also able to fly: wing-propelled thick-billed murres (Uria lomvia) and foot-propelled pelagic cormorants (Phalacrocorax pelagicus). We predicted that murres would have elevated flight costs compared with nondiving birds, but would have low costs of swimming, although not as low as penguins, which have lost the ability to fly.

Results and Discussion

The compromises inherent in the tradeoffs in form for function in water and air were evident in murres, which have the highest wing loading (mass per unit area of the wing) of any flying bird. Our estimated flight cost for the 1-kg murre, 141 ± 18 W (146 W/kg or 0.83 J/Nm) of cost of transport; Fig. L4 and Figs. S1 and S2 surpassed the previous record of 135 W recorded in the 2.6-kg bar-headed goose (Anser indicus). Flight costs were also high for the 1.8-kg cormorant, 158 ± 51 W (87 W/kg or 0.70 J/Nm). The
murre value is more than double that for a 980-g bird predicted by the maximum output line proposed to represent maximum aerobic capacity in a flying bird (16). Presumably, the apparent limit to aerobic capacity simply reflects the narrow range of body plans of species measured to date (Fig. 1A)—specifically, the absence of measurements from species that also swim with their wings underwater (16).

Flight costs for murres were also 33% higher than expected from biomechanical modeling based on the murre body plan (Fig. S1), implying low muscle or mechanical efficiency. The high energy costs exhibited by murres in flight suggest why most animals use their wing locomotion in a single medium. In addition to having the highest absolute cost of sustained flight (Fig. 1A), murres also exhibit the highest flight costs in terms of multiples of basal metabolic rate (BMR; Fig. 1C). Because murres spend up to 5 h per day flying during reproduction (15), such a high level of sustained energy output (31x BMR) is remarkable; other vertebrates do not exceed ~20x BMR during intense activity (Fig. 1C).

In contrast to the high cost of flight, average daily energy expenditure of murres (25 ± 7 W or 5.7x BMR) and cormorants (23 ± 5 W or 4.1x BMR) during the chick-rearing period was well within the 7x BMR upper limit described for vertebrates (16). Thus, different mechanisms appear to limit maximum sustained energy expenditure over periods of days (e.g., internal properties related to heat dissipation and digestive or excretory abilities) compared with minutes (e.g., oxygen intake or use by muscles) (17). By altering time budgets, such that less time is spent doing energetically costly activities (flight), birds achieve an average daily energy expenditure that does not exceed the reported maximum and that is presumably limited by constraints that act over the scale of days, such as the amount of food that can be assimilated in a day.

Dive costs increase rapidly with body mass for flying divers, compared with nonflying divers (penguins; Fig. 1F). This fact, combined with the high flight costs of flying divers, likely explains the evolution of flightlessness in large, wing-propelled divers, such as penguins and extinct larger auks (15, 18). Flying, wing-propelled divers occupy a separate morphological space from other flying birds; foot-propelled divers occupy a space intermediate between flyers and wing-propelled divers (Fig. 1D and Fig. S3). Penguins are restricted to an entirely disjunct portion of the morphological space, well beyond the region occupied by flying, wing-propelled divers (Fig. 1D). The morphological differences in wing design are reflected in flight costs (Fig. S1). The space between flying and nonflying wing-propelled divers appears to represent a fitness valley in the adaptive landscape, with flying wing-propelled divers at the edge of the flying bird morphological space, supporting a biomechanical explanation for the origin of flightlessness in penguins. Within the auk family (Alcidae), the great auk (Pinguinus impennis) bounds the flightless edge of the valley, and murres bound the opposite, flighted edge of the valley (Fig. 1D). Some wing-propelled divers,
such as albatrosses and shearwaters, and plunge divers, like boobies and gannets, which have long wings, overlap with the morphological space of nondiving flyers. The underwater performance of these species is likely limited in terms of energy economy or achievable depth.

We speculate that great auks and flightless cormorants (Phalacrocorax harrisii) are closer to the flying bird morphological space because of their more recent evolution (19). For instance, the humerus of the great auk is flattened compared with that of other auks, but the wing area still consists largely of primary feathers and the humerus is not nearly as large as that of penguins (Fig. S4). In penguins, the wing has flattened into a whale-like flipper. The great auk and its closest extant relatives, razorbills (Alca torda) and murres, shared a common ancestor roughly 20 Mya, whereas the penguin order (Sphenisciformes) split from their closest extant relatives, flying seabirds in the order Procellariiformes—an order that also includes flying, wing-propelled divers—over 80 Mya (9).

The wing area of the great auk, ~58 cm², is similar to its closest flighted relatives, the razorbill (Alca torda; 42 cm², 0.72 kg) and murres (46 cm², 1.0 kg), although the great auk was much heavier (5 kg). Body size is more labile over evolutionary time than wing area—which in penguins has included remodeling the entire humerus—as exhibited by the rapid change in body size of flightless animals on recently colonized islands (3, 4, 10, 15).

Larger body size increases dive duration and efficiency (14, 15). Thus, loss of flight in diving birds appears to be followed by the rapid evolution of large body size, as shown by the presence of early giant penguins (20). Several observations support this scenario: (i) extant and fossil wing-propelled pursuit-diving birds are flightless above ~1 kg and capable of flight below ~1 kg (10, 15)—the largest auk and smallest penguin are both ~1 kg; (ii) dive costs increase more rapidly with body size in flying birds than in flightless birds (Fig. 1B); and (iii) the flying, wing-propelled diving bird morphological space is strongly tilted along discriminant axis 1, which represents body mass (Fig. 1D), suggesting particularly strong morphological tradeoffs for heavy birds that use their wings for both flying and diving. Furthermore, along with extant and recently extinct flightless cormorants, auks, penguins, and ducks, a fifth lineage of flightless birds (order Hesperornithiformes) from the Cretaceous were large and showed a temporal progression from flying to flightlessness accompanied by a progression from foot-propelled to wing-propelled diving (21).

Metabolic costs for murres freely diving in 5.9 ± 0.3 °C water declined with dive depth and duration, with the best explanatory model being an exponentially declining function similar to that obtained for free-diving penguins: (3.64 ± 1.33) \(\Sigma \left[1 - e^{-duration/1.23}\right]\) kJ (22). On average, murre dive costs were 27% lower than expected for a similarly sized foot-propelled diver, but 30% higher than expected for a similar-sized penguin (Fig. 1B). In contrast, foot-propelled cormorant dive costs were over threefold greater than expected for a similar-sized penguin (Fig. 1B). In general, flightless wing-propelled divers (penguins) have lower underwater energy costs and higher underwater efficiency than flying divers, whether wing or foot propelled (Fig. 1B). Presumably, “flying” underwater with long wings (murres) creates extra drag, whereas greater body mass (penguins) allows for insulation via adipose tissue rather than via a thick air layer in the feathers, reducing buoyancy costs in shallow water (23). In mammals, drag-based propulsion in semiaquatic animals is less efficient than lift-based propulsion in marine mammals (9, 24); foot-propelled avian divers use drag-based propulsion whereas wing-propelled divers rely on lift-based propulsion.

Unlike mammals and foot-propelled diving birds, many wing-propelled seabirds regularly exceed their calculated aerobic dive limit, the theoretical time limit for the exhaustion of oxygen stores during dives (25, 26). The calculated aerobic dive limit depends heavily on diving metabolic rate, which is usually measured in shallow dive tanks. Those calculations are problematic for deep-diving birds because metabolic rate likely decreases with dive depth due to reduced buoyancy, temperature, heart rate, and blood perfusion at depth (22, 23, 25–27). Using our exponentially declining model for oxygen consumption rate, and updating murre oxygen store estimates so that 90% of respiratory oxygen stores are used during the dive (as recently found in penguins) (25), leads to a calculated aerobic dive limit of 162 s, which is highly similar to the 150-s limit when surface pauses increase to allow for breakdown of accumulated lactate (25, 29). Thus, the low energy costs for deep-diving birds resolves the paradox that many deep dives exceed the calculated aerobic dive limit for birds using dive costs measured in shallow tanks (28, 29).

Large penguins have low metabolic costs during deep dives (21, 23), and our observation of hypometabolism in small, wing-propelled divers shows that the savings are not unique to large birds and may be characteristic of wing-propelled divers in general (Fig. 1B). Although mechanical costs during deep dives have been precisely measured and increase approximately linearly with dive depth (30–32), actual metabolic costs measured in the field decelerated as dive depth increased (Table 1). We suggest that physiological processes during the dive, such as oxygen store management and thermoregulation, are the dominant processes determining costs in wing-propelled divers diving to depths where buoyancy costs are minimal (22, 25, 27).

In contrast, dive costs for foot-propelled divers are correlated with dive depth and duration, reflecting biomechanical costs that increase linearly with dive duration (Table 1). In foot-propelled divers, the leg muscles are separate from the body core, whereas in wing-propelled divers, the breast muscles encase the body core, enabling them to finely regulate body core temperature, muscle temperature, and oxygen availability in the muscles to minimize costs during deep dives (22, 25, 27). Progressive reductions in dive costs may explain why aquatic birds developed wing-propelled diving and finally flightlessness in response to foraging opportunities at increasing depths. In particular, high flight costs and low dive costs in auks illustrate why penguins, and other diving birds, have lost the ability to fly.

Materials and Methods

Between July 15 and August 10, 2006, we captured 41 thick-billed murres at the Coats Island west colony, Nunavut, Canada, and equipped those birds with time-depth temperature recorders (TDRs). Between July 25 and August 14, 2012, we captured 22 pelagic cormorants at Middleton Island, Alaska, and equipped those birds with temperature-depth accelerometers. We simultaneously measured energy expenditure in the same birds using doubly labeled water and recorded weather. When murres injected with doubly labeled water were equipped with TDRs, we compared paired control murres equipped at the same time and breeding stage only with TDRs, there was no difference in maximum dive depth (\(t_{24} = 0.40, P = 0.69\), average dive depth (\(t_{24} = 0.99, P = 0.33\), time spent submersed (\(t_{24} = 0.71, P = 0.48\), time spent flying (\(t_{24} = 0.32, P = 0.75\), time spent at the colony (\(t_{24} = 0.35, P = 0.96\), or time to switch-over (\(t_{24} = 1.83, P = 0.08\). We concluded that the doubly labeled water injections had little impact on behavior.

To calculate activity-specific metabolic rate, we regressed energy expenditure against activity times for diving, flying, at the water surface, and on land. By using a multivariate approach and forcing the intercept to be zero (no energy was expended when no time elapsed), we overcame the problem of incorporating some of the slope value into the intercept. We subtracted the high cost of preening from cormorant activities. Because our depth recorders provided more details on time partitioning during the dive, we also considered three other models: (i) cost of diving was proportional to costs associated with buoyancy; (ii) cost of diving was proportional to total mechanical work during a dive; and (iii) cost of diving followed the oxygen depletion curve developed for deep-diving penguins. We compared the effectiveness of different models using Akaike’s information criterion, which penalizes models with increased numbers of parameters without improvement in fit. We compared the morphological space of flying birds and flightless divers by conducting a discriminant analysis on the first principal component of the log-transformed morphological traits (wing area, wingspan, body mass, and three functions derived from those...
Table 1. Comparison of models that estimate energy expenditure in watts (± SE) for several activities in thick-billed murres and pelagic cormorants

| At the colony Flying Water surface Diving Other ΔAIC |
|-----------------|-------|---------|---------|---------|--------|
| **Thick-billed murre** |       |         |         |         |        |
| 9.2 ± 3.1        | 141 ± 18 | 26 ± 6  | (1.01 ± 0.36) ![Equation](\(1 - e^{-\text{duration}/0.23}\)) | Wind = −16 ± 15 kJ/h/km | 0.00 |
| 11.4 ± 3.3       | 143 ± 18 | 29 ± 7  | (0.93 ± 0.37) ![Equation](\(1 - e^{-\text{duration}/0.23}\)) | Mass loss = 2.57 ± 6.89 kJ/d/g | 1.10 |
| 9.7 ± 3.1        | 140 ± 18 | 28 ± 7  | (0.99 ± 0.37) ![Equation](\(1 - e^{-\text{duration}/0.23}\)) | Temperature = −1.5 ± 8.1 kJ/°C | 1.45 |
| 8.6 ± 5.8        | 140 ± 18 | 27 ± 7  | (1.09 ± 0.44) ![Equation](\(1 - e^{-\text{duration}/0.23}\)) | Sex | 1.89 |
| 7.2 ± 8.1        | 140 ± 15 | 27 ± 6  | (0.80 ± 0.26) ![Equation](\(1 - e^{-\text{duration}/0.23}\)) | Mass = 0.30 ± 0.83 kJ/g | 1.92 |
| 7.5 ± 6.1        | 140 ± 18 | 24 ± 9  | (1.01 ± 0.37) ![Equation](\(1 - e^{-\text{duration}/0.23}\)) | All activities | 6.82 |
| 9.2 ± 3.1        | 143 ± 19 | 27 ± 6  | (0.95 ± 0.39) ![Equation](\(\Sigma \ln(\text{depth})\)) | Away from the colony | 31.47 |
| 8.9 ± 3.1        | 148 ± 18 | 28 ± 6  | 27 ± 12 | 3.09 |
| 9.4 ± 3.3        | 158 ± 18 | 29 ± 6  | (3.1 ± 1.9) mechanical costs | 6.82 |
| 7.5 ± 3.3        | 165 ± 18 |         |         |         |        |

| **Pelagic cormorant** |       |         |         |         |        |
| 7.8 ± 2.5          | 168 ± 51 | 7.7 ± 12 | 66 ± 11 |         | 0.00 |
| 6.4 ± 2.9          | 180 ± 52 | 3.4 ± 13 | 75 ± 15 |         | 0.92 |
| 4.4 ± 19.6         | 174 ± 56 | −4.0 ± 22.4 | 57 ± 18 |         | 1.50 |
| 8.2 ± 3.8          | 170 ± 55 | 10 ± 12 | 68 ± 13 |         | 1.81 |
| 11 ± 4.5           | 175 ± 68 | 11 ± 9 | 68 ± 12 |         | 1.95 |
| 8.4 ± 3.1          | 167 ± 93 | 194 ± 72 | (0.15 ± 0.06) ![Equation](\(1 - e^{-\text{duration}/0.89}\)) |         | 7.12 |
| 5.2 ± 5.1          | 168 ± 9 | 53 ± 20 | (0.010 ± 0.004) ![Equation](\(\Sigma \ln(\text{depth})\)) |         | 23.60 |
| 5.1 ± 5.1          | 168 ± 9 | 53 ± 20 |         |         | 26.39 |

Models include different functional relationships between energy costs and dive depth or duration and terms for individual body mass, sex, average ambient air temperature, mass loss during the deployment period, and average wind speed. Water surface values for cormorants exclude preening. Models with ΔAIC values > 2 are generally considered inferior.

12. Sato K, et al. (2007) Stroke frequency, but not swimming speed, is related to body size and morphological analyses, details of statistical modeling, and raw data.


SI Materials and Methods

Our observations were made at the Coats Island west colony (62°57′N, 82°00′W), Nunavut, Canada, July 15 to August 10, 2006 (n = 40). Parental thick-billed murres (Uria lomvia) were caught at their nests with a noose pole. All birds were placed in a cloth bag and weighed using a Pesola spring balance (±10 g) during each capture period. Handling time, including bleeding time, was always less than 10 min and usually less than 5 min. In addition to doubly labeled water experiments, during each experimentation period 5–10 birds were simultaneously equipped with time/depth/temperature recorders (TDRs), but not injected or blood-sampled to examine the effect of these activities on murre time budgets.

TDR Observations. We attached Lotek LTD 1100 (Lotek Marine Technologies, Inc.) TDRs to plastic bands attached to the legs of murres. The TDRs were cylindrical (mass = 4.5 g; diameter = 1 cm; length = 3.3 cm; sampling interval = 3 s; depth precision = ±0.1 m; accuracy = ±2 m; 128-kB memory, recording continuously for 55 h). Back-mounted TDRs are known to impact murre provisioning rates, number of foraging trips, adult attendance, mass loss, and dive depth and duration. Though heavier, our leg-mounted TDRs had no impact on provisioning rates, trip duration, or mass loss (1–9). We used the pressure log from the TDR to determine time spent under water, and the temperature log from the TDR to determine time in air, on land, and on water (7–9). In 2009, as part of a separate project, we attached TDRs and accelerometers to the same individuals. Activity budgets obtained from the two different instruments agreed within 3% for each activity.

Doubly Labeled Water. During the incubation and chick-rearing periods (July 18 to August 10, 2006), we injected intraperitoneally either 0.5 mL or 1.0 mL of doubly labeled water (50% H2 18O and 25% D2O; see data table deposited in the Dryad database for exact enrichment) ~1 cm deep into the brood patch of 51 parental murres using a 27-gauge needle and a 1-mL syringe. Eight of the birds were weighed and killed after 240 min. The entire body was then freeze-dried to determine the body water content. Using those data, we were able to directly measure percent body water and show that it closely corresponded with the plateau method used for doubly labeled water experiments when values obtained 90–180 min after injection (10). The remaining birds were released following injection, and immediately returned to their breeding site. We failed to recapture 13 of the 51 birds within 90–180 min of injection. Equilibrium isotope values for these 13 birds were found using a regression between 18O/deuterium and body mass for the 38 birds yielding plateau (initial) blood samples (R2 = 0.98).

Initial blood samples were taken from the tarsal vein. Final blood samples were obtained 48 h (n = 39), 72 h (n = 7, including six birds sampled at 48 h), 96 h (n = 5, including three birds sampled at 48 h), and 120 h (n = 1) after the initial blood sample, and were taken from the brachial vein. Background blood samples were obtained from the tarsal vein in a separate group of 14 birds (seven on July 14 and seven on August 5) to avoid entering the brachial vein multiple times. The average background concentrations were 1993.4 ± 0.5 ppm of 18O and 1529.2 ± 0.5 ppm of deuterium, with no difference between dates (P > 0.6). All blood samples were collected into one, two, or three 60-μL capillary tubes that were flame-sealed and sent to the University of Aberdeen for laboratory processing after the field season. Only samples with final measurements above 20 ppm of 18O were used (this cutoff is 10 SD above background; deuterium depleted at a slower rate and was consequently always >10 SD above background during final measurements when 18O was >10 SD above background) (11). Final blood samples (F1 samples) >10 SD above background were obtained from 41 birds (of 43 birds that were injected—one bird was recaptured but not blood-sampled, and another was not recaptured), including 1 bird at 24 h, 38 at 48 h, 1 at 72 h, and 1 at 96 h. Second final blood samples (F2 samples) >10 SD above background were obtained from 10 birds, including 1 at 48 h, 5 at 72 h, and 4 at 96 h. For average estimates of murre energy consumption across the season or within breeding periods, single values for each individual (either F1 values or F2 values, where available) were averaged across all individuals.

We used a value of 0.809 for the respiratory quotient (25 kJ/L CO2) based on a diet of 85% protein, 10% lipids, and 5% carbohydrates, as directly measured via nesting and adult dietary sampling (12). In contrast, the respiratory quotient for 11 thick-billed murres fasting for 4 h at the colony (i.e., using primarily lipids) measured using open-flow respirometry was 0.70 ± 0.05. Thus, it is unlikely that we underestimated the average respiratory quotient across activities, and therefore it is unlikely that daily energy expenditure was overestimated.

Activity-Specific Metabolic Rate. Complete activity records (i.e., birds recaptured before TDRs stopped recording) were obtained for 38 F1 records and four F2 records. Three of those birds provided both F1 and F2 measurements. Thus, our sample size for independent measurements of activity and metabolic rate was 39. We completed a second set of analyses (n = 42) that included the second set of F2 measurements as independent samples, assuming that most of the variation in metabolic rate is due to variation in activity rather than individual variation in activity-specific metabolic rates. Because we were interested in activity-specific metabolic rates (and had measured activity independently), we assumed that activity-specific metabolic rates did not vary significantly with time of day (i.e., we assumed that diel variation in energy expenditure was related primarily to diel variation in activity) and did not adjust values that were collected up to ±3 h from 24-h cycles.

To calculate activity-specific metabolic rates, we modeled total energy expenditure by the equation

\[
EE = DMR \times T_d + FIMR \times T_f + RMR_{ac} \times T_{ac} + RMR_{ab} \times T_{ab},
\]

where EE is energy expenditure during the sampling period (measured by doubly labeled water), DMR is diving metabolic rate, FIMR is flying metabolic rate, RMRab is resting metabolic rate on the water, RMRac is resting metabolic rate on land, Td is time spent diving, Tf is time spent flying,谭 is time spent on the water, and T_{ab} is time spent at the colony (the latter four measurements derived from TDR records). The metabolic rates for each activity were determined using a multiple linear regression to provide DMR, FIMR, RMRab, and RMRac. The values are therefore averages within each activity, and each includes relatively energy-intensive periods. For example, preening on the water would be included in RMRab, preening at the colony in RMRac, and active prey chasing in DMR. We also considered three other variations on the basic model, because dive costs are known to be nonlinearly related to dive duration and depth. By including all activities within the model and forcing an intercept of zero, we avoided underestimating activity-specific metabolic rates, which can occur when examining the slope of energy expenditure against each activity separately (13).
First, we considered a model where dive costs were proportional to the mechanical costs associated with overcoming buoyancy (14) rather than total dive time. To model those costs we used the equation

\[
EE = \eta \times \sum \text{mechanical costs} + \text{FIMR} \times T_i + \text{RMR}_d \times T_w + \text{RMR}_a \times T_w ,
\]

where \(\eta\) is the efficiency of converting metabolic energy into mechanical energy (estimated as a coefficient in the multiple regression) and \(\Sigma \text{Ln(dive depth)}\) is the summation taken over all dives for the relevant period and represents mechanical buoyancy-related costs. Work done to overcome buoyancy is proportional to the expression \(\text{Ln(dive depth)}\) up to a constant \(C\), which was obtained assuming a surface buoyancy of 0.518 L (total buoyancy including air) as measured previously for thick-billed murres (15). We considered equations with and without the constant; we allowed the equation’s intercept to vary to allow for the constant \(C\).

Second, we considered a model where dive costs were proportional to the mechanical costs measured by accelerometers, and including drag, inertial, and buoyancy (5, 15) rather than total dive time. To construct that model, we used the equation

\[
EE = \eta \times \sum \text{mechanical costs} + \text{FIMR} \times T_i + \text{RMR}_d \times T_w + \text{RMR}_a \times T_w ,
\]

where \(\eta\) is the efficiency of converting metabolic energy into mechanical energy (estimated as a coefficient in the multiple regression) and \(\Sigma \text{mechanical costs}\) is the summation taken over all dives for the relevant period for the predicted mechanical costs associated with diving to the depth of each dive. We estimated mechanical costs in 2009 by attaching accelerometers to 10 murres. Mechanical costs were estimated as the summation of buoyancy, inertial, and drag forces over each wing stroke (see ref. 15 for details of calculation and estimation of drag forces). The depth-specific value for the mechanical component of diving matched that obtained from the literature (5, 15). The model assumed that the mechanical costs for diving were primarily associated with descent.

Third, we considered a model where dive costs were determined by rate of oxygen use in the air sacs, with oxygen consumption rate declining exponentially through the dive because of shunting of blood away from nonessential organs. We used the functional relationship (an exponential model) described for penguins (16) and modeled the relationship using the equation

\[
EE = \eta \times \sum (1-e^{-t/(1.23)}) + \text{FIMR} \times T_i + \text{RMR}_d \times T_w + \text{RMR}_a \times T_w ,
\]

where \(t\) is dive duration and \(\eta\) is the conversion factor from oxygen consumption rate to watts (energy expenditure rate); note that the equation keeps the form of oxygen utilization during the dive described by Knower Stockard (16), but the coefficient representing the absolute rate of oxygen consumption is subsumed into \(\eta\). The summation was taken over all dives for the relevant period. The functional relationship assumed that the oxygen depletion rate followed the form described for deep-diving emperor penguins (Aptenodytes forsteri), the only deep-diving species where the relationship has been directly measured (figure 10 in ref. 16). We altered the time constant for emperor penguins from 1/2.91 (figure 10 in ref. 16) to 1/1.23 because the maximum dive duration for murres is ~4.6 min (6), compared with 11 min for emperor penguins (16). Thus, oxygen stores in murres are used up approximately at 11/4.6 = 2.39 times the rate of emperor penguins, and we adjusted the time constant to represent that change in rate (2.39/2.91 = 1/1.23). Because we were able to model directly the nonlinear change in oxygen consumption during the dive based on ref. 16 measurements, we did not need to combine oxygen consumption rates over the course of the entire bout (17).

We compared the effectiveness of different models using Akaike’s information criterion (AIC), which penalizes models that increase the number of parameters without improvement in fit. Fig. S2 shows the expected shape of each of the four different models for diving metabolic rate.

To compare measured flight costs with mechanical flight costs predicted from aerodynamic theory, we calculated the mechanical flight costs for thick-billed murres using the program Flight.EXE 1.24 (18). We used the values 1 kg for body mass, 0.71 m for wingspan, and 0.051 m² for wing area. Results are shown in Fig. S3.

**Weather and Controlling for the Effect of Temperature.** Wind speed and ambient temperature were measured at sea level at the colony each day at 0800 and 2000 hours during the doubly labeled water experiment. Wind speed and temperature were also measured hourly at the Environment Canada weather station at the Coral Harbor airport, 100 km to the northwest. The foraging range of murres from the Coats Island colony encompasses the region bounded the foraging area used by murres in this study (19). We used the temperature logs from the TDRs to determine average temperature during flight and diving. We averaged both wind speed and temperature across each deployment and examined the regression of wind speed and temperature on energy expenditure. We also examined the residual of energy expenditure rate for each activity (i.e., after accounting for the other activities) and regressed that against wind speed and temperature.

To compare our values against other estimates of diving metabolic rate obtained for other species at other temperatures, we used linear relationships to convert all measured metabolic rates to the equivalent metabolic rate at 13 °C, following the protocol established by Enstipp et al. (figure 7 in ref. 20). We followed Enstipp et al. (20)—who actually chose 12.6 °C—because more previous studies used that water temperature than any other temperature; it is often equivalent to the typical temperature in a shallow pool used for respirometry, and therefore required transforming the fewest data. Diving metabolic rate decreases linearly with increasing temperature in endotherms diving in cold water (20–22), and we used linear relationships established for each taxonomic group to interpolate the value at 13 °C for those species that had not been previously measured at 13 °C. We also removed one early data point for African penguins (Spheniscus demersus) that had been flagged by other authors as too high (23).

**Statistical Analyses.** All statistical analyses were completed in R 3.2.1. Before using parametric statistics, we tested for normality (Shapiro–Wilk test) and homogeneity of variance (Levine’s test). All percentages were arcsine-transformed before analysis. We only analyzed dives with maximum depth >3 m due to imprecision inherent in our TDRs. Body mass does not vary substantially in breeding murres (±15% from average value), and >90% of that variation occurs in metabolically inert lipid stores and total body water (10), so we did not expect a strong impact of body mass on metabolic rate. Nonetheless, we included body mass and body mass loss as covariates in the model. All values reported are means ± SE.

**Are Our Values Realistic?** Because past researchers have claimed that using doubly labeled water to measure activity-specific metabolic rates underestimates activity costs (e.g., ref. 13 criticized a different mathematical approach than the one we used),
we carefully considered whether our values obtained were realistic. First, we used a multiple regression rather than a simple regression to determine activity-specific metabolic rates (13). Simple regression tends to underestimate activity-specific metabolic rates by assuming that the intercept represents average activity-specific rates when none of the particular activity occurs; if different activities are intercorrelated, the slope can lead to erroneous measurements using simple regression. Second, we directly confirmed equations relating isotopic dilution to total body water content, and we obtained dietary samples from dozens of individuals so that we have a good representation of the respiratory quotient. Because energy expenditure measured via doubly labeled water is linearly (and mathematically) proportional to percent body water and respiratory quotient, by directly measuring those parameters in murres (10) we were able to reduce the absolute error in our energy expenditure measurements; most past studies that found discrepancies between doubly labeled water and direct respirometry did not measure percent body water directly (11). Third, we compared our values to those obtained on captive thick-billed murres diving in shallow dive tanks (21).

Croll and McLaren (21) measured metabolic rates of ~8 W for postabsorptive murres in the thermoneutral zone and resting in air, 28 W for murres preening in the thermoneutral zone, and 19 W for postabsorptive murres resting in water 0–5 °C (with an elevation to 24 W assuming additive effects of digestion and temperature, as much of the time resting on the surface in murres is believed to be associated with digestion). In comparison, our values of 9.2 W for resting on land and 26 W for resting on the water surface are quite similar, especially given that our measurements also included activities (such as preening). Dividing metabolic rate in the thermoneutral zone for murres diving in shallow dive tanks was 21 W (21). Assuming a 14-W increase associated with diving in 0–5 °C water (21), the total diving metabolic rate in cold water was 35 W for an average dive duration of 41 s. In contrast, our value in the model assuming dive costs were proportional to dive time was 27 W, and our value (at 41 s) for the model assuming dive costs decreased exponentially with dive time was 50 W at 41 s, 40 W at 68 s (the average dive duration for our study population), and 13 W at 280 s, the maximum duration recorded for our study population (9). The differential effects of buoyancy in a shallow dive tank, the potential substitution of heat from exercise for thermoregulation, and the associated costs of returning peripheral body temperature to core temperature values meant that it was difficult to directly compare the diving metabolic rate from the earlier study (21) with our own, but values seemed to be relatively similar. Because there have been no past measurements of flight costs, we could not make comparisons between our costs and those measured previously.

**Morphological Analyses.** We use a multivariate approach to describe and compare the morphological space of 452 species of diving and nondiving birds (see ref. 24 for a similar approach). We collected data on wingspan, wing area, and body mass from two sources (ref. 25 and data appendix in ref. 26). We calculated aspect ratio (wingspan squared/wing area), wing loading (body mass/wing area), and a parameter we call the Pennycuick-induced drag coefficient ($P_{ind}$). $P_{ind}$ is the induced drag coefficient from the Pennycuick model (18) divided by body mass:

$$ P_{ind} = \frac{2.4 \text{ body mass}}{1.23 \pi \text{ wingspan}^2} $$

![Equation S5](https://www.pnas.org/cgi/content/short/1304838110)

The main difference between $P_{ind}$ and wing loading is that $P_{ind}$ uses the square of wingspan rather than wing area (18). Wingspan is a simpler parameter to measure than wing area. Because many of these parameters are intercorrelated and/or constructed from one another, we log-transformed each of the six parameters (body mass, wing area, wing-span, wing loading, aspect ratio, and $P_{ind}$) and performed a principal components analysis (PCA) to reduce the parameters to a single set of three orthogonal vectors that described over 99% of the morphological space. We used a multigroup discriminant analysis and a multivariate analysis of variance to determine whether five groups could be distinguished: flightless divers, plunge divers, foot-propelled divers, wing-propelled divers, and nondiving flyers. We classified shallow, wing-propelled divers, such as albatrosses and shearwaters (27), as nondiving flyers. A separate analysis that included both of those two groups showed no morphological differences between those two groups and other nondiving flyers.

In the PCA (Fig. S3E), PC1 explained 84.9% of the variation, PC2 explained 14.0% of the variation, and PC3 explained 1.1% of the variation. We ignored all further principal components because PC1–PC3 explained 99.99% of the variation. In brief, PC1 represented variation in body mass, whereas PC2 represented variation associated with the variation in wing area that is independent of body mass. Thus, wing- and foot-propelled divers (flightless or flying) were distinguished from nondivers primarily along PC2.

The discriminant analyses on PC1–PC3 correctly classified 100% of penguins, 96% of flyers, 93% of flying, wing-propelled divers, 46% of foot-propelled divers, but none of the plunge divers or shearwaters (Fig. S3F). A multivariate analysis of variance was highly significant ($F_{15,993} = 31.0, P < 0.0001$) with Tukey’s post hoc tests showing no difference between shearwaters and flyers ($p_{adj} = 0.70$) or plunge divers and flyers ($p_{adj} = 0.94$), but significant differences among flyers and foot-propelled divers, flying wing-propelled divers, and penguins (all $p_{adj} < 0.0001$). Shearwaters, plunge divers, and flyers did not differ significantly from one another (all $p_{adj} > 0.5$), whereas foot-propelled divers, flying wing-propelled divers, and penguins did differ significantly from one another ($p_{adj} < 0.001$). We used the statistical similarity of shearwaters, plunge divers, and flyers as further justification for considering all three as flyers. The discriminant analyses on PC1–PC3, with shearwaters and plunge divers categorized as flyers, correctly classified 100% of penguins, 98% of flyers, 93% of flying, wing-propelled divers, and 52% of foot-propelled divers (Fig. S3C). A multivariate analysis of variance was highly significant ($F_{27,15993} = 46.4, P < 0.0001$) with Tukey’s post hoc tests showing significant differences among flyers and foot-propelled divers, flying wing-propelled divers, and penguins (all $p_{adj} < 0.0001$).

We then correlated the morphological data (PC2 and $P_{ind}$) for species using primarily flapping flight against the residual of flight costs on the maximum output line (from Fig. 1A) for species using primarily flapping flight. We used the maximum output line as the expected value for flapping flight to account for body mass, because many birds using flapping flight cluster along that line (Fig. 1A). We chose $P_{ind}$ because of its close connection to aerodynamic theory.

**Scenario for the Evolution of Flightlessness in Diving Birds.** Large body size allows for increased dive duration via (i) reduced mass-specific oxygen consumption rate; (ii) increased volume of oxygen stores; and (iii) increased density of oxygen stores (myoglobin concentrations) (28, 29). At the level of the individual, there is therefore a strong selective pressure for greater body size in diving birds, allowing for exploitation of deeper prey and more efficient diving (more time underwater for each surface pause) at all dive depths (28). In flying birds, those selective pressures are
balanced by selective pressures for reduced body mass to reduce flight costs (18). Species that nest on predator-free islands close to consistent, dense prey aggregations can become flightless and evolve higher body mass (27).

Flightlessness in diving birds appears to be associated with rapid evolution of large body size, as shown by the presence of early giant penguins. In support of this scenario, (i) extant and fossil wing-propelled diving birds are flightless above ~1 kg, and flying below ~1 kg; (ii) dive costs increase more slowly with body size in flightless than in flying birds (Fig. 1B); and (iii) the flying-wing-propelled diving bird space is strongly tilted along discriminant axis 1, which represents body mass (Fig. 1D), suggesting particularly strong morphological tradeoffs for heavy birds that use their wings both for flying and diving. We therefore speculate that once the benefits of deep-diving outweigh the benefits of flying, populations rapidly evolved larger individual body size, leading to increasing benefits in terms of dive efficiency. Development of a wing and feather structure optimized for underwater locomotion would have taken longer evolutionary time, partly because remodeling the wing may have required alterations in the kinematics of underwater flight. In contrast, flightless foot-propelled divers that have little use for wings (e.g., such as flightless cormorants Phalacrocorax harrisi, 3–5 kg) are only slightly larger than flying cormorants (e.g., usually 1–3 kg; Fig. 1B) and lack highly remodeled wing bones, but have very reduced primary feathers. We propose that flightlessness in wing-propelled diving birds therefore involves (i) inefficient foot-propelled or wing- and foot-propelled swimming at shallow depths among the first small, shallow-diving seabirds; (ii) the evolution of increased body size, reduced wing size, and increased dive efficiency in flying birds as they move toward the boundary of the adaptive valley in morphological space; (iii) the rapid “jump” to flightlessness, permitting an increase in body size and resulting dive efficiency (the great auk); and (iv) the slow remodeling of the wing architecture (bone and feathers) toward a form optimized for swimming.

Fig. S1. Models used to estimate biomechanical costs and their comparisons with morphology. (A) Comparison among four different models describing how metabolic rate changes with dive depth. Buoyancy-related costs and total mechanical costs are represented as mechanical work (left scale) and converted to metabolic energy (right scale) assuming a constant efficiency. The exponentially declining and constant metabolic rate models are represented directly in metabolic energy consumed (in kilojoules, right scale). Because the data themselves generate the coefficient associated with converting the different functions into metabolic rates, the shapes of the functions rather than their absolute values are being compared. (B) Measured flight costs across 29 bird species compared with flight costs estimated from aerodynamic theory for the same bird species (Pennycuick model, as generated by the program Flight 1.24). The thin line represents the least-squares linear regression, which is curvilinear on the log-log graph. The solid line is the 1:1 line: the value if measured values were equal to the values predicted by the Pennycuick model. Murres had the highest absolute residual from the best-fit linear trend line (65 W) and, among large birds, had the highest absolute residual from the 1:1 line (33 W). Relative deviation from maximum output line [(measured value – expected value from maximum output line)/measured value] increased with both (C) $P_{ind}$ and (D) PC2 (Fig. S4).
Fig. S2. Cost of transport for different modes of locomotion for birds and mammals. Open symbols represent animals that face functional tradeoffs with movement in different media: flying wing-propelled birds (when flying), human swimmers, sea otters and flying, diving birds (when swimming) and penguins and geese (when running).
Fig. S3. (Continued)
Fig. S3. Comparison of morphology across bird species. (A) Wing area; (B) wing-loading; (C) aspect ratio; and (D) \( P_{\text{out}} \) as a function of body weight for bird species. (E) PCA of log-transformed avian morphology. The 95% confidence ellipses are shown for each group. The lengths of the arrows representing morphological variables were divided by 7.5 for ease of presentation. Penguins are excluded from wing-propelled divers. (F and G) Discriminant analysis of log-transformed avian morphology, with Puffinus shearwaters and plunge divers treated as separate groups (F), and Puffinus shearwaters and plunge divers treated as flyers (G). *Murre* and *Great auk* are printed immediately below the appropriate point.
Fig. S4. Humerus and wing of (Left) thick-billed murre; (Center) great auk; and (Right) Galapagos penguin. The value of PC2 (Fig. S3E) is low for the murre and high for the auk and penguin.