Supplementary Material for

Morphology, muscle capacity, skill, and maneuvering ability in hummingbirds

Roslyn Dakin, Paolo S. Segre, Andrew D. Straw, Douglas L. Altshuler*

*Corresponding author. Email: doug@zoology.ubc.ca

Published 9 February 2018, *Science* **359**, 653 (2017)
DOI: 10.1126/science.aao7104

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/359/6376/653/suppl/DC1)

Table of Contents

Materials and Methods

Maneuvering trials

We studied the maneuvering performance of 25 neotropical hummingbird species at four capture sites with distinct avian communities: the Los Amigos Biological Station, Peru (-12.568983ºS, -70.100132ºW, 280 m elevation) in September-October 2010; the La Selva Biological Station, Costa Rica (10.431254ºN, -84.003600ºW; 60 m elevation) in February-March 2012; the Cerro de la Muerte, Costa Rica (9.559141ºN, -83.725318ºW; 3100 m elevation) in April 2012; and the Hacienda Guaytara, Ecuador (-0.466286ºS, -78.307731ºW, 3523 m elevation) in June-July 2013. All research was approved by the Institutional Animal Care and Use Committee of the University of California, Riverside, the University of British Columbia Animal Care Committee, the Ministerio de Agrigultura of Peru, the Ministerio de Ambiente y Energia of Costa Rica, and the Ministerio del Ambiente of the Pichincha province of Ecuador.

For maneuvering trials, resident hummingbirds were captured with mist-nets and identified using plumage and morphological traits (Los Amigos: 53 individuals from 8 species; La Selva: 70 individuals from 8 species; Cerro de la Muerte: 54 individuals from 4 species; Hacienda Guaytara: 36 individuals from 5 species; note that each species was captured at only one site). We released any subadult individuals and individuals that could be definitively identified as female, except for 6 female *S. flammula* at Cerro de la Muerte. These definitive females are omitted from further analysis because we were not able to account for sex differences in any of the other species, as they were either all males, or were individuals from sexually monomorphic species. Note that because this is a comparative study, the sample sizes were not predetermined; instead, we selected sites that are known to have distinct bird communities and sought to collect as many unique species at each site and as many individuals as possible. Therefore, the sample sizes for each species vary depending on the rate of capture and our ability to run trials within the time available.

After capture, birds were housed in netted enclosures (0.6 m x 0.6 m x 0.9 cm, Live Monarch, Boca Raton, FL) supplied with perches and hummingbird food *ad libitum* (Nektar-Plus, Pforzheim, Germany). When possible, birds were housed separately; however, if necessary, birds of different species were kept together after verifying that they would not exclude each other from the feeders. Hummingbirds were housed until they demonstrated typical perching and flight behavior in captivity, at which point they were recorded in the maneuvering trials.

For the maneuvering trials, each individual was recorded alone in a 3 m x 1.5 m x 1.5 m flight chamber (Fig. 1A). The birds were motivated to fly by their captive environment, use of the feeder, and perception of environmental stimuli (e.g., the presence of their captors as well as sounds from outdoors). A previous study using the same method showed that for nearly all performance metrics, the maneuvering performance of birds flying alone in the chamber was at least as great as that of birds flying in the presence of a competitor (*17*). This demonstrates that our assay captures a broad range of performance in the flight chamber that is, in general, not lower than that of

vigorously competing birds in the same environment (*17*). Body position and orientation were automatically recorded at 200 f.p.s. using a computer vision multi-camera tracking system (*16*, *17*, *25*). After completing the study, we conducted additional analyses described below under *Correlation analyses* to verify that the chamber was large enough to capture a wide range of maneuvers.

Trials began between 0715 and 1715 hours local time. The flight chamber was made of garden mesh and included multiple perches and one artificial feeder with *ad libitum* sucrose in water (wt/vol = 225 g/1 L). We used the first 30 minutes of flight recorded from each trial to evaluate maneuvering performance, after verifying that performance of the different maneuvers did not change over time during a 2-hour period in a previous study (*17*). We chose this duration because it yielded a large number of maneuvers of each type (average sample sizes in this study ranged from 45-435 maneuvers per bird; Fig. 1B).

After completing the maneuvering trial, we recorded each bird's body mass and took photographs of its left and right outstretched wings (*17*, *25*). Birds were then marked with an identification band on the leg and released.

Morphology and muscle capacity

We measured the area and aspect ratio (AR) of each bird's left and right wing from photographs. Following Ellington's definition (*43*), we measured AR for each wing as $4R^2/S$ where R is the length of one wing and S is twice the area of that wing. We performed this calculation separately for each wing, and then took the average of the left and right wings as a bird's AR. Wing loading was defined as body mass divided by the sum of the area of the two wings. We chose to analyze wing loading over other potential measures of wing size because of its relevance to aerodynamic theory and previous empirical work (*1*, *44, 45*), and because we wished to interpret the effects of wing loading separately from body mass as these traits vary independently (fig. S1D).

Some of the species we studied had other exaggerated traits that may influence flight, including elongated tails (e.g., *Lesbia victoriae;* 5 *Phaethornis* spp.), elongated and decurved bills (2 *Glaucis* spp.; 5 *Phaethornis* spp.; 2 *Threnetes* spp.), feathers that are modified to produce sound (*Selasphorus flammula (46)*), or otherwise modified plumage (e.g, tufted plumage on the legs and face in *Eriocnemis luciani* and *Colibri thalassinus*; forked tails in *Thalurania furcata*). These modifications could have diverse effects on flight, both aerodynamically, and because the exaggerated traits may have coevolved with particular display maneuvers in which traits and sounds are displayed. Given the variable expression of these traits among different lineages, their effects may be taxon specific. We consequently focused our analyses on three morphological traits expected to influence performance that are common to all birds: body mass, wing AR, and wing loading.

We have previously found that an individual's muscle capacity can explain its performance on several different maneuvers (*17*). Muscle capacity is measured using a load-lifting assay that requires a bird to lift successively more beads during its natural

escape response (fig. S2), eliciting 200-400% of the mechanical power output of hovering flight (*32*, *46*, *47*). To obtain species-typical values for load-lifting, we used a large database of 263 individuals from load-lifting studies (*32*, *48*) (see *Data processing* below for further details). Due to time constraints, we were only able to obtain load-lifting data from 18 out of 207 individuals in the maneuvering trials. While it would have been preferable to have data on individual-level variation in muscle capacity from all of the birds in the maneuvering trials, we were limited by the challenges of simultaneously mist-netting and monitoring captive individuals from sensitive tropical species. Therefore, given that our main goal was to test among-species variation, we prioritized our efforts toward increasing statistical power via the sample size of maneuvering trial birds, rather than obtaining individual-level load-lifting from the same individuals.

We also used the database of 263 individuals above to obtain species-average values for wing morphology and body mass for our 25 study species (fig. S1). It was important to use a different sample for species-level analysis of these traits, because when sample sizes are small within species, sampling error can cause an intraspecific (within-species) trend to appear to be an interspecific (among-species) trend, even if the species do not actually differ. For example, suppose 5 individuals are sampled from each of two species, and suppose all 5 individuals from one species happen to have relatively low muscle capacity relative to their unsampled conspecifics, whereas all 5 of the individuals from the other species happen to have relatively high muscle capacity relative to their unsampled conspecifics. Then we might misinterpret this as a species difference, when in reality all of the muscle capacity variation is within-species. To guard against this possible confound, we used a more conservative approach to our analysis for species-level effects, such that we used different individuals to obtain species-average values for biomechanical traits and maneuvering performance whenever possible. As a result, there was only slight overlap of 18 individuals between the maneuvering (n = 207) and load-lifting (n = 263) datasets used to calculate species averages, as follows: 11 *Lesbia victoriae*, 2 *Amazilia lactea* and 5 *Oreotrochilus chimborazo* individuals.

Many of the studied individuals had wings that were noted to have some subtle wing damage, or were missing one primary feather due to molt (29/207 or 14% of the maneuvering birds, and 35/263 or 13% of the load-lifting birds). Because these subtle imperfections were fairly common, and because they also determine both an individual's wing area and wing AR (*42*), we did not omit these individuals from study in either dataset. However, if a bird had more than one of the primary feathers P8, P9 or P10 missing, it was considered to have excessive molt, and was not studied.

Data processing

To process the maneuvering trials, we used an extended Kalman filter to smooth flight trajectory data from the tracking system (*16*, *17*, *25*). We chose conservative smoothing parameters for body position (R_{pos}) and body orientation (R_{ori}) (*17*, *25*) such that six dropped objects had a peak acceleration due to gravity (mean ± SE) of 9.78 ± 0.12 m/s^2. The resulting time series were then processed to identify 8 different stereotyped maneuvers including body translations, body rotations, and complex turns (Fig. 1B and movie S1), using algorithms defined in a previous study (*17*) and Table S1.

Each of the 8 maneuver types was exhibited by all 25 species. In total, 333,135 maneuvers were identified for further analysis. We calculated performance metrics for each maneuver as defined in (*17*, *25*) and Table S1. An advantage of this holistic approach of defining multiple maneuvers is that our results should not depend on any one definition – instead, they reveal broad patterns in translational, rotational, and turning maneuvers.

Each maneuver is defined by the geometry of behavior formalized in Table S1: translations represent translational geometries, rotations represent rotational geometries, and the complex turns represent the combination of translational and rotational changes. For the complex turns, the relative contribution of translation and rotation depends on the specific performance metric, the kinematics of how the turns are performed, and how this varies among individuals (*10*, *30*). For example, if an animal performs an arcing turn by banking while maintaining velocity, it is redirecting thrust from a forward direction to a lateral direction, making $Arc_{\text{cent,max}}$ chiefly a measure of translational performance. In contrast, if the animal performs an arcing turn using asymmetrical wing kinematics to affect a yawing motion, then $Arc_{\text{cent,max}}$ also captures rotational performance. These differences in the underlying physics for the same performance metric illustrate why it is useful to distinguish the complex turns from simpler translational and rotational maneuvers.

Animals often exhibit abnormal stereotypy in captivity, defined as fixed (invariant) patterns of behavior caused by the captive environment (e.g. pacing, grooming, pecking, or moving in circles, in the same way repeatedly). To determine whether the hummingbirds in our study displayed fixed behavior patterns, we looked at the sources of variation in the 11 performance metrics, calculated at the level of individual maneuvers (i.e., from approx. 9,300 – 90,000 maneuvers per type). If stereotypy is high, we expect low variation within trials, such that individuals repeatedly perform a given maneuver using a narrow range of performance values. Contrary to this scenario, we found that 70-89% of the total variation could be attributed to variation within individuals. This demonstrates high within-individual plasticity in the use of these maneuvers. Next, we examined the within-trial coefficients of variation (CV) for each performance metric. The CV is a scale-free measure of dispersion calculated as the sample standard deviation divided by the sample mean. The advantage of this approach is that CVs can be compared across different studies and different types of data. In this study, the typical within-trial CVs ranged from 24-68%, per performance metric. These results were virtually identical to the corresponding CVs obtained from a previous study of Anna's hummingbirds tested alone (22-65%) and in competition (25-70%) (*17*). Furthermore, a paired *t*-test confirmed that there was no significant difference in the within-trial dispersion when comparing these two environmental conditions (alone vs. with a competitor present; $p = 0.71$). Together, these results demonstrate that (i) hummingbirds in the flight chamber exhibit widespread plasticity in the performance of maneuvers, and (ii) the plasticity observed for lone birds is equivalent to that of a more complex environment where a competitor is also present.

For further analysis, we took the mean performance metric for each bird using the sample of maneuvers that the bird performed. The exception was the metric *PRT%*, which was not a mean, but was instead calculated as the percentage of sharp pitch-roll turns out of all complex turns that a bird performed. In a previous study, we showed that a bird's performance on all of these metrics in Table S1 is highly repeatable in separate trials (*17*). We expressed the two negative metrics, $DecHor_{max}$ and $PitchD_{vel,avg}$, as positive values by multiplying them by –1 prior to further analysis. Therefore, increased values for $DecHor_{max}$ and $PitchD_{vel,avg}$ represent faster decelerations and downward rotations, respectively. We expressed the turn duration, PRT_{time}, as negative values prior to statistical modeling and correlation analyses, such that higher values represent faster pitch-roll turns. For clarity, scatterplots of PRT_{time} durations are shown using the original positive values.

To determine muscle capacity independent of other traits and atmospheric conditions, we used the load-lifting database to model the mass of beads lifted in relation to elevation, individual body mass, and sex (male, female or unknown), with a random effect of species in a linear mixed-effects model (LME) (*49*). We took the random intercepts for species from this model as the species-average muscle capacity scores (mean 0, range –5.02 to 4.86; fig. S2). Therefore, a species with a higher value on this score has the capacity to lift more weight on average, given the traits and test elevation of individuals representing that species; this is illustrated in fig. S2.

We calculated individual wing and body mass trait values for the birds in our maneuvering flight trials, relative to their conspecifics, by subtracting the species average from each bird's trait value (e.g., relative body mass, range –2.32 to 2.17 g; relative wing loading, –0.076 to 0.185 g/cm^2; relative AR, –1.58 to 1.99).

Complete wing trait and body mass records were available for 187 out of 207 individuals in the maneuvering analysis. Data were missing from 20 individuals, as follows: 6 individuals from 6 species were missing body mass, 12 individuals from 3 species were missing wing photographs, and 2 individuals from 2 species were missing both body mass and wing photographs. These data were missing because the birds were released unintentionally before they could be measured, and thus we consider them missing at random. Another 7 out of 207 individuals (3%) did not perform any arcing turns (*Arc*) during their maneuvering trials, and were thus missing values for Arc_{radius}, $Arc_{vel,avg}$ and $Arc_{cent,max}$. Although these arcing turns were not missing at random, these individuals were distributed across 4 different species that together had 38 other individuals. Therefore, the *Arc* analyses should be robust to the small number of missing values distributed across several species. We nonetheless included the sample size of maneuvers as a predictor in all regression analyses to account for potential associations between performance and the number of maneuvers performed (see *Statistical analysis* below for details).

Because maneuvering performance was tracked by an automated computer vision system and defined algorithmically, it was determined blind to a bird's traits and species

identity. All other measurements were taken prior to, and thus blind to, the results of the algorithms that determined maneuvering performance.

Sensitivity analysis for smoothing

We chose the Kalman smoothing parameters R_{pos} and R_{ori} by re-projecting the body position and body axis onto sample videos, and then verified that the measured accelerations of dropped objects matched the expected acceleration due to gravity (mean ± SE = 9.78 ± 0.12 m/s^2; n = 6) (*16*, *17*, *25*). Because these smoothing parameters can influence derived metrics such as velocities and accelerations, we also conducted a sensitivity analysis. The purpose of this analysis was to test if altering the smoothing parameters would influence the relative values of the performance metrics that were the subject of our study.

We re-processed videos of four individuals, representing four different species, using eight alternate smoothing parameters (movie S2). These alternate parameters represented ten- and two-fold over- and under-smoothing (0.1*x*, 0.5*x*, 2*x* and 10*x*, for each of R_{pos} and R_{ori}). The results are shown in figs. S3 and S4. This analysis demonstrated several key points. First, the performance values obtained using the alternate smoothing parameters were all highly correlated with the values obtained using R_{pos} and R_{ori} (average r = 0.96, n = 352 correlations; $p < 10^{-4}$ for all except two that were 0.005 and 0.02). This indicates that the relative rankings of the metrics are robust to large changes in smoothing. Second, even a two-fold change in the smoothing parameters (shown in the 0.5*x* and 2*x* columns of figs. S3 and S4) yielded only slight changes in the performance values (mean fold-change ratio for 2*x* undersmoothing = 1.02, range = 0.87-1.15; mean fold-change ratio for 2*x* oversmoothing = 0.98, range = 0.83-1.16). Moreover, this was true across all four species tested. Consistent with a previous study (*25*), we also observed that the magnitudes of the translational maneuvers (*Vel*, *AccHor*, *DecHor*) and complex turns (*Arc* and *PRT*) were most sensitive to R_{pos}, and virtually unchanged by R_{ori}. In contrast, the rotational maneuvers (*PitchU*, *PitchD* and *Yaw*) were most sensitive to R_{ori} and virtually unchanged by R_{pos}. However, even large changes in the influential smoothing parameters still preserved the relative differences among the four tested individuals. This is illustrated in the far right column of figs S3 and S4. Together, these results indicate that the relative differences among individuals and among species are robust to the smoothing procedure.

Statistical analysis

All statistical analyses were performed in R 3.3.3 (*50*). The data and scripts to reproduce the analysis and figures are available at figshare at: https://doi.org/10.6084/m9.figshare.5307136.v3

Correlation analyses

To determine how the different maneuvers are correlated with one another, we generated an association network for the performance metrics (fig. S5A). Each node represents a performance metric, and each edge represents the correlation between two different metrics, weighted as the Pearson's correlation coefficient (*51*) (n = 200 complete records). These results are also used to generate the summary in Fig. 2A.

The correlations in fig. S5A are overwhelmingly positive. An alternative explanation for this strong positive covariance is that maneuvers in the same bout of flight were constrained to be correlated (coupled), either biomechanically, because of the limited space available, or as a result of a given bird's fluctuating motivation. To test this possibility, we derived within-flight bout correlations by filtering the raw data to pair distinct maneuvers that occurred close in time within the same bout of flight (defined as < 1 s apart). There are 51 possible paired combinations of performance metrics derived from distinct maneuvers. We calculated these 51 within-flight-bout correlation coefficients from each of the 207 individuals, omitting any correlations that had $n \leq 10$ paired maneuvers, and then took the average and 95% CI across the population. These results are shown in fig. S5B and summarized in Fig. 2B. This analysis demonstrated that virtually all of the strong among-individual correlations cannot be explained by constraints that cause coupling within flight bouts. For example, it is possible that space constraints could cause birds to decelerate more rapidly after a particularly fast acceleration. However, although $AccHor_{max}$ and $DecHor_{max}$ are strongly correlated among individuals ($r = 0.88$), they are only moderately correlated within flight bouts ($r = 0.34$ [95% CI 0.32-0.37]), demonstrating that this constraint cannot completely explain the phenotypic correlation among birds. Moving faster also allows for greater turning forces, because faster air movement over the wings allows for quicker force asymmetries (via pronation/supination) as compared to flight at slow speeds when force asymmetries are generated by stroke asymmetries (*10*). Accordingly, we checked $Arc_{cent,max}$, Vel_{max} and $AccHor_{max}$. The centripetal acceleration $Arc_{cent,max}$ is much more strongly correlated to Vel_{max} among individuals ($r = 0.66$) than it is to Vel_{max} in the same flight bout ($r = 0.28$ [0.25-0.32]). The same is true for $Arc_{cent,max}$ and $AccHor_{max}$ ($r = 0.52$ vs. 0.15 [0.11-0.20]). This indicates that these maneuver correlations are caused by intrinsic differences among birds, and not merely due to coupling within bouts of flight.

To evaluate whether larger birds were more constrained by the chamber size than smaller birds, we examined two further analyses. First, we tested whether the within-flight bout correlations were associated with individual body mass (fig. S5C). In no case did the within-flight bout correlation coefficient increase significantly with body mass. This demonstrates that small and large birds experienced the same degree of coupling for maneuvers paired within the same bout of flight. Second, we examined the full distributions of *Arc* turn radii for each bird as a measure of maneuver size. The maximum Arc_{radius} and the span of Arc_{radius} values were both unrelated to body size (fig. S5D-E, all $p > 0.52$). As shown in fig. 5D-E, both small and large birds expressed a wide range of *Arc* turn sizes. We cannot rule out the possibility that in a larger chamber, we would have observed even larger differences among birds. However, collectively, these analyses demonstrate that the chamber met our goal of capturing variation within and among the studied species.

Species differences

To evaluate how well species could be discriminated on the basis of their flight style, we used a discriminant function and cross-validation analysis. The discriminant analysis finds linear combinations of predictor variables, called discriminant functions

(DFs), that best predict species identification (based on species labels known *a priori*). We used the bioinformatics package adegenet (*52*) to perform this analysis because of its flexibility and graphical options. The 'dapc' function in this package works by first transforming the data using a PCA, to allow optional dimension reduction and unsupervised cluster finding, if desired (*53*). However, in our case, no dimension reduction was performed because we wanted to preserve the maximum information available; furthermore, our analysis was supervised (in that the grouping variable, species, was already known). For the maneuvering analysis, we entered the 12 performance metrics into the model, as well as the sample size of each of the 8 maneuvers (number of maneuvers performed in a 30 minute trial, log transformed to meet normality assumptions), for a total of 20 variables. It was important to include the sample sizes to maximize the amount of information for discriminating the species, and because frequency of use can be associated with the performance metrics (see fig. S12 and (*25*)). For comparison, we also performed a separate discriminant analysis using 3 morphological traits: body mass, wing area, and wing AR. All variable contributions in the resulting models were based on variables that had been back-transformed onto the axes and scale of the original data for interpretation (*52*).

We ran a cross-validation procedure for each discriminant analysis to evaluate how well it could categorize the species. We first fit the discriminant model using a partially random subset of 72% of the complete records (n = 129 out of 180 individuals), and then used the resulting model to predict the species labels for the remaining 51 samples. The subset for model building included all individuals from species with fewer than 3 individuals and randomly selected 2/3 of the individuals from species with ≥ 3 individuals. This procedure was repeated 10,000 times to determine the average percent correctly classified. To derive a null distribution for this statistic, we also performed this procedure using data with randomly permuted species labels. These results were used to calculate Cohen's□ □ statistic, which is a measure of agreement relative to the rate expected to occur by chance. A □ of 1 indicates that the model could determine species classification accurately for all individuals, whereas a □ of 0 means that the classification accuracy was no better than expected under random allocation.

To visualize the DFs and quantify the contribution of different variables (fig. S6), we ran the discriminant analyses using all individuals with complete data (n = 180). To evaluate the extent to which closely related species resemble each other, we calculated the species-average score on the first two DF axes for each species (n = 25). We then estimated Pagel's□ λ for each DF axis as a continuous trait using the 'phyloSignal' function in the phylosignal R package (*54*, *55*). A Pagel's□λ significantly greater than 0 indicates that closely related species resemble each other more so than randomly selected species.

Note that when DFs are used in further analyses with additional character(s) that can also be mapped onto a phylogeny (such as habitat type), it is important to account for phylogenetic non-independence of the residuals before making statistical inference (*56*). However, in our case, the inference in the cross-validation analysis is simply how well

the overall classification model performs, making no assumptions of independence or use of additional characters.

Biomechanical traits and maneuvering performance

Our analysis of biomechanical traits and maneuvering performance had four goals: (i) to describe how the performance of different maneuvers scales with body mass; (ii) to identify whether other biomechanical traits can explain this scaling relationship; (iii) to compare the effects of biomechanical traits at two levels of diversity: across species, and among conspecific individuals within species; and (iv) to evaluate the strength of these effects after accounting for shared phylogenetic history.

To achieve (iii), we used within-group centering in a mixed-effects regression model framework, entering the species average as well as the individual trait values (relative to the species average) as two separate fixed effects (*49*, *57–59*). This was done for each trait except for muscle capacity, because load-lifting information was not available for the same individuals that were studied in the maneuvering trials (see 'Morphology and muscle capacity' above for details). Thus, muscle capacity is examined at the species level only.

We first fit a 'scaling' model for each performance metric. These were linear mixed-effects (LME) models that included species body mass and individual body mass as the only fixed effects (*49*). Species was entered as a random effect to account for non-independence of the residuals for birds of the same species. The sample size was n = 187, except for performance metrics derived from *Arc* (n = 180). The next step was to update this 'scaling' model to the 'full' model by adding the other traits, while still accounting for body mass at both the species and individual levels. These other traits were: species muscle capacity, species wing loading, species AR, individual wing loading, and individual AR. We verified that the residuals for all LME models met the assumptions of normality and that the variance inflation factors (VIFs), a measure of multicollinearity, were all < 2.7.

Next, we refit the 'scaling' and 'full' models using a Markov chain Monte Carlo (MCMC) algorithm for phylogenetic mixed-effects regression models (*58*). Like the LME models, these MCMC models included a random effect to account for repeated measures within species; they also included an additional random effect to account for non-independence of the residuals owing to shared evolutionary ancestry (*59*). This additional step is important to distinguish between associations that may be attributed to shared evolutionary history versus those that result from statistically independent evolutionary events. We used the most recent multilocus phylogeny for the family Trochilidae (*23*) for this phylogenetic random effect. We used priors for both the random effect and residual variances corresponding to an inverse-Gamma distribution with shape and scale parameters equal to 0.01 (*59*), and used the default uninformative priors for the fixed effects (*58*). Two chains were run for each model and every 500^{th} sample was stored after a burn-in of 3,000, stopping after 1,003,000 iterations, for a total of 2,000 stored iterations per chain. This procedure yielded autocorrelation values between successive stored iterations that were ≤ 0.09. To check that the separate chains had

converged, we used the Gelman-Rubin diagnostic procedure and determined that the upper 95% confidence limit for the Gelman-Rubin statistic was ≤ 1.03 in every case.

All models included site elevation and the number of maneuvers performed as additional fixed effects. There is a general decline in aerial maneuverability at higher altitudes as a result of lower air density (*25*). Therefore, we entered elevation as a categorical factor (high or low), because our test sites were either below 300 m or above 3000 m elevation (fig. S1). The number of maneuvers performed was also included to determine whether performance differs for birds that use a given maneuver more frequently (fig. S12). As defined in the main text, a negative association between the level of performance and the number of maneuvers performed indicates an effect of maneuvering fatigue. In contrast, a positive association indicates an effect of maneuvering skill, such that birds performing a maneuver more often tend to perform it at a higher level.

Prior to mixed-effects modeling, we centered and standardized all trait and response variables (mean 0, SD 1) so that the regression slopes could be compared among traits and among maneuvers. Thus, the regression slope or 'effect size' gives the change in standard deviations of the performance metric for one standard deviation increase in the predictor. These effect sizes are presented in fig. S7 as well as the marginal R^2_{GLMM} estimates of the variance explained by the fixed effects in the whole model (*60*). The R^2_{GLMM} values ranged from an average of 0.13 (complex turns) to 0.20 (translations) and 0.37 (rotations). When comparing the results from LME and MCMC models, we found that the effect sizes were nearly identical except for a few species-level effects (fig. S8). If the magnitude of a species-level effect was reduced in the MCMC model, it suggests that the association is at least partly due to shared evolutionary history. Conversely, if the magnitude increased in the MCMC model, the trait is more strongly associated with performance after accounting for species ancestry.

To visualize the results of the LME models, we also generated partial effect plots showing residual values that adjust for the other parameters in the LME model (*61*) (e.g., Figs. 5A,B and 6B,C, fig. S12). These plots show the relative effect of one variable, on the scale of the original data, given the other predictors in the statistical model. For all partial effect plots, we took the average of the residuals for the high and low elevation categories.

Maneuver-trait clustering

We used a hierarchical clustering analysis to test the hypothesis that different maneuvers rely on different biomechanical traits (*62*). The input matrix was the set of effect sizes for 7 traits on 12 performance metrics, derived from the regression models described above. We first generated a distance matrix for the performance metrics from the Pearson's correlations among the effect sizes subtracted from 1, and then ran the bottom-up UPGMA agglomerative clustering algorithm (*50*). We derived two measures of support for the resulting dendrogram: the agglomerative coefficient, wherein higher values indicate a more clustered structure (*63*, *64*); and the multiscale bootstrap p-values quantifying the percent support for each cluster node (*62*). We compared the

agglomerative coefficient with the null distribution obtained from 10,000 random permutations of the input matrix. For the node support, we obtained approximately unbiased (AU) p-values after 5,000 bootstrap iterations (*62*). As an alternative way to visualize the maneuver-trait clusters, we also plotted the effect size matrix in fig. S9 as a bipartite graph (*51*) illustrating the results of the MCMC models.

Use of complex turns

To understand how both behavior and biomechanics jointly determine the use of the two complex turn maneuvers, we fit additional mixed-effects regression models of *PRT%* that included candidate predictors identified by the previous analyses as being important (figs. S5 and S9). The model (n = 180 individuals) included elevation as well as species-average values for body mass, wing loading, and $Arc_{cent,max}$, as well as individual values (relative to the species-average) for wing aspect ratio (AR) and $Arc_{cent,max}$. We followed the same procedure for fitting MCMC and LME models described above. The R^2_{GLMM} value for the fixed effects in this model was substantially higher, at 0.30, than it was for the initial full model of *PRT%* (R^2_{GLMM} = 0.16). The results of this behavioral model of *PRT%* are shown in Fig. 6.

Inspection of these results revealed that one species, *Eriocnemis luciani* with n = 1 individual, is an outlier for $Arc_{cent,max}$ with high leverage in this analysis (Fig. 6B). Therefore, we re-ran the analysis with *Eriocnemis luciani* removed (n = 179 individuals). All of the model estimates were virtually unchanged (Pearson's correlation comparing the original and updated estimates > 0.99, $p < 1 \times 10^{-12}$), including the R^2_{GLMM} (which decreased only slightly to 0.28), demonstrating that the analysis is robust to removing this one individual/species.

Species-specific analyses

To interpret our results in the context of single species studies, we conducted two additional analyses. First, to determine if significant individual-level trait effects could be detected with much smaller sample sizes, we selected 8 species with the most complete data (range of n = 10-17), and re-fit linear regressions of the performance metrics and individual-level traits for each species separately (fig. S10). The results are used in fig. S10 to show that significant trait effects are only recovered about 12.5% of the time in these single-species studies with low statistical power.

Second, to test whether individual wing aspect ratio (AR) has species-specific effects on the use of complex turns (*17*), we re-fit the full LME model for *PRT%* (fig. S11). In the new analysis, we allowed the slope for individual AR to vary in different species as a random slope. We then compared the deviation or spread of the species-specific slope values to that obtained from a randomization test (permuting individual AR 10,000 times). If AR has species-specific effects, the spread or variation in these slope values should be greater than expected from randomized data.

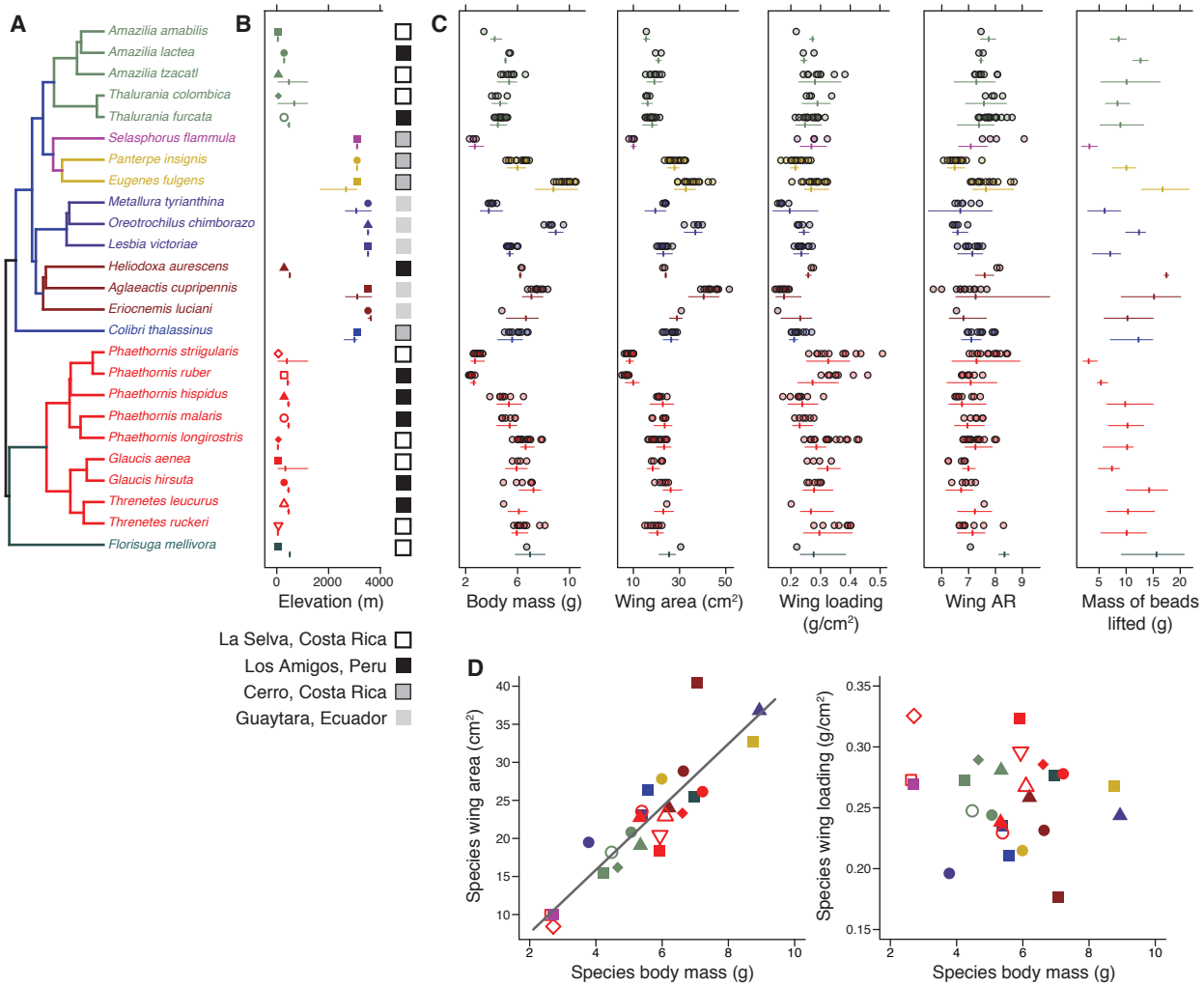

fig. S1. Study sites and morphological traits. (A-B) Each species in the maneuverability dataset was studied at a single capture site that was either below 300 m or greater than 3,000 m elevation. (C) Birds were also assessed for morphological traits and muscle capacity. The individuals in the maneuverability trials are shown using filled circles (n = 207 individuals). Vertical ticks in (B-C) show the means and ranges for a sample of 263 individuals tested for muscle capacity by measuring the mass of beads they could lift in the load-lifting assay. See fig. S2 for more details. (D) Species wing area scales positively with body mass. Species wing loading is not related to species body mass. Each species is represented by a unique color-symbol combination in (B) and (D).

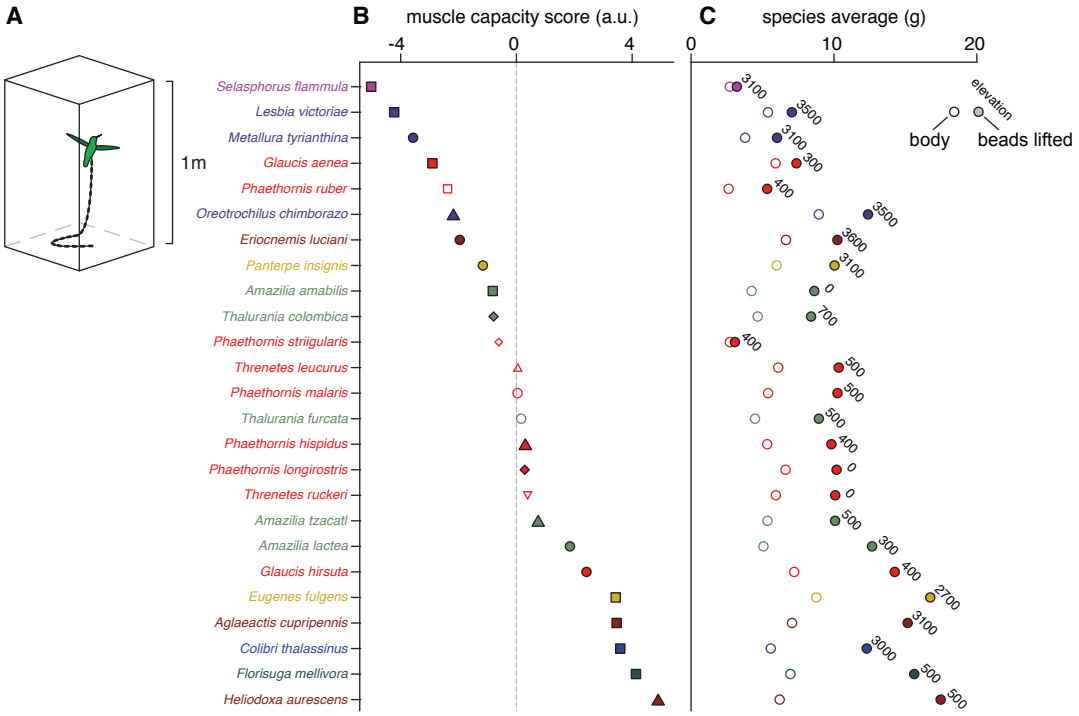

fig. S2. Species muscle capacity scores. (A) Muscle capacity was measured using a load-lifting assay that takes advantage of the hummingbird's natural escape response to measure the mass of beads the bird can lift in vertical flight. (B) Species-average scores for muscle capacity (arbitrary units) were obtained by modeling the mass of beads lifted, accounting for elevation, sex, and body mass (n = 263 individuals). The species with higher scores lift more weight, given their body mass and test elevation. (C) Species-average values for body mass and the mass of beads lifted, with the average test site elevation rounded to the nearest 100 metres. Note that sex is an additional source of variation in the analysis that is not illustrated on the figure.

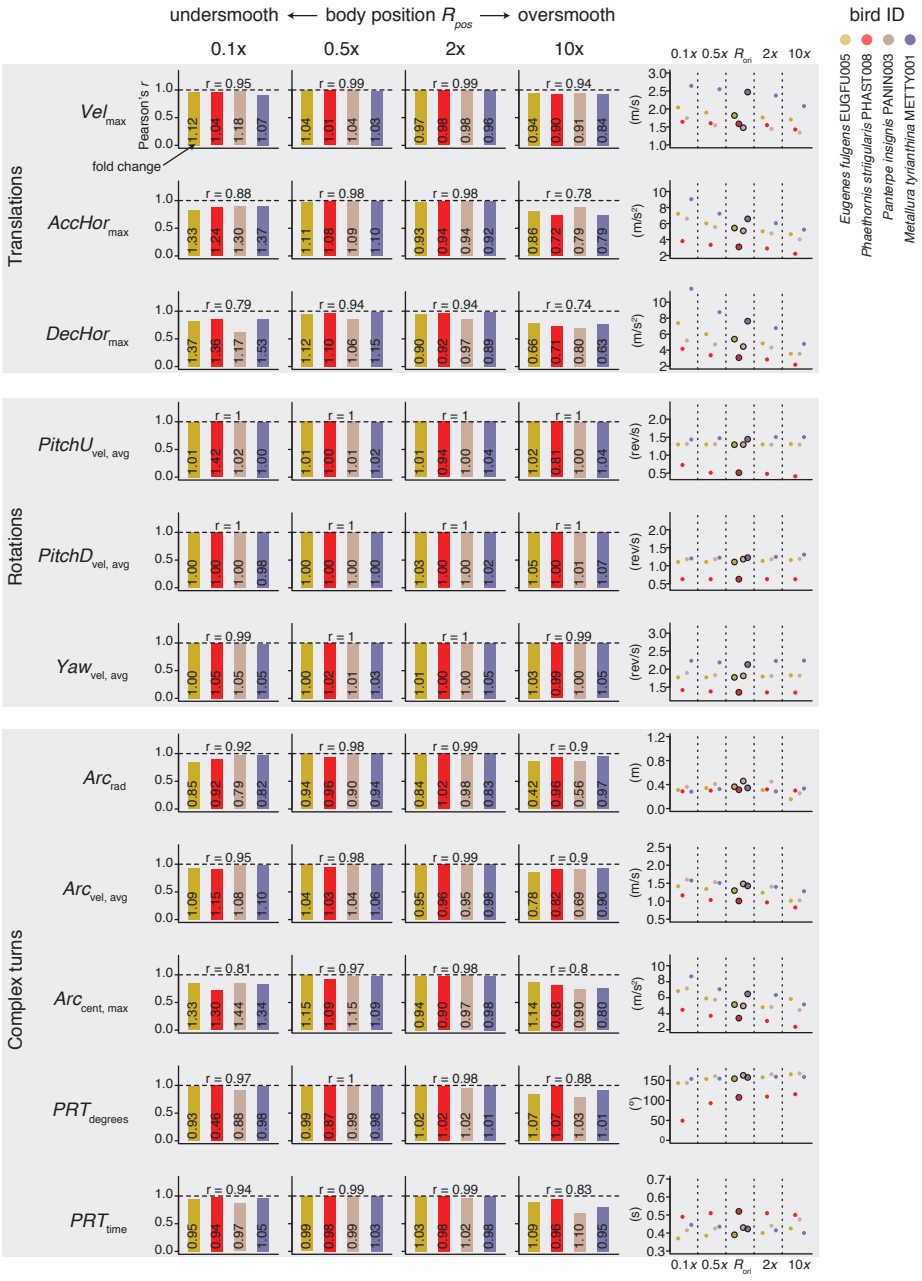

fig. S3. Sensitivity analysis for body position smoothing parameters. Videos of four birds (each representing a different species) were re-processed with alternative body position smoothing parameters (0.1x, 0.5x, 2x, and 10x, where x is R_{pos}, the parameterization used for the main analyses). The bar plots show the correlation coefficients, *r*, for each bird, for correlations comparing the performance metrics obtained at R_{pos} to the values obtained with the alternate smoothing parameters. The average correlation coefficient is presented above the bars. The numbers within each colored bar give the average fold change of the performance metric, expressed as a ratio. A fold change of 1 is no change; >1 is an increase, and <1 is a decrease. The scatterplots at right show the trial medians for each individual under different smoothing scenarios. The central datapoints outlined in black were obtained using R_{pos}.

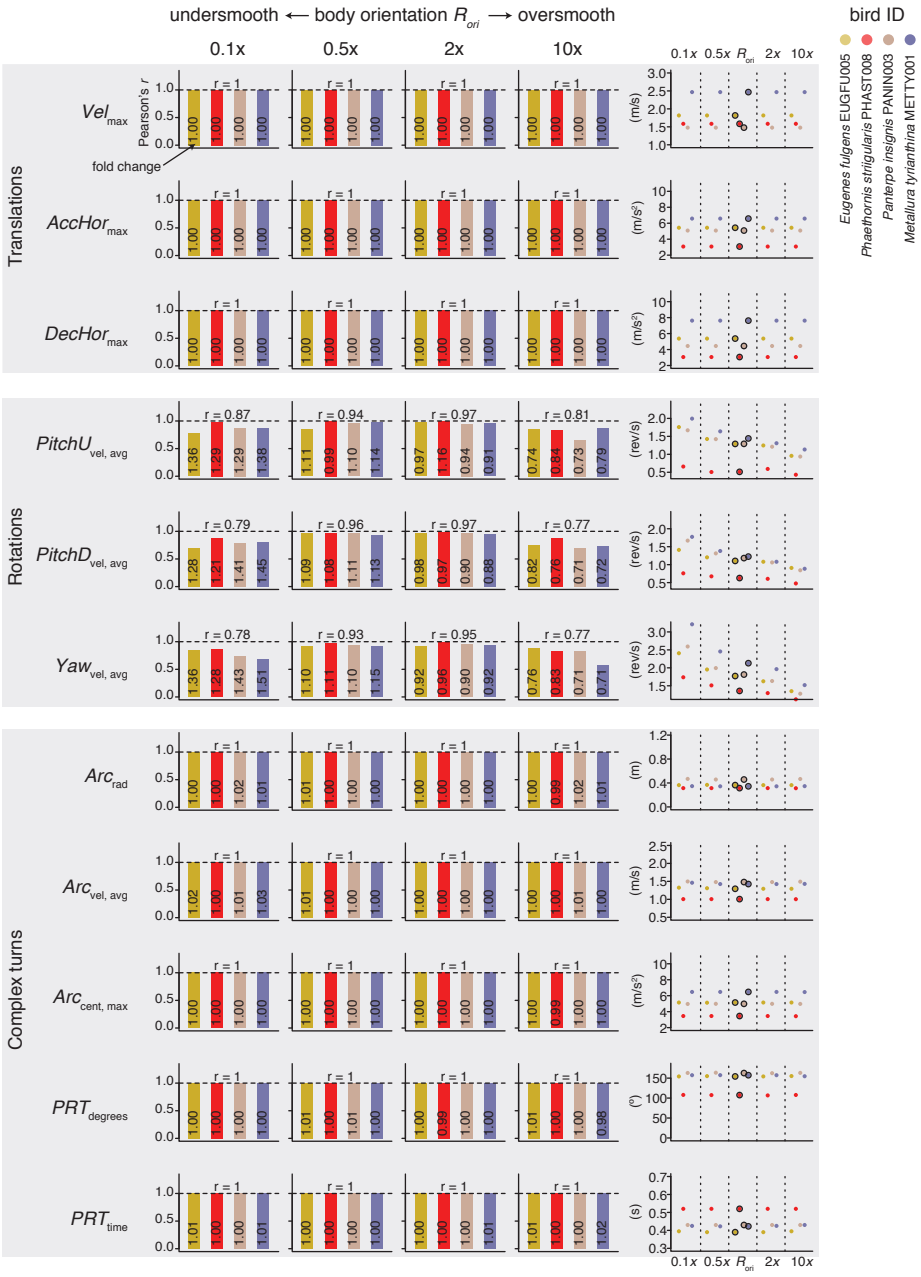

fig. S4. Sensitivity analysis for body orientation smoothing parameters. Videos of four birds were re-processed with alternative body orientation smoothing parameters (0.1x, 0.5x, 2x, and 10x, where x is R_{ori}, the orientation smoothing parameter used for the main analyses). All other features in this figure follow fig. S3.

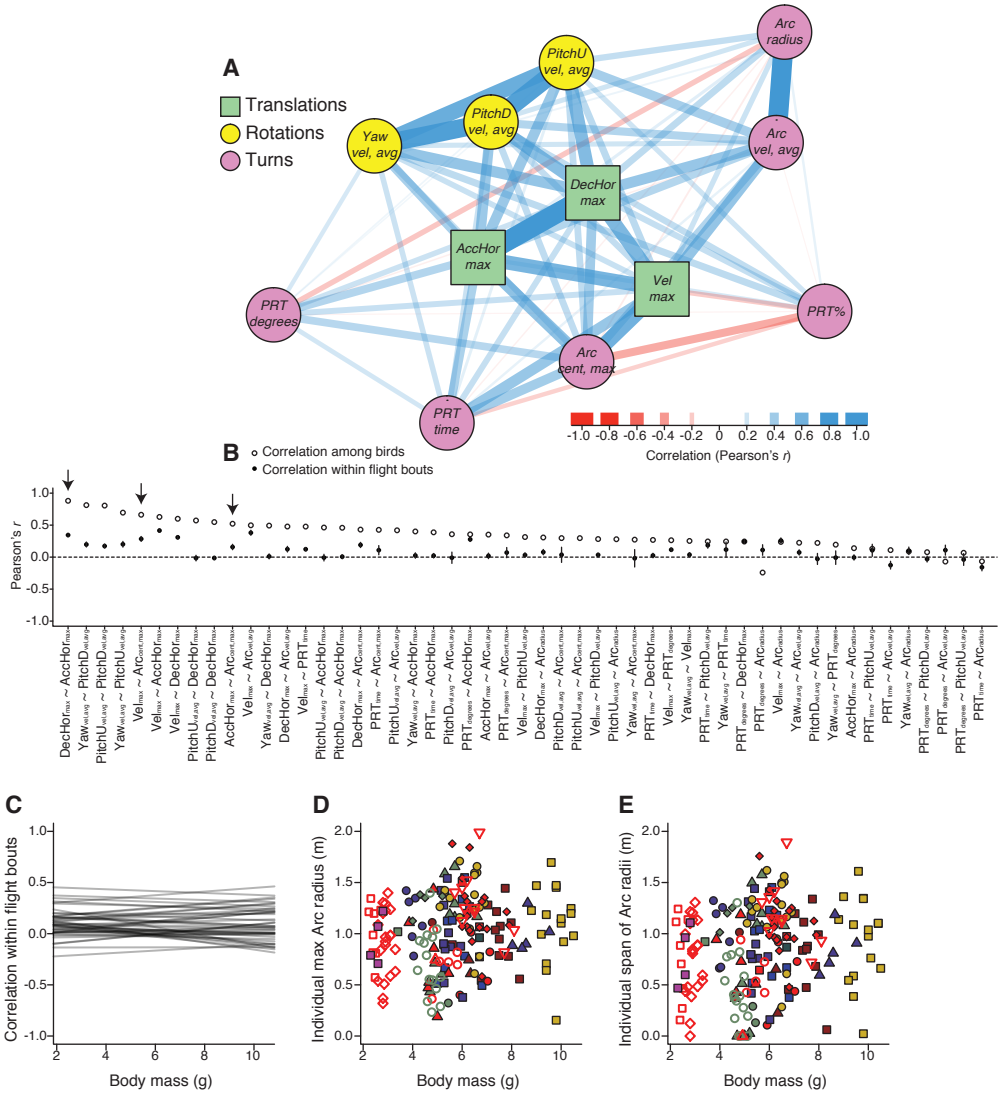

fig. S5. Maneuver correlations among individuals and within flight bouts. (A) Each edge in the network represents a Pearson's correlation coefficient (r) for the correlation between two performance metrics, among individuals, calculated from n = 200 individuals with complete maneuvering data. Edge thickness and color indicate the correlation strength and direction, respectively. The network layout is based on the Fruchterman-Reingold algorithm to minimize edge crossings. Note that the negative metrics ($DecHor_{max}$ and $PitchD_{vel,avg}$) were transformed to positive values prior to the correlation analysis, whereas turn duration (PRT_{time}) was transformed to negative time values, so that higher values represent faster pitch-roll turns. These data are used to derive Fig. 2A. (B) Comparison of among-individual and within-individual/within-flight bout correlations. The three arrows point to specific comparisons for: (1) $DecHor_{max} \sim AccHor_{max}$, (2) $Vel_{max} \sim Arc_{cent,max}$, and (3) $AccHor_{max} \sim Arc_{cent,max}$. Note that within-flight bout correlations are not available if maneuvers did not occur within 1 s of each other. These data are also shown in Fig. 2B. (C) Regression fits for the within-flight bout correlations in relation to body mass. Each line is a best-fit for a separate pair of performance metrics as in (B). (D-E) The maximum Arc_{radius} per trial and span of Arc radii per trial, in relation to body mass. Each species is represented by a unique color-symbol combination as in fig. S1.

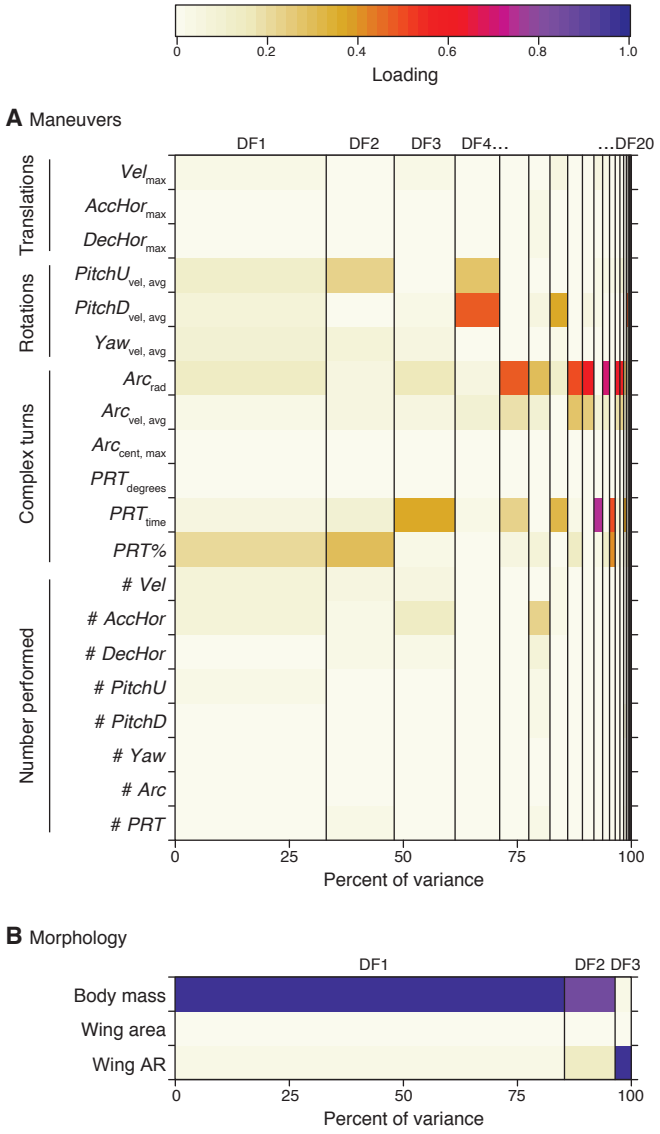

fig. S6. Loadings for the discriminant function (DF) analyses. Color indicates the contribution made by each (A) behavior or (B) trait to a DF, on a scale from 0 (no contribution) to 1 (a very strong contribution). The column widths are scaled to represent the percent of the total variance explained by each DF. (A) The variation among species in maneuvering style is attributed mainly to their performance complex turns and rotational maneuvers. The translational maneuvers (top rows) and the # of maneuvers performed (bottom rows) have much weaker loadings. This indicates that the species do not differ as much in translational performance or the absolute number of maneuvers performed. (B) Nearly all of the morphological variation among species is attributed to body mass.

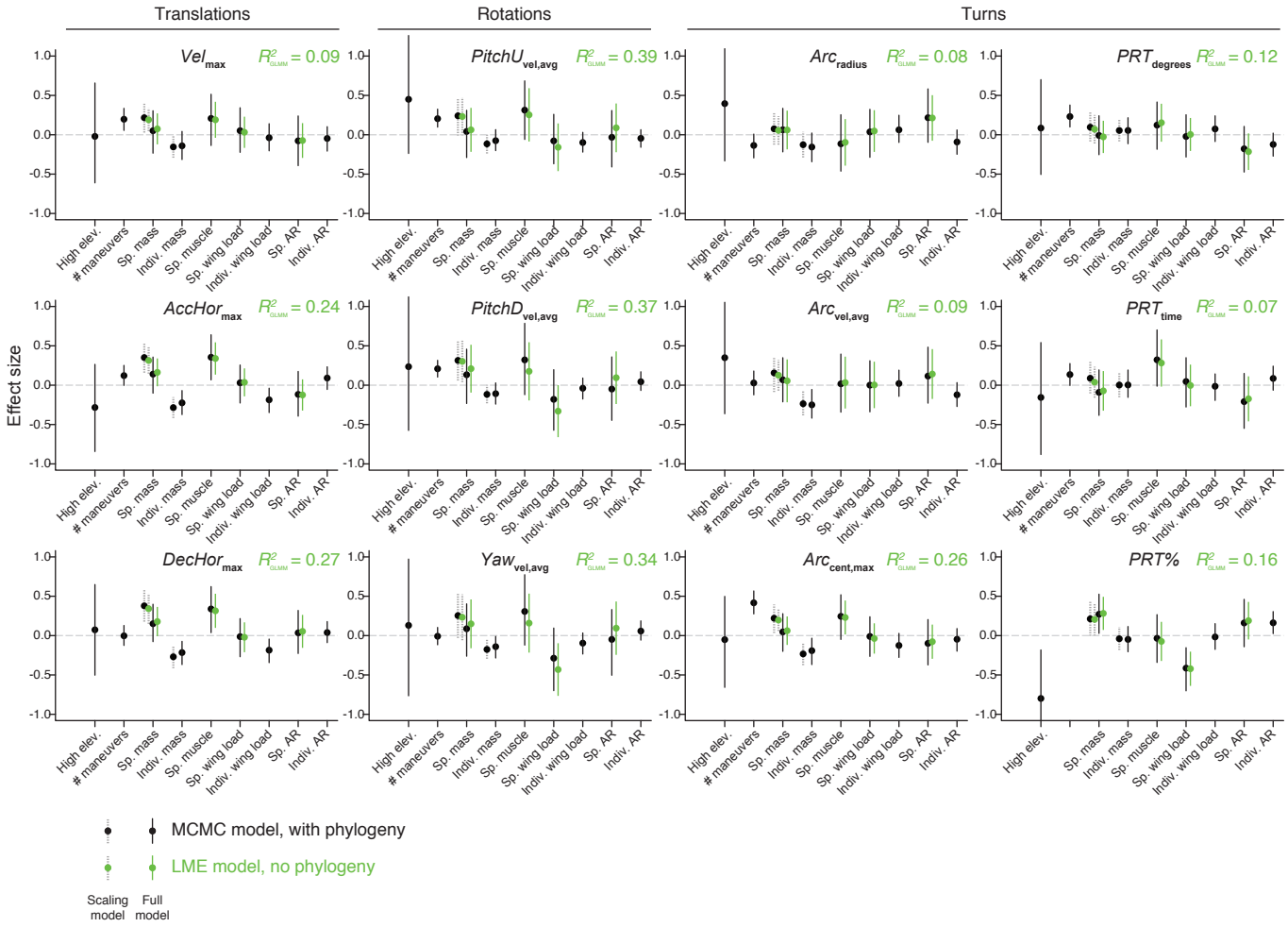

fig. S7. Models of biomechanical traits and maneuvering performance. Results are shown in black for MCMC analyses (±95% highest posterior density intervals) that account for phylogenetic history. Results of LME analyses are also shown in green (±95% confidence intervals), to the right of the MCMC results. For simplicity, LME results are only shown for the species level traits, because the individual level results closely match the MCMC effect sizes (see also fig. S8). The results for the scaling models of body mass are shown to the left of the full model results. "Sp." = species; "Indiv." = individual. Note that PRT_{time} is modeled using negative time values so that it corresponds to turn speed for comparison with $Arc_{vel,avg}$, such that higher values indicate faster turns for both of these metrics. R^2_{GLMM} values provide an estimate of the variance explained by all fixed effects in the full LME model.

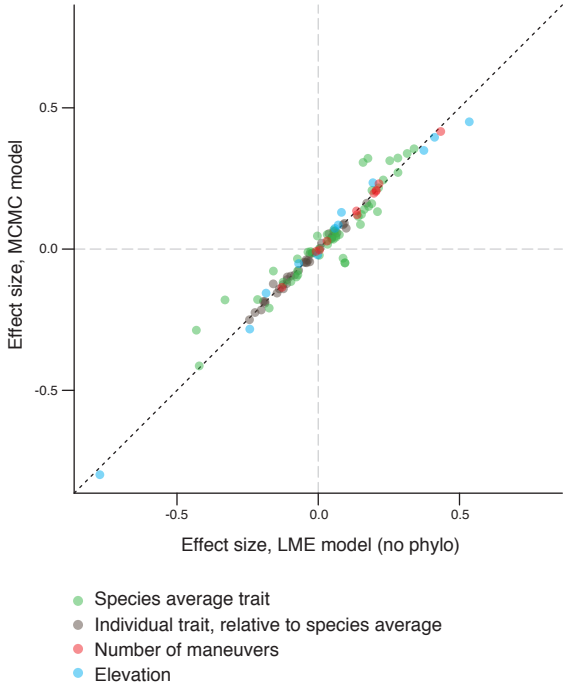

fig. S8. Comparison of MCMC and LME results. Effect sizes from MCMC analyses (y-axis) are broadly consistent with the results of LME analyses (x-axis), with the exception of some traits at the species level (green).

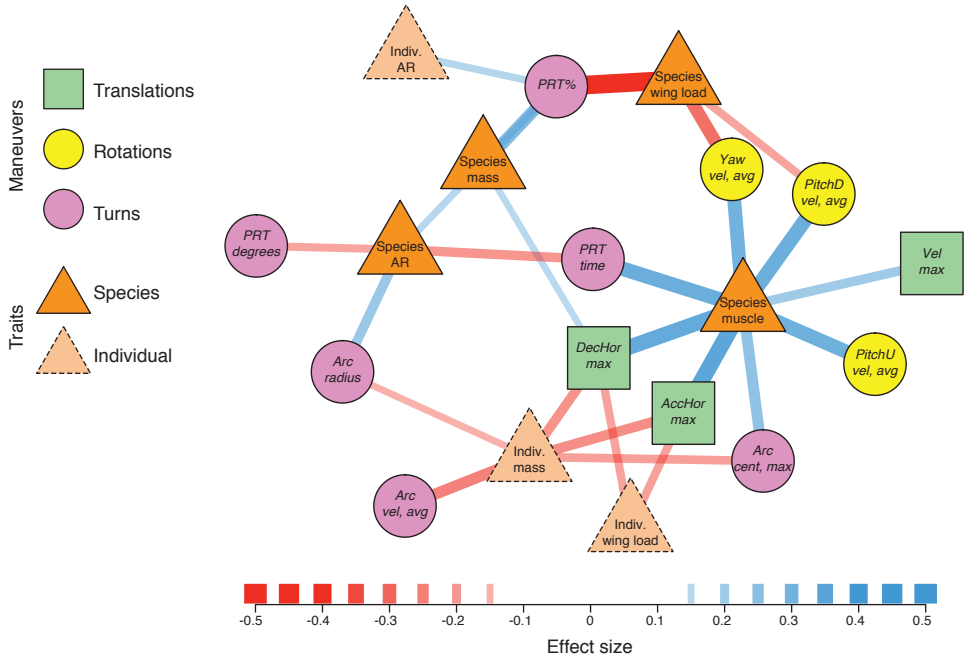

fig. S9. Different maneuvers are associated with different biomechanical traits. Each edge in the network represents the effect of a trait on a particular maneuver, with edge thickness and color indicating the magnitude and direction of the results from the MCMC analysis. The network layout is based on the Fruchterman-Reingold algorithm to minimize edge crossings. All data were standardized to allow the effect sizes to be compared. For clarity, only edges with a magnitude > 0.15 are shown. This is an alternative way to visualize the results shown in Fig. 4C.

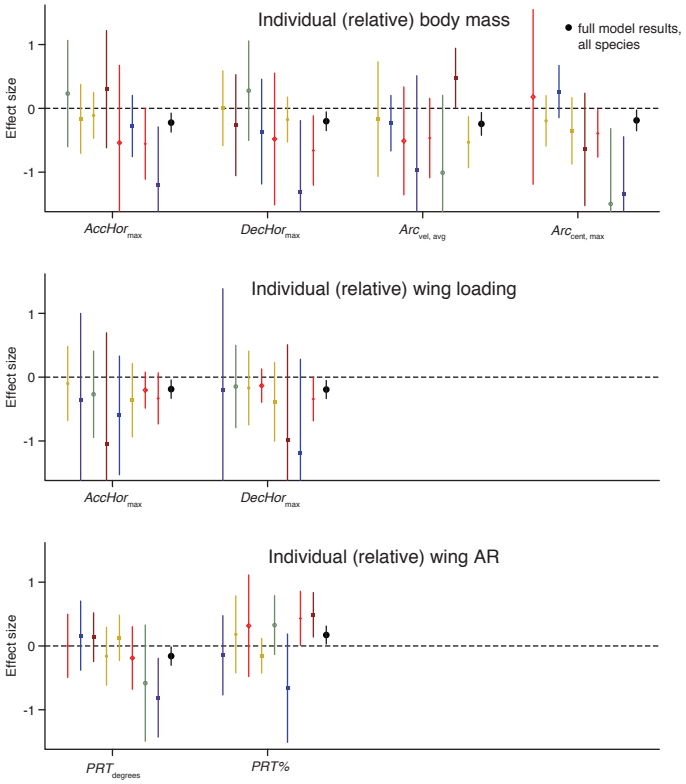

fig. S10. Effects of morphology when single species are analyzed separately. An individual bird's body mass and wing morphology have relatively subtle effects on maneuvers that are not detected in low-power studies. The large black points show the results from the full models including all data (±95% confidence interval; n = 187, except 180 for *Arc*). Colored datapoints show the effect sizes for single-species analyses of species with n ranging from 10-17 individuals; data are sorted by p-value and colored according to clade. Significant effects in the same direction as the full results were detected in only 8 out of 64 single-species analyses (i.e., only 12.5% of the time).

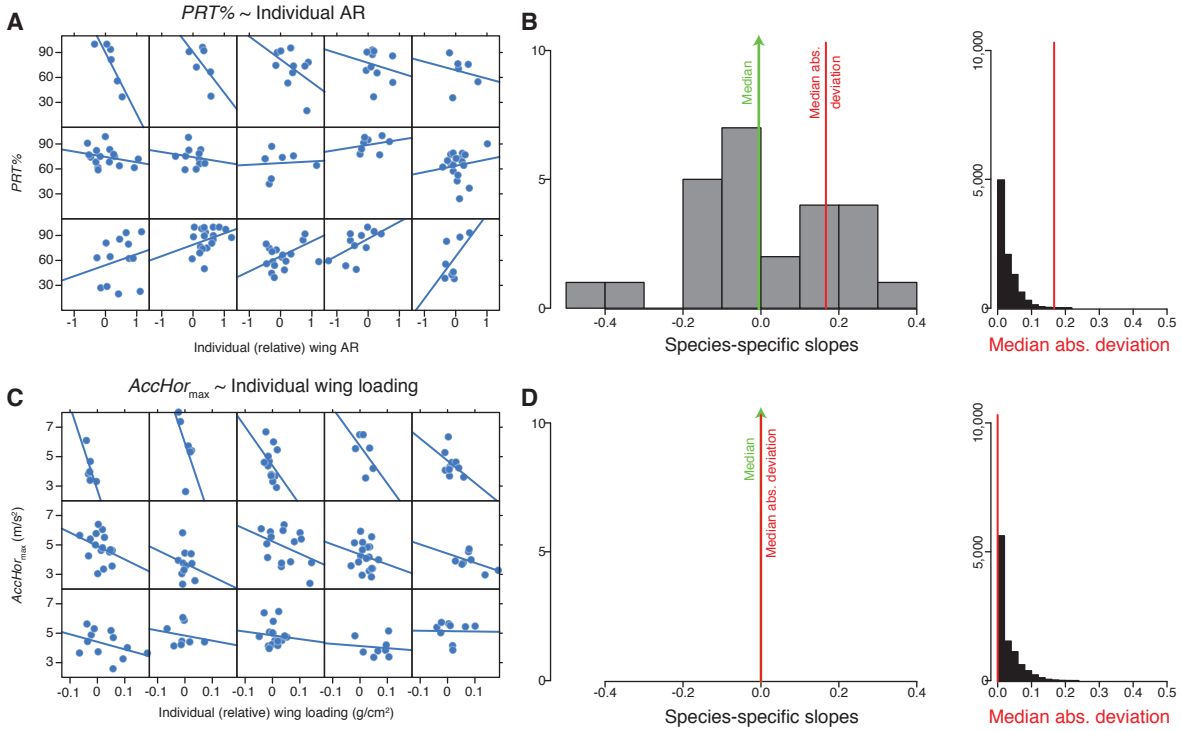

fig. S11. Effects of wing shape vary in different species. (A) Individual wing aspect ratio (AR) has either a negative, neutral, or positive association with *PRT%*, depending on the species (note that its overall effect is weakly positive; Fig. 4C). Each panel in (A) shows the least-squares fit for a single species with n ≥ 6, ordered by slope from negative (top) to positive (bottom). (B) Histogram for the effects of individual AR on *PRT%* from a model where AR slopes can vary by species. The species-specific slopes range from –0.40 to 0.34. The median absolute deviation (red) is a measure of spread (higher values indicate more spread). On the right, the deviation in slopes of 0.17 is greater than that obtained from randomized data (black histogram) more than 99% of the time. This indicates that the effects of wing shape vary among species significantly more than expected by chance. (C-D) For comparison, the same analysis is presented for individual wing loading and horizontal acceleration ($AccHor_{max}$), because it has a similar overall effect size (i.e., 0.19 as compared to 0.17 for AR and *PRT%*). (D) The species-specific slopes are all 0.00 and the deviation is 0.00, because the effect of wing loading is consistently negative. This deviation is less than that obtained from randomized data 99% of the time. Thus, the effects of wing loading are more consistent than expected by chance.

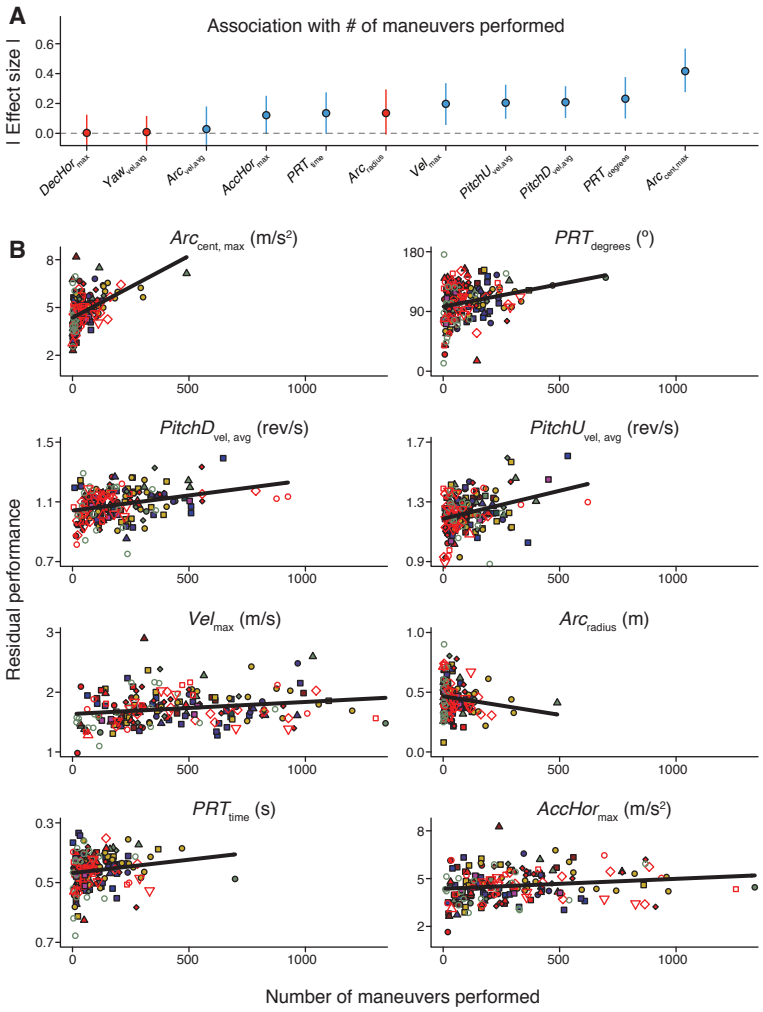

fig. S12. Performance is associated with the frequency of use of several maneuvers. (A) Effect sizes ±95% HPD interval from the MCMC analyses. Blue indicates a positive association between the performance metric and the use of that particular maneuver; red is negative. (B) Partial effect plots for the performance metrics associated with the number of maneuvers performed. Each plot shows the residual values after accounting for the other effects in the statistical model. Birds that make more smooth arcing turns (*Arc*) perform them with higher centripetal accelerations and smaller radii. Birds that make more sharp pitch-roll turns (*PRT*) use them for larger heading changes and perform them in less time. Birds that make more downward rotations (*PitchD*), more upward rotations (*PitchU*), and more total velocity increases (*Vel*) perform these respective maneuvers more rapidly. Lastly, birds that make more horizontal accelerations (*AccHor*) tend to do so with slightly greater magnitudes of acceleration.

Table S1.

Definitions of the maneuvers and their performance metrics. The definitions were used to identify maneuvers from body position and orientation data streams. Modified from (*17*).

Category	Symbol	Maneuver	Definition	Performance Metric	Units
Translations	*Vel*	3D acceleration	Start: velocity *xyz* min End: velocity *xyz* max Distance *xyz* > 25 cm	• Max *xyz* velocity Vel_{max}	m/s
	AccHor	Horizontal acceleration	Start: velocity *xy* min End: velocity *xy* max Distance *xy* > 25 cm Distance *z* < 10 cm	• Max *xy*-plane acceleration $AccHor_{max}$	m/s^2
	DecHor	Horizontal deceleration	Start: velocity *xy* max End: velocity *xy* min Distance *xy* > 25 cm Distance *z* < 10 cm	• Max *xy*-plane deceleration $DecHor_{max}$	m/s^2
Rotations	*PitchU*	Pitch-up body rotation	Start: pitch min End: pitch max Degrees rotated > 45º Distance *xyz* < 10 cm	• Average pitch velocity $PitchU_{vel,avg}$	rev/s
	PitchD	Pitch-down body rotation	Start: pitch max End: pitch min Degrees rotated > 45º Distance *xyz* < 10 cm	• Average pitch velocity $PitchD_{vel,avg}$	rev/s
	Yaw	Yaw turn	Start: yaw vel = 0º/s End: yaw vel = 0º/s Degrees rotated > 90º Pitch max < 75º Distance *xyz* < 10 cm	• Average yaw velocity $Yaw_{vel,avg}$	rev/s
Complex turns	*PRT*	Sharp, pitch-roll turn in the horizontal plane	Start: velocity max End: velocity max Pitch max > 75º Distance *xy* before vel min > 12.5 cm Distance *xy* after vel min < 12.5 cm Distance *z* < 10 cm	• * Degrees turned $PRT_{degrees}$ • * Duration PRT_{time}	º s
	Arc	Smooth, arcing turn in the horizontal plane	Start: Δ heading > 0.25 rev/s End: Δ heading < 0.25 rev/s Velocity *xy* min > 50 cm/s Distance *xy* > 25 cm Distance *z* < 10 cm	• ‡ Turn radius Arc_{radius} • ‡ Average *xy*-plane velocity $Arc_{vel,avg}$ • ‡ Centripetal acceleration, (velocity/radius^2) $Arc_{cent,max}$	m m/s m/s^2

* For a 25 cm segment centered at the minimum *xyz* velocity

‡ For a 25 cm segment centered at the sharpest point of the *Arc* turn

Movie S1

Maneuvers of two hummingbird species recorded in Costa Rica. The first clip shows an individual *Eugenes fulgens*, a high elevation specialist with a body mass of 8.8g. Body position (blue) and orientation (red) are reprojected onto the videos of four synchronized cameras. The sequence of maneuvers illustrated in Fig. 1A occurs at 11-13s. The second clip shows an individual *Phaethornis striigularis*, a much smaller lowland species with a body mass of only 2.9g. All maneuvers are shown in real time. Note that while the cameras in the tracking system occasionally drop frames (indicated by white frames), continuous 3D tracking was obtained using the other cameras.

Movie S2

Comparison of smoothing parameters. The matrices Q, R_{pos}, and R_{ori} are implemented in the extended Kalman filters used to smooth body position and orientation data. These parameter values were chosen by reprojecting hummingbird body position and orientation onto sample videos, and validated by verifying the gravitational acceleration of dropped objects. Next, a sequence of maneuvers is shown by an individual *Eugenes fulgens* under nine alternate smoothing scenarios. The first clip uses R_{pos} and R_{ori} smoothing parameters implemented in the main text. The sequence repeats for each alternate smoothing scenario (two-fold and ten-fold over- and under-smoothing for R_{pos} and R_{ori}, for a total of eight alternate scenarios). The original 200 fps video is modified so that the bird's motion is shown in real time. Note that while one of the cameras occasionally dropped frames, continuous 3D tracking was obtained from the other cameras.

References and Notes

1. R. Dudley, Mechanisms and implications of animal flight maneuverability. *Integr. Comp. Biol.* **42**, 135–140 (2002). doi:10.1093/icb/42.1.135 Medline

2. M. F. Land, T. S. Collett, Chasing behaviour of houseflies (*Fannia canicularis*). *J. Comp. Physiol.* **89**, 331–357 (1974). doi:10.1007/BF00695351

3. S. A. Combes, D. E. Rundle, J. M. Iwasaki, J. D. Crall, Linking biomechanics and ecology through predator-prey interactions: Flight performance of dragonflies and their prey. *J. Exp. Biol.* **215**, 903–913 (2012). doi:10.1242/jeb.059394 Medline

4. K. M. Sholtis, R. M. Shelton, T. L. Hedrick, Field flight dynamics of hummingbirds during territory encroachment and defense. *PLOS ONE* **10**, e0125659 (2015). doi:10.1371/journal.pone.0125659 Medline

5. T. Y. Moore, A. A. Biewener, Outrun or outmaneuver: Predator–prey interactions as a model system for integrating biomechanical studies in a broader ecological and evolutionary context. *Integr. Comp. Biol.* **55**, 1188–1197 (2015). Medline

6. T. Garland Jr., P. A. Carter, Evolutionary physiology. *Annu. Rev. Physiol.* **56**, 579–621 (1994). doi:10.1146/annurev.ph.56.030194.003051 Medline

7. V. Careau, T. Garland Jr., Performance, personality, and energetics: Correlation, causation, and mechanism. *Physiol. Biochem. Zool.* **85**, 543–571 (2012). doi:10.1086/666970 Medline

8. K. P. Dial, E. Greene, D. J. Irschick, Allometry of behavior. *Trends Ecol. Evol.* **23**, 394–401 (2008). doi:10.1016/j.tree.2008.03.005 Medline

9. T. L. Hedrick, A. A. Biewener, Low speed maneuvering flight of the rose-breasted cockatoo (*Eolophus roseicapillus*). I. Kinematic and neuromuscular control of turning. *J. Exp. Biol.* **210**, 1897–1911 (2007). doi:10.1242/jeb.002055 Medline

10. T. J. G. Read, P. S. Segre, K. M. Middleton, D. L. Altshuler, Hummingbirds control turning velocity using body orientation and turning radius using asymmetrical wingbeat kinematics. *J. R. Soc. Interface* **13**, 20160110 (2016). doi:10.1098/rsif.2016.0110 Medline

11. S. Sefati, I. D. Neveln, E. Roth, T. R. T. Mitchell, J. B. Snyder, M. A. Maciver, E. S. Fortune, N. J. Cowan, Mutually opposing forces during locomotion can eliminate the tradeoff between maneuverability and stability. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 18798–18803 (2013). doi:10.1073/pnas.1309300110 Medline

12. S. J. Arnold, Morphology, performance and fitness. *Integr. Comp. Biol.* **23**, 347–361 (1983).

13. R. B. Srygley, R. Dudley, Correlations of the position of center of body mass with butterfly escape tactics. *J. Exp. Biol.* **174**, 155–166 (1993).

14. D. R. Warrick, The turning- and linear-maneuvering performance of birds: The cost of efficiency for coursing insectivores. *Can. J. Zool.* **76**, 1063–1079 (1998). doi:10.1139/z98-044

15. B. Cheng, B. W. Tobalske, D. R. Powers, T. L. Hedrick, S. M. Wethington, G. T. C. Chiu, X. Deng, Flight mechanics and control of escape manoeuvres in hummingbirds. I. Flight kinematics. *J. Exp. Biol.* **219**, 3518–3531 (2016). doi:10.1242/jeb.137539 Medline

16. A. D. Straw, K. Branson, T. R. Neumann, M. H. Dickinson, Multi-camera real-time three-dimensional tracking of multiple flying animals. *J. R. Soc. Interface* **8**, 395–409 (2011). doi:10.1098/rsif.2010.0230 Medline

17. P. S. Segre, R. Dakin, V. B. Zordan, M. H. Dickinson, A. D. Straw, D. L. Altshuler, Burst muscle performance predicts the speed, acceleration, and turning performance of Anna's hummingbirds. *eLife* **4**, e11159 (2015). doi:10.7554/eLife.11159 Medline

18. C. J. Clark, Courtship dives of Anna's hummingbird offer insights into flight performance limits. *Proc. Biol. Sci.* **276**, 3047–3052 (2009). doi:10.1098/rspb.2009.0508 Medline

19. T. Alerstam, M. Rosén, J. Bäckman, P. G. P. Ericson, O. Hellgren, Flight speeds among bird species: Allometric and phylogenetic effects. *PLOS Biol.* **5**, e197 (2007). doi:10.1371/journal.pbio.0050197 Medline

20. C. J. Pennycuick, Fifteen testable predictions about bird flight. *Oikos* **30**, 165–176 (1978). doi:10.2307/3543476

21. J. H. Marden, Maximum lift production during takeoff in flying animals. *J. Exp. Biol.* **130**, 235–258 (1987).

22. C. H. Greenewalt, The flight of birds: The significant dimensions, their departure from the requirements for dimensional similarity, and the effect on flight aerodynamics of that departure. *Trans. Am. Philos. Soc.* **65**, 1–67 (1975). doi:10.2307/1006161

23. J. A. McGuire, C. C. Witt, J. V. Remsen Jr., A. Corl, D. L. Rabosky, D. L. Altshuler, R. Dudley, Molecular phylogenetics and the diversification of hummingbirds. *Curr. Biol.* **24**, 910–916 (2014). doi:10.1016/j.cub.2014.03.016 Medline

24. D. A. Skandalis, P. S. Segre, J. W. Bahlman, D. J. E. Groom, K. C. Welch Jr., C. C. Witt, J. A. McGuire, R. Dudley, D. Lentink, D. L. Altshuler, The biomechanical origin of extreme wing allometry in hummingbirds. *Nat. Commun.* **8**, 1047 (2017). doi:10.1038/s41467-017-01223-x Medline

25. P. S. Segre, R. Dakin, T. J. G. Read, A. D. Straw, D. L. Altshuler, Mechanical constraints on flight at high elevation decrease maneuvering performance of hummingbirds. *Curr. Biol.* **26**, 3368–3374 (2016). doi:10.1016/j.cub.2016.10.028 Medline

26. A. M. Wilson, J. C. Lowe, K. Roskilly, P. E. Hudson, K. A. Golabek, J. W. McNutt, Locomotion dynamics of hunting in wild cheetahs. *Nature* **498**, 185–189 (2013). doi:10.1038/nature12295 Medline

27. R. P. Wilson, I. W. Griffiths, M. G. L. Mills, C. Carbone, J. W. Wilson, D. M. Scantlebury, Mass enhances speed but diminishes turn capacity in terrestrial pursuit predators. *eLife* **4**, e06487 (2015). doi:10.7554/eLife.06487 Medline

28. H. C. Howland, Optimal strategies for predator avoidance: The relative importance of speed and manoeuvrability. *J. Theor. Biol.* **47**, 333–350 (1974). doi:10.1016/0022-5193(74)90202-1 Medline

29. A. Sih, A. M. Bell, J. C. Johnson, R. E. Ziemba, Behavioral syndromes: An intergrative overiew. *Q. Rev. Biol.* **79**, 241–277 (2004). doi:10.1086/422893 Medline

30. I. G. Ros, L. C. Bassman, M. A. Badger, A. N. Pierson, A. A. Biewener, Pigeons steer like helicopters and generate down- and upstroke lift during low speed turns. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 19990–19995 (2011). doi:10.1073/pnas.1107519108 Medline

31. J. M. Cheverud, Developmental integration and the evolution of pleiotropy. *Integr. Comp. Biol.* **36**, 44–50 (1996).

32. D. L. Altshuler, R. Dudley, S. M. Heredia, J. A. McGuire, Allometry of hummingbird lifting performance. *J. Exp. Biol.* **213**, 725–734 (2010). doi:10.1242/jeb.037002 Medline

33. N. A. Wright, D. W. Steadman, C. C. Witt, Predictable evolution toward flightlessness in volant island birds. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4765–4770 (2016). doi:10.1073/pnas.1522931113 Medline

34. A. Lindström, A. Kvist, T. Piersma, A. Dekinga, M. W. Dietz, Avian pectoral muscle size rapidly tracks body mass changes during flight, fasting and fuelling. *J. Exp. Biol.* **203**, 913–919 (2000). Medline

35. K. C. Welch Jr., D. L. Altshuler, Fiber type homogeneity of the flight musculature in small birds. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **152**, 324–331 (2009). doi:10.1016/j.cbpb.2008.12.013 Medline

36. O. B. O. Savile, Adaptive evolution in the avian wing. *Evolution* **11**, 212–224 (1957). doi:10.1111/j.1558-5646.1957.tb02889.x

37. A. Hedenström, M. Rosén, Predator versus prey: On aerial hunting and escape strategies in birds. *Behav. Ecol.* **12**, 150–156 (2001). doi:10.1093/beheco/12.2.150

38. J. G. Burns, R. C. Ydenberg, The effects of wing loading and gender on the escape flights of least sandpipers (*Calidris minutilla*) and western sandpipers (*Calidris mauri*). *Behav. Ecol. Sociobiol.* **52**, 128–136 (2002). doi:10.1007/s00265-002-0494-y

39. F. L. Carpenter, D. C. Paton, M. A. Hixon, Weight gain and adjustment of feeding territory size in migrant hummingbirds. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 7259–7263 (1983). doi:10.1073/pnas.80.23.7259 Medline

40. K. Lilliendahl, Daily patterns of body mass gain in four species of small wintering birds. *J. Avian Biol.* **33**, 212–218 (2002). doi:10.1034/j.1600-048X.2002.330302.x

41. J. W. Kruyt, E. M. Quicazán-Rubio, G. F. van Heijst, D. L. Altshuler, D. Lentink, Hummingbird wing efficacy depends on aspect ratio and compares with helicopter rotors. *J. R. Soc. Interface* **11**, 20140585 (2014). doi:10.1098/rsif.2014.0585 Medline

42. P. Chai, Hummingbird hovering energetics during moult of primary flight feathers. *J. Exp. Biol.* **200**, 1527–1536 (1997). Medline

43. C. Ellington, The aerodynamics of hovering insect flight. II. Morphological parameters. *Philos. Trans. R. Soc. B.* **305**, 17–40 (1984). doi:10.1098/rstb.1984.0050

44. C. J. Pennycuick, Gliding flight of the white-backed vulture *Gyps Africanus*. *J. Exp. Biol.* **55**, 13–38 (1971).

45. C. J. Clark, T. J. Feo, I. Escalante, Courtship displays and natural history of scintillant (*Selasphorus scintilla*) and volcano (*S. flammula*) hummingbirds. *Wilson J. Ornithol.* **123**, 218–228 (2011). doi:10.1676/10-076.1

46. P. Chai, J. S. Chen, R. Dudley, Transient hovering performance of hummingbirds under conditions of maximal loading. *J. Exp. Biol.* **200**, 921–929 (1997). Medline

47. P. Chai, D. Millard, Flight and size constraints: Hovering performance of large hummingbirds under maximal loading. *J. Exp. Biol.* **200**, 2757–2763 (1997). Medline

48. D. L. Altshuler, R. Dudley, J. A. McGuire, Resolution of a paradox: Hummingbird flight at high elevation does not come without a cost. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17731–17736 (2004). doi:10.1073/pnas.0405260101 Medline

49. J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, *nlme 3.1-131: Linear and Nonlinear Mixed Effects Models* (2017); https://cran.r-project.org/web/packages/nlme/index.html.

50. R Core Team, *R 3.3.3: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2017); www.R-project.org.

51. S. Epskamp *et al*., *qgraph 1.4.3: Graph Plotting Methods, Psychometric Data Visualization and Graphical Model Estimation* (2017); https://cran.rproject.org/web/packages/qgraph.

52. T. Jombart *et al*., *adegenet 2.0.1: Exploratory Analysis of Genetic and Genomic Data* (2015); https://cran.r-project.org/web/packages/adegenet.

53. T. Jombart, S. Devillard, F. Balloux, Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94 (2010). doi:10.1186/1471-2156-11-94 Medline

54. F. Keck, *phylosignal 1.1: Exploring the Phylogenetic Signal in Continuous Traits* (2015); https://cran.r-project.org/web/packages/phylosignal.

55. M. Pagel, Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884 (1999). doi:10.1038/44766 Medline

56. W. A. Barr, R. S. Scott, Phylogenetic comparative methods complement discriminant function analysis in ecomorphology. *Am. J. Phys. Anthropol.* **153**, 663–674 (2014). doi:10.1002/ajpa.22462 Medline

57. M. van de Pol, J. Wright, A simple method for distinguishing within- versus between-subject effects using mixed models. *Anim. Behav.* **77**, 753–758 (2009). doi:10.1016/j.anbehav.2008.11.006

58. J. Hadfield, *MCMCglmm 2.24: MCMC Generalised Linear Mixed Models* (2016); https://cran.r-project.org/web/packages/MCMCglmm/index.html.

59. P. de Villemereuil, S. Nakagawa, in *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*, L. Z. Garamszegi, Ed. (Springer, 2014), pp. 287–303; http://link.springer.com/chapter/10.1007/978-3-662-43550-2_11.

60. S. Nakagawa, H. Schielzeth, A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* **4**, 133–142 (2013). doi:10.1111/j.2041-210x.2012.00261.x

61. P. Breheny, W. Burchett, *visreg 2.2-2: Visualization of Regression Models* (2016); https://cran.r-project.org/web/packages/visreg.

62. R. Suzuki, H. Shimodaira, *pvclust 2.0-0: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling* (2015); https://cran.r-project.org/web/packages/pvclust.

63. M. Maechler *et al*., *cluster 2.0.5: "Finding Groups in Data": Cluster Analysis Extended Rousseeuw* et al. (2016); https://cran.r-project.org/web/packages/cluster.

64. L. Kaufman, P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis* (John Wiley & Sons, 2009).