Title: Wing length canalisation and behaviour across birds: a phylogenetic meta-analysis of variance

Authors: Klaus Reinhold, Alfredo Sánchez-Tójar

Affiliation: Department of Evolutionary Biology, Bielefeld University, 33615, Bielefeld, Germany

Corresponding author: Klaus Reinhold (klaus.reinhold@uni-bielefeld.de)

Abstract

Stronger stabilising selection is expected to lead to a decrease in trait variation (i.e., in higher canalisation). We examined this prediction across species by investigating individual variation in wing length across measured as the coefficient of variation (*CV*). We hypothesised that species that heavily rely on aerial feeding as well as long-distance migratory species should show higher canalisation (lower *CV*) in wing length than non-aerial feeders and non-migratory species. We collected species- and population-specific summary statistics on wing length for males (*k* = 340 *CV* values) and females (*k* = 310) from the literature (172 species, 314 populations) and analysed them using recently developed meta-analytic metrics for integrating phenotypic variance estimates. Our phylogenetic multilevel meta-analysis showed that heterogeneity among *CV* values was relatively low, indicating generalisability and replicability of the overall CV value (2.6%). Nonetheless, about a third of the heterogeneity was associated with phylogenetic relationships (23%) and differences among species (8%), a precondition to test for species-specific drivers of those differences. Although not all pairwise comparisons were statistically significant, all our analyses both at the species and the population level robustly and consistently showed higher canalisation in aerial (*CV* = 2.5% [2.2 to 2.9]) compared to non-aerial feeders (*CV* = 2.7% [2.4 to 3.0]), and in migratory (*CV* = 2.5% [2.2 to 2.9]) compared to non-migratory species (*CV* = 2.8% [2.4 to 3.2]). We conclude that wing length in bird species relying on their wings more heavily is likely under stronger (stabilising) selection, which in turn would have led to the observed higher canalisation on this trait for those species. Our study showcases how to combine already available descriptive statistics for phenotypic traits with underused meta-analysis of variance approaches to test often-neglected evolutionary predictions at the variance level. We hope to inspire others to expand our phylogenetic study to more species and life-history traits as well as to other wing traits of higher dimensionality such as wing area and shape, and more generally, to study canalisation across species and traits.

Keywords: meta-analysis of variance; evidence synthesis; avian; comparative analysis; coefficient of variation; canalization; replicability; generalizability; life-history evolution

Introduction

Much progress has been made in understanding phenotypic traits in recent years; however, most research has focused on understanding mean phenotypic differences (e.g., Fattorini et al. 2023, Richardson et al. 2023, Nokelainen et al. 2024) and less so on understanding differences in phenotypic variance (but see Kleven et al. 2008, Lifjeld et al. 2010, Sánchez-Tójar et al. 2020, Moran et al. 2021). This lack of research on phenotypic variance is surprising given the key role that it plays in evolution and its importance in the study of selection, evolvability and canalisation (Mitchell-Olds et al. 2007, Willmore et al. 2007, Geiler-Samerotte et al. 2013, Bolnick et al. 2011). All else being equal, including indirect genetic effects, which can substantially alter adaptive potential (Santostefano et al. 2025), traits that differ in additive genetic variance will evolve at different speeds when exposed to the same strength of selection, regardless of whether directional or stabilising selection is taking place (Price 1972). Nonetheless, the amount of variance in a quantitative trait depends on both the type of selection occurring and the effects of the environment on the trait (i.e., phenotypic plasticity). First, both stabilising and directional selection are expected to decrease additive genetic variance, whereas disruptive and fluctuating selection should increase it. Second, phenotypic plasticity can potentially increase trait variance if individuals experience different environments and if canalisation is imperfect. Similarly, bet-hedging strategies, independent of whether these refer to within- or among-individual developmental variance, should also lead to an increase in phenotypic variance. Despite our expectation that trait variance is expected to be mostly driven by selection and affected by phenotypic plasticity, trait variance can increase due to other processes. This includes sources such as developmental stochasticity but also sources such as measurement error. Indeed, trait variance might be increased due to random deviations occurring during development, even when strong stabilising selection is expected to reduce this developmental load, leading to stronger canalisation (i.e., less variance) during trait development. Lastly, although measurement error might be estimated and accounted for via repeated measurement designs, it is an unavoidable source of trait variance.

From the processes listed above, canalisation has received comparatively little attention. For example, variation in sperm length in birds seems negatively correlated with the rate of extra-pair paternity (Kleven et al. 2008, Lifjeld et al. 2010), whereas variation in conspicuous plumage traits under strong selection is surprisingly large (Delhey et al. 2017). In insects and frogs, traits used in acoustic communication that are under strong directional selection via female choice have been shown to be less variable than acoustic traits under weaker directional selection (Reinhold 2011). Nonetheless, studies on directional selection have shown compelling evidence for both a decrease in genetic trait variance (and, thus, also in phenotype trait variance) and an increase in phenotypic variance via disrupting developmental canalisation (Groth et al. 2020; but see Hayden et al. 2014). Whereas disruptive selection has been shown to generally lead to lower canalisation (Pelabon et al. 2010), several experimental studies suggest that trait canalisation increases with stabilising selection (e.g., Prout 1962), and that traits closely correlated with fitness are usually strongly canalised (Stearns et al. 1995; but see Young 2006). Overall, few studies have examined the extent of trait canalisation among species and for those that have, results remain contradictory. Consequently, the importance of this variance-reduction process remains unclear.

Here, we make use of recently developed meta-analytic metrics for integrating variance estimates across studies (Nakagawa et al. 2015; Noble et al. 2021) to study trait canalisation on an essential bird trait: wing length. There are several reasons that make bird wing length a great model trait to study canalisation. First, the heritability of bird wing length is relatively high, ranging from 0.4 to 0.6 (Cava 2019, Jensen 2003, Tarka et al 2014), and this trait has been shown to be under stabilising selection (Hall et al. 2004) as well as to respond to environmental effects (Yom-Tov et al. 2006). Second, wing length is comparatively little affected by measurement error (Subasinghe et al. 2021) and often measured and reported in articles studying questions other than canalisation (e.g., sexual-dimorphism or population differentiation; Selander 2017, James 1970), which reduces the likelihood that our conclusions would be affected by publication bias. Our main hypothesis was that bird wing length variance is smaller in species that are more likely to experience strong stabilising selection on this trait. Specifically, we tested whether phenotypic variance in wing length is smaller in species showing aerial foraging and in long-distance migratory species compared to non-aerial feeding and resident species, respectively. We chose those two flight-related behaviours because deviations from an optimal wing length are expected to be more costly for species that rely on their wings more heavily, consequently leading to strong stabilising selection on wing length. Besides our main hypothesis, we used the collected data to assess the replicability of the sex-chromose hypothesis for wing length (Reinhold & Engqvist 2013, Nakagawa et al. 2015).

Methods

Search strategy

We collected wing length measurements (i.e., mean, standard deviation [SD], sample size [n]) for male and female birds from two sources: (1) a dataset used to test the sex-chromosome hypothesis (Reinhold and Engqvist 2013), and (2) an update of such dataset which we performed using similar keyword search terms, specifically: (“coefficient of variation” OR “CV?” OR “variation in size”) AND (“wing size“ OR “wing length“) AND (“bird$” OR [*a list of all bird genera*]) [for the complete list of bird genera see Capilla-Lasheras et al. 2022]. We searched for studies published between 2011 and 2019 in the Web of Science Core Collection (databases: Science Citation Index Expanded (SCIE); Social Sciences Citation Index (SSCI); Arts & Humanities Citation Index (AHCI); Emerging Sources Citation Index (ESCI)), and an adjusted combination of those keywords were used to perform additional searches in Biological Abstracts and Google Scholar. Last, to reduce a potential bias towards species with strong sexual dimorphism, we also performed additional searches specifically targeting studies using molecular sex determination by adding the keywords ("molecular sexing" OR "molecular sex determination") to the search string detailed above. Importantly, our search was not intended to be fully systematic nor comprehensive but representative and unbiased across the bird phylogeny, as well as built on our previous work (Reinhold and Engqvist 2013). We believe that the full search strategy detailed above, which led to a representative sample of 172 bird species (Figure 1) accomplished those goals.

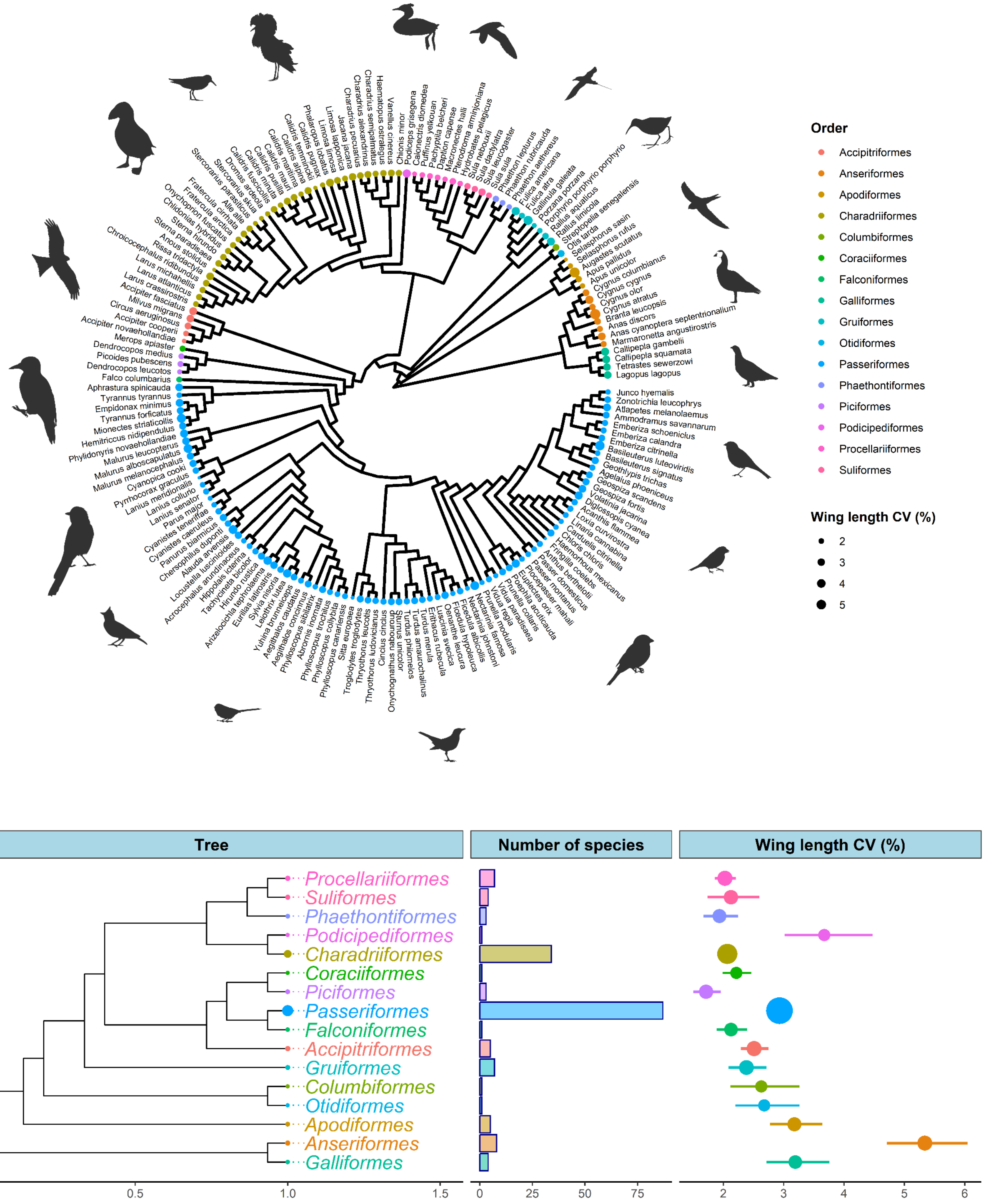

Figure 1. The panel on top shows the full phylogenetic tree of all 172 bird species included in the phylogenetic multilevel meta-analyses performed to understand canalisation in wing length. The three panels below show the corresponding phylogenetic tree at the order level (16 orders) with its corresponding number of species, and the mean and 95% confidence interval for wing length *CV* (%). Dot size is relative to the number of observations. Silhouette images under Creative Commons licenses obtained from Phylopic (https://www.phylopic.org/).

Data description

As in our previous study (Reinhold and Engqvist 2013), our effect size of interest was the coefficient of variation (hereafter *CV*), which is calculated as the SD of wing length divided by its mean wing length. Therefore, we extracted wing length measurements (mean, SD, n) for male and female birds separately for all studies (N = 181). Whenever necessary, standard errors (SE) were multiplied by the square root of their sample size to calculate their corresponding SD. We only considered measurements for which the SD had a two-decimal point accuracy to avoid inaccurate extreme outliers. We prioritized measurements taken only from adults, and thus, juvenile-only measurements and measurements combining both juveniles and adults were not extracted. Juveniles often systematically differ from adults in their wing length and including them would lead to an uninformative increase in trait variance. In addition, whereas estimates were extracted as averages across populations in our previous study (Reinhold and Engqvist 2013), for our current study, we revisited all the references included in Reinhold and Engqvist (2013) to extract separate estimates from different populations or sampling times from the same reference whenever possible, and model that non-independence in our multilevel models. We did so to avoid potentially inflating *CV* values due to geographical variation (i.e., clines) and/or temporal variation in wing length (see James 1970, Yom-Tov et al. 2006). As an additional safeguard, we did not extract data from museum collections because these usually combine measurements taken in different years and from several populations, which would also likely lead to uninformative inflated *CV* values. Last, we excluded estimates based on a single individual, for which SD or n were missing, and estimates with an SD equal to zero and based on three individuals or less (e.g., *Malurus alboscapulatus*; Swaddle et al. 2000).

To better understand canalisation in wing length across bird species, we extracted species-specific and population-specific information for two behavioural traits that we hypothesised should lead to different levels of selective pressure on wing length among species. For each species, we classified its feeding type (three levels: fully aerial feeding, partially aerial feeding, non-aerial feeding) and its migratory behaviour (three levels: fully migratory, partially migratory, non-migratory), and we refer to analyses using these two variables as being at the “species-specific level”. In addition to this species-specific level, we used a population-specific classification in which, for those species categorised as “partially migratory”, we classified the migratory behaviour of each specific population as either non-migratory or fully migratory. To arrive at a comparable binary classification, we created an additional feeding type variable (two levels: fully aerial feeding, non-aerial feeding) in which we only classified birds as feeding on the wing, if they show mostly aerial feeding. Although the latter was not strictly speaking at the population level, we refer to analyses using these two binary variables as being at the “population-specific level”.

To classify migratory behaviour, we referred to maps provided by the IUCN (https://datazone.birdlife.org) and Wikipedia (https://www.wikipedia.org/), and if insufficient information was available from those sources, we also searched for information published in the scientific literature by performing searches in Google Scholar. By using the distance in the maps, we classified bird species as migratory if most individuals can be expected to have two annual migratory periods within which they regularly travel more than 500 km between the breeding and wintering grounds in each direction, and if not, those species were classified either as partially migratory or residents (see below). Birds species that were all-year resident in less than ca. 20% of their breeding distribution were classified as migratory (e.g., *Sterna hirundo* or *Ficedula hypoleuca*) whereas species that only migrated in an area that covered between ca. 20% and ca. 80% of the breeding distribution were classified as partially migratory (e.g., *Accipiter cooperii* or *Linaria cannabina*) and species that were residents over more than ca. 80% of their distribution were classified as non-migratory (e.g., *Pyrrhocorax graculus* or *Parus major*). Bird species that nest close to the sea and are vagrants during the non-breeding period were not classified as migrants if they did not show directed and temporally coherent movements (i.e., two annual migratory periods; e.g., *Puffinus yelkouan* or *Larus atlanticus*), as we assumed that stabilising selection should only be strong for longer periods of continuous migration. However, since vagrants usually fly over the ocean for extended periods of time, we reran all our analyses after classifying vagrants as fully migratory rather than non-migratory. These sensitivity analyses not only confirmed our main conclusions but tended to show stronger evidence for the existence of higher wing length canalisation in fully and partial migratory vs non-migratory at both the species and the population level (see "Supplementary Material S5").

To classify feeding type, we referred to the same sources as for the previous (i.e., IUCN and Wikipedia), and if insufficient information was available from those sources, we also searched for information published in the scientific literature by performing searches in Google Scholar. We classified bird species as showing fully aerial feeding if they mainly collected their food by using their wings, which included species that mainly catch flying prey (e.g., *Hirundo rustica* or *Tyrannus tyrannus*), plunge dive (e.g., gannets), are flying surface skimmers (e.g., *Puffinus yelkouan*), fly underwater (e.g., *Fratercula cirrhata*) or hover in front of flowers (e.g., hummingbirds). Birds species showing such feeding behaviours frequently but also frequently feeding while climbing, walking, hopping, or swimming with their feet (e.g., *Phylloscopus canariensis or Cinclus cinclus*) were classified as showing partially aerial feeding, whereas the remaining bird species, which may only sometimes or never feed 'on the wing' were classified as showing non-aerial feeding.

Statistical analyses

For all analyses, we calculated the log coefficient of variation (hereafter *lnCV*; equation 1) and its associated sampling variance (equation 2) using the function 'escalc()' from the R package 'metafor' v.4.6-0 (Viechtbauer 2010) in R v.4.3.1 (R Core Team 2023), which includes a small-size bias correction (more in Nakagawa et al. 2015). Although we used *lnCV* for all the analyses, we present back-transformed percentage values throughout (i.e., exp(*lnCV*)*100) to aid biological interpretation. Our visualizations were performed following general recommendations in Yang et al. (2023) and primarily using the R packages 'ggtree' (Yu et al. 2017) and 'orchaRd' v.2.0 (Nakagawa et al. 2023).

$$lnCV = log(\frac{SD}{mean}) + \frac{1}{2*(n-1)},$$ (eqn 1)

$$V_{lnCV} = \frac{SD^2}{n*mean^2} + \frac{1}{2*(n-1)},$$ (eqn 2)

We ran phylogenetic multilevel meta-analytic and meta-regression models with *lnCV* as the response variable and where sampling variances were fit as a variance-covariance matrix that assumed a correlation of 0.5 between sampling variances obtained from the same study following recommendations in Noble et al. (2017). To further model nonindependence, we included the following six random effects in all models: (1) study identity, which encompassed estimates extracted from the same primary study, (2) pair identity, which encompassed male-female pair estimates obtained from the same population and sampling event, (3) population identity, which encompassed estimates obtained from the same geographic location, (4) phylogenetic correlation matrix, which modelled the phylogenetic relationships among the species included in our dataset, (5) species, which accounted for among-species variation additional to phylogeny (Cinar et al. 2022), and (6) unit-level observation identity, which modelled within-study variance. For the phylogenetic correlation matrix among species, we extracted the phylogenetic information from the Open Tree of Life using the R package 'rotl' v.3.1.0 (Michonneau et al. 2016) and computed branch lengths using the Grafen method as implemented in the R package 'ape' v.5.7-1 (Paradis and Schliep 2019). We followed similar steps to generate a phylogenetic tree at the order level to aid in visualizing general differences in wing length *CV* among orders (Figure 1). We performed several sensitivity analyses that modelled the variance-covariance matrix assuming a 0.25, 0.5 and 0.75 correlation between sampling variances with the same study identity, pair identity and population identity, and results remained (virtually) the same. We provide all those sensitivity models in our data and code repository ("models/sensitivity_analysis/").

We first ran a phylogenetic intercept-only multilevel meta-analytic model to (i) estimate the mean wing length *CV* across estimates while accounting for all the sources of nonindependence detailed above and (ii) explore heterogeneity among effect sizes. We then ran five phylogenetic multilevel uni-moderator meta-regressions (i.e., only one moderator included at a time) with the goal of testing our predictions regarding canalisation and the replicability of the sex-chromosome hypothesis. First, we tested whether wing length *CV* values differed between sexes (levels: male, female; "sex-chromosome hypothesis"). Then, we tested whether some of the heterogeneity among effect sizes could be explained by differences among species and populations in feeding type or migration status. Last, we also ran two additional models including both moderators (i.e., feeding and migratory type), one at the species-specific level (Table S2) and another at the population-specific level (Table S3), to estimate the total amount of heterogeneity explained when both moderators are included in the same model.

For the intercept-only models, we estimated heterogeneity across estimates using the pluralistic approach suggested by Yang et al. (2024) and implemented in the R package 'orchaRd' v.2.0 (Nakagawa et al. 2023). This pluralistic approach involves estimating total heterogeneity (σ^2), its source (or source of heterogeneity; I^2) and two different metrics for its magnitude – a mean-standardized metric (*CVH2*) and a variance-mean-standardized metric (*M2*). For the meta-regressions, we report the percentage of variance explained by the moderator(s) as R^2_{marginal} (Nakagawa & Schielzeth, 2017), which we also calculated using the R package 'orchaRd'. Throughout, we present mean estimates with their associated 95% confidence intervals (CI) and prediction intervals (PI).

Results

Our final dataset contained 340 male estimates and 310 female estimates obtained from 181 studies from 314 different geographic locations. We obtained data from a total of 172 bird species for males (median = 1 estimate/species, range: 1-23; mean = 1.98, SD = 2.44; median = 42 individuals/estimate, range: 2-3179) and 163 species for females (median = 1 estimate/species, range: 1-25; mean = 1.90, SD = 2.47; median = 36 individuals/estimate, range: 2-4116; Figure 1). Our dataset covered 119 genera from 63 families belonging to 16 orders of birds, with the most common order being Passeriformes (87 species), and the two most common families being Scolopacidae (11 species) and Laridae (10 species). In all, 407 estimates came from our previous compilation (Reinhold and Engqvist 2013) to which we added 243 additional estimates through our searches (see Supplementary Material S1-2 for additional descriptive statistics).

Overall canalisation and heterogeneity

Our phylogenetic multilevel meta-analysis showed that the mean wing length *CV* value across species was 2.63% (95% CI = 2.32 to 2.98, 95% PI = 1.55 to 4.45) with most differences among effect sizes coming from sources other than sampling error (I^2_{total} = 94.2%). Although I^2_{total} was high and indicated that heterogeneity in our dataset is, on average, around 16 times larger than statistical noise, absolute heterogeneity was relatively small (σ^2 = 0.068). Indeed, the other two heterogeneity metrics showed that the magnitude of the heterogeneity in our data set is small ($CVH2_{total}$ = 0.005; $M2_{total}$ = 0.005; Table S1, Figure S4), suggesting that the meta-analytic mean obtained is generalisable and replicable across different contexts. The discrepancy between I^2 with both *CVH2* and M2 is likely the consequence of the small typical sampling variance found in our dataset ($\bar{\nu}$ = 0.004), which can make the interpretation of I^2 challenging – something often encountered with similar effect size measures such as the log coefficient of variation ratio (*lnCVR*; more in Yang et al. 2024) as well as the much larger meta-analytic mean (*lnCV* = -3.69) compared to σ^2.

Sex-chromosome hypothesis

Our phylogenetic multilevel meta-regression showed that, although in the expected direction, the mean wing length *CV* did not differ statistically (p-value = 0.176) between males (mean *CV* = 2.61%, 95% CI = 2.30 to 2.97) and females (mean *CV* = 2.65%, 95% CI = 2.34 to 3.02) and, indeed, the variance explained by this moderator was negligible (R^2_{marginal} = 0.10%; Figure S5).

Species-specific level

We ran two uni-moderator phylogenetic multilevel meta-regressions to understand among-species differences in wing length canalisation. First, non-aerial feeding was associated with larger wing length *CV* values (mean *CV* = 2.68%, 95% CI = 2.39 to 3.02) than those found in species showing partially (mean *CV* = 2.42%, 95% CI = 2.09 to 2.79) and fully aerial feeding (mean *CV* = 2.54%, 95% CI = 2.23 to 2.88), but the difference was only statistically significant for partially aerial feeding (p-value = 0.044) and not for fully aerial feeding (p-value = 0.231; Figure 2A). Second, non-migratory behaviour was associated with larger wing length *CV* values (mean *CV* = 2.77%, 95% CI = 2.41 to 3.19) than those found in partially migratory (mean *CV* = 2.64%, 95% CI = 2.28 to 3.04) and fully migratory species (mean *CV* = 2.51%, 95% CI = 2.19 to 2.89), but the difference was only statistically significant for fully migratory species (p-value = 0.017) and not for partially migratory species (p-value = 0.266; Figure 2B). The percentage of variance explained by feeding (R^2_{marginal} = 2.08%) and migratory type (R^2_{marginal} = 2.63%) separately were relatively small, but both moderators combined explained up to 4.05% of the heterogeneity (Table S2).

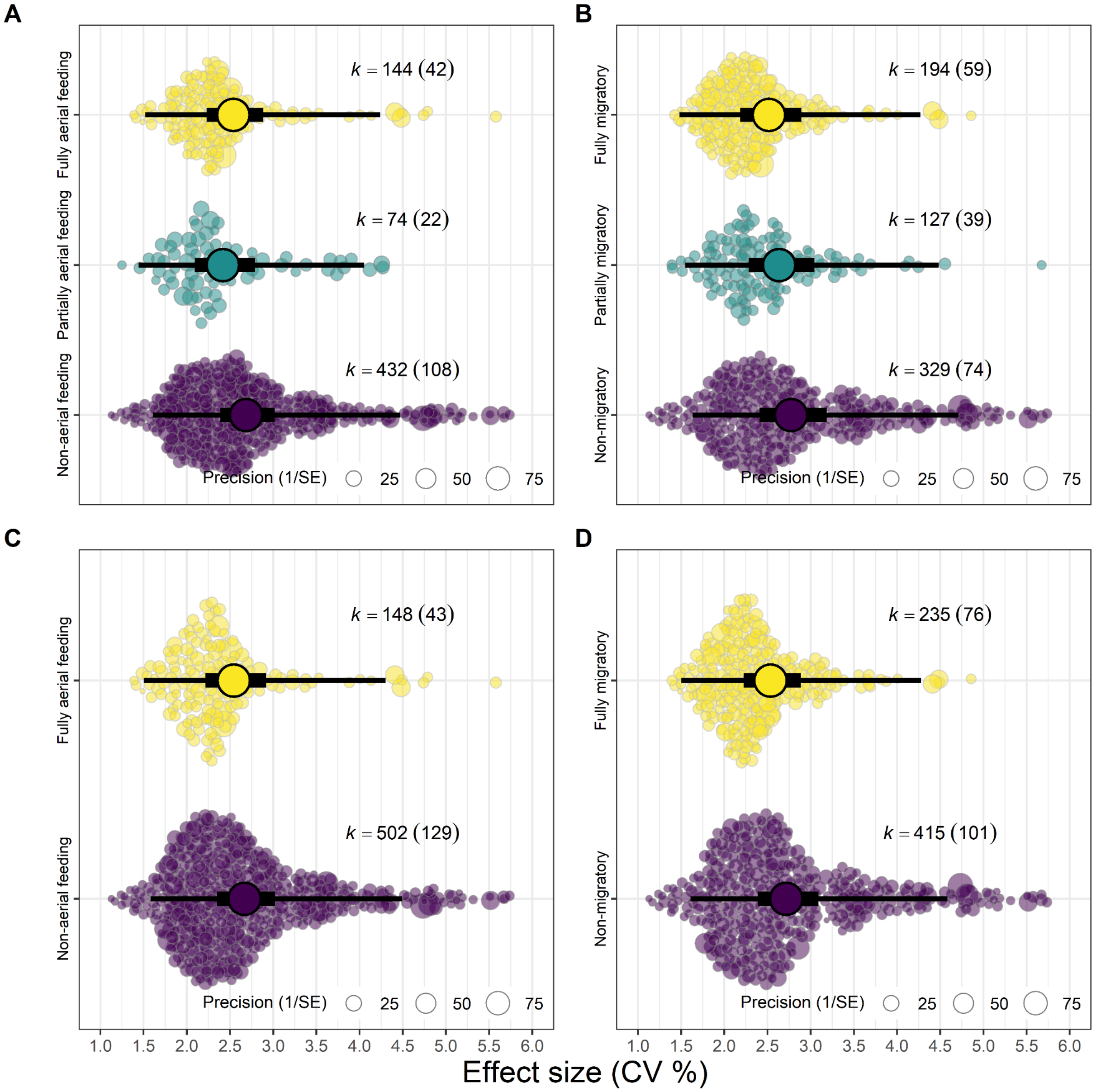

Figure 2. Flight-related behaviours such as partial or aerial feeding and migration are generally associated with higher wing length canalisation (i.e., smaller values of coefficient of variation, *CV*) across 172 bird species, but only some of the pairwise comparisons were statistically significant (see text). Orchard plots show the mean estimate, 95% confidence intervals (thick whisker) and 95% prediction intervals (thin whisker), with dot size being scaled by effect size precision (i.e., 1/SE). The x-axes are cropped between *CV* values of 1 and 6 to aid visualization, see section "S5. Supplementary tables and figures" for the full version.

Population-specific level

We ran two additional uni-moderator phylogenetic multilevel meta-regressions to understand if differences in wing length canalisation remained when species' feeding and migratory type were explored at the population level (e.g., different populations of a partially migratory species may differ in whether they are full or non-migratory). First, non-aerial feeding was associated with larger wing length

CV values (mean *CV* = 2.67%, 95% CI = 2.35 to 3.02) than those found for fully aerial feeding (mean *CV* = 2.54%, 95% CI = 2.22 to 2.92), but the difference was not statistically significant (p-value = 0.310; Figure 2C). Second, non-migratory behaviour was associated with statistically significantly larger wing length *CV* values (mean *CV* = 2.72%, 95% CI = 2.39 to 3.09) than those found for fully migratory populations (mean *CV* = 2.53%, 95% CI = 2.23 to 2.89; p-value = 0.040; Figure 2D). The percentage of variance explained by feeding (R^2_{marginal} = 0.58%) and migratory type (R^2_{marginal} = 1.63%) separately were small and both moderators combined only explained up to 1.95% of the heterogeneity (Table S3).

Discussion

The overall variation in bird wing length across species was small, suggesting high canalisation on this trait. The mean *CV* value was estimated to have a 95% confidence interval between 2.3. and 3.0%, values that are clearly smaller than those observed in human body height and sperm length (Mossman et al. 2013), and somewhat larger than those observed for sperm length in birds (Kleven et al. 2008) and dung flies (Sharma et al. 2013). Although heterogeneity among *CV* values was relatively small, phylogeny and species combined explained a third of it, confirming the necessary condition for expecting that among-species differences in wing use can be important drivers in explaining among-species differences in wing length *CV* values.

Indeed, our results showed evidence indicating that among-species differences in wing use explain some of the observed among-species differences in wing length *CV* values, where species that rely on their wings more heavily showing higher canalisation on wing length. Despite that not all differences were statistically significant, bird species that mostly or partially feed on the wing showed, on average, around 5 to 10% lower wing length *CV* values than species not feeding on the wing. Along the same lines, and again despite not all differences being statistically significant, bird species that perform full or partial migration showed, on average, around 7 to 9% lower wing length *CV* values than resident species. Importantly, the results regarding migratory behaviour became clearer when we categorised vagrant species such as *Puffinus yelkouan*, *Phaethon aethereus* and *Sula sula* that fly over the ocean for extended periods of time as migratory rather than resident (Supplementary Material S5), further confirming that species heavily relying on their wings show lower wing length *CV* values. Feeding and migratory type combined explained 4 to 6% of the observed differences in CV values, and in all, provide preliminary support for our hypothesis that wing length should be under stronger stabilising selection in bird species strongly relying on their wings, leading to higher canalisation on this trait. These results would, therefore, agree with previous among-species studies on other animal groups showing that traits under strong selection show low variation (i.e., higher canalisation). For example, previous among-species studies have shown that acoustic courtship traits under strong selection in amphibians and insects show lower variation (Reinhold 2011), and that canalisation in sperm length is stronger in bird species that show higher rates of extra-pair matings (Lifjeld et al. 2010). In line with these results, a recent meta-analysis of variance (Noble et al. 2021) has shown small between-individual variation in body size of hatchlings in snakes, lizards and turtles. Strong canalization seems to be present, as no effect of incubation temperature could be observed for this trait that likely is under strong selection.

Finally, using our updated dataset we assessed the replicability and generalizability of the "sex-chromosome hypothesis", which predicts higher variability in body size in the heterogametic sex compared to the homogametic sex, and for which evidence has been found in mammals, insects and birds (Reinhold & Engqvist 2013). Our results did not show clear evidence for the "sex-chromosome hypothesis" for bird wing length, as the difference between males and females was small and not statistically significant, and the moderator explained a negligible amount of heterogeneity. The observed effect was, nonetheless, in the expected direction, with males (CV_{males} = 2.61) showing, on average, about 1.5% lower wing length *CV* values than females ($CV_{females}$ = 2.65), which is lower than the mean effect size previously found for general body size in birds (Reinhold & Engqvist 2013: 4.0%). Our estimate for the difference in *CV* between males and females was also smaller than what was found in a recent re-analysis of Reinhold & Engqvist (2013)'s data using a phylogenetic multilevel meta-analysis similar to the one we used in our study (Nakagawa et al. 2015). The main difference between our data and that of Reinhold & Engqvist (2013) and its reanalysis by Nakagawa et al. (2015) is that ours focuses on wing length only, whereas Reinhold & Engqvist's (2013) focused on general body size (i.e., body mass and size, wing, head and beak size). To confirm that the difference in trait coverage likely led to the difference in results, we re-analysed the subset of Nakagawa et al. (2015) corresponding to bird wing length. This re-analysis, which included only a third of the studies and about half of the bird species included in our study, confirmed little support for the "sex-chromosome hypothesis": males showed, on average, a non-statistically significant 1.5% lower wing length *CV* values than females (Supplementary Material S6). Our replication attempt showing that the sex-chromosome hypothesis explains negligible variation among wing length CV values suggests that the previous evidence for the sex-chromosome hypothesis has been likely underestimated by including this trait. Therefore, we call for future research reassessing the sex-chromosome hypothesis across animals with a special focus on more variable traits such as body mass.

Conclusions

Our analyses indicate that coefficients of variation in wing length are about 7-9% smaller in migratory bird species compared to non-migratory bird species, and that birds feeding on the wing have about 5-10% smaller values than species that do not usually feed on the wing. Although not all comparisons were statistically significant, they were robust and consistent, supporting the expectation that stabilising selection on wing length is stronger for species using their wings more intensely during foraging and migration. The observed differences between these groups only explained a small proportion of the total variation, likely because a large proportion of the remaining variance is attributed to phylogenetic differences. This suggests that different levels of canalisation have evolved within the different bird taxa. We suggest that future research should make use of the meta-analytic techinques showcased in our study together with large-scale datasets to study trait canalisation across taxonomic groups.

Authors contributions

KR: Conceptualization; Data curation; Funding acquisition; Investigation; Methodology; Project Administration; Supervision; Writing – original draft; Writing – review & editing. AST: Data curation; Formal Analysis; Investigation; Methodology; Project Administration; Software; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

Acknowledgements

We would like to thank Oliver Heerwig for his help with data extraction, and Daniel W.A. Noble for kindly adding the possibility to easily back-transform *lnCV* values to the R package 'orchaRd'. This study was partially funded by the German Research Foundation (DFG: Deutsche Forschungsgemeinschaft) as part of the SFB TRR 212 (NC3)—Project No. 316099922 and 396782608.

Data and code availability statement

All data and code are available at the following GitHub repository (https://github.com/ASanchez-Tojar/meta-analysis_canalisation_bird_wing). Upon acceptance, all data and code will be provided with a DOI via Zenodo.

Conflict of interest statement

The authors declare no competing interests.

References

Alonso, J.C., Magaña, M., Alonso, J.A., Palacín, C., Martín, C.A. & Martín, B. (2009) The Most Extreme Sexual Size Dimorphism Among Birds: Allometry, Selection, and Early Juvenile Development in the Great Bustard (*Otis tarda*). *The Auk* **126**, 657–665.

Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M., Novak, M., Rudolf, V.H.W., Schreiber, S.J., Urban, M.C. & Vasseur, D.A. (2011) Why intraspecific trait variation matters in community ecology. *Trends in Ecology & Evolution* **26**, 183–192.

Capilla-Lasheras, P., Thompson, M.J., Sánchez-Tójar, A., Haddou, Y., Branston, C.J., Réale, D., Charmantier, A. & Dominoni, D.M. (2022) A global meta-analysis reveals higher variation in breeding phenology in urban birds than in their non-urban neighbours. *Ecology Letters* **25**, 2552–2570.

Cava, J.A., Perlut, N.G. & Travis, S.E. (2019) Heritability and evolvability of morphological traits of Savannah sparrows (*Passerculus sandwichensis*) breeding in agricultural grasslands. *PLOS ONE* **14**, e0210472.

Cinar, O., Nakagawa, S. & Viechtbauer, W. (2022) Phylogenetic multilevel meta-analysis: A simulation study on the importance of modelling the phylogeny. *Methods in Ecology and Evolution* **13**, 383–395.

Delhey, K., Szecsenyi, B., Nakagawa, S. & Peters, A. (2017) Conspicuous plumage colours are highly variable. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20162593.

Fattorini, N., Costanzo, A., Romano, A., Rubolini, D., Baillie, S., Bairlein, F., Spina, F. & Ambrosini, R. (2023) Eco-evolutionary drivers of avian migratory connectivity. *Ecology Letters* **26**, 1095–1107.

Geiler-Samerotte, K., Bauer, C., Li, S., Ziv, N., Gresham, D. & Siegal, M. (2013) The details in the distributions: why and how to study phenotypic variability. *Current Opinion in Biotechnology* **24**, 752–759.

Groth, B.R., Huang, Y., Monette, M.J. & Pool, J.E. (2018) Directional selection reduces developmental canalization against genetic and environmental perturbations in *Drosophila* wings. *Evolution* **72**, 1708–1715.

Hall, K.S.S., Ryttman, H., Fransson, T. & Stolt, B.-O. (2004) Stabilising selection on wing length in reed warblers *Acrocephalus scirpaceus*. *Journal of Avian Biology* **35**, 7–12.

Hayden, E.J., Bratulic, S., Koenig, I., Ferrada, E. & Wagner, A. (2014) The Effects of Stabilizing and Directional Selection on Phenotypic and Genotypic Variation in a Population of RNA Enzymes. *Journal of Molecular Evolution* **78**, 101–108.

James, F.C. (1970) Geographic Size Variation in Birds and Its Relationship to Climate. *Ecology* **51**, 365–390.

Jensen, H., Sæther, B. -E., Ringsby, T.H., Tufto, J., Griffith, S.C. & Ellegren, H. (2003) Sexual variation in heritability and genetic correlations of morphological traits in house sparrows (*Passer domesticus*). *Journal of Evolutionary Biology* **16**, 1296–1307.

Kleven, O., Laskemoen, T., Fossøy, F., Robertson, R.J. & Lifjeld, J.T. (2008) Intraspecific variation in sperm length is negatively related to sperm competition in passerine birds. *Evolution* **62**, 494–499.

Lifjeld, J.T., Laskemoen, T., Kleven, O., Albrecht, T. & Robertson, R.J. (2010) Sperm Length Variation as a Predictor of Extrapair Paternity in Passerine Birds. *PLOS ONE* **5**, e13456.

Michonneau, F., Brown, J.W. & Winter, D.J. (2016) rotl: an R package to interact with the Open Tree of Life data. *Methods in Ecology and Evolution* **7**, 1476–1481.

Mitchell-Olds, T., Willis, J.H. & Goldstein, D.B. (2007) Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics* **8**, 845–856.

Moran, N.P., Sánchez-Tójar, A., Schielzeth, H. & Reinhold, K. (2021) Poor nutritional condition promotes high-risk behaviours: a systematic review and meta-analysis. *Biological Reviews* **96**, 269–288.

Mossman, J.A., Pearson, J.T., Moore, H.D. & Pacey, A.A. (2013) Variation in mean human sperm length is linked with semen characteristics. *Human Reproduction* **28**, 22–32.

Nakagawa, S., Johnson, P.C.D. & Schielzeth, H. (2017) The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface* **14**, 20170213.

Nakagawa, S., Lagisz, M., O'Dea, R.E., Pottier, P., Rutkowska, J., Senior, A.M., Yang, Y. & Noble, D.W.A. (2023) orchaRd 2.0: An R package for visualising meta-analyses with orchard plots. *Methods in Ecology and Evolution* **14**, 2003–2010.

Nakagawa, S., Poulin, R., Mengersen, K., Reinhold, K., Engqvist, L., Lagisz, M. & Senior, A.M. (2015) Meta-analysis of variation: ecological and evolutionary applications and beyond. *Methods in Ecology and Evolution* **6**, 143–152.

Noble, D.W.A., Lagisz, M., O'dea, R.E. & Nakagawa, S. (2017) Nonindependence and sensitivity analyses in ecological and evolutionary meta-analyses. *Molecular Ecology* **26**, 2410–2425.

Noble, D.W.A., Senior, A.M., Uller, T. & Schwanz, L.E. (2021) Heightened among-individual variation in life history but not morphology is related to developmental temperature in reptiles. *Journal of Evolutionary Biology* **34**, 1793–1802.

Nokelainen, O., Silvasti, S.A., Strauss, S.Y., Wahlberg, N. & Mappes, J. (2024) Predator selection on phenotypic variability of cryptic and aposematic moths. *Nature Communications* **15**, 1678.

Paradis, E. & Schliep, K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528.

Pélabon, C., Hansen, T.F., Carter, A.J.R. & Houle, D. (2010) Evolution of Variation and Variability Under Fluctuating, Stabilizing, and Disruptive Selection. *Evolution* **64**, 1912–1925.

Price, G.R. (1972) Fisher's 'fundamental theorem' made clear. *Annals of Human Genetics* **36**, 129–140.

Prout, T. (1962) The effects of stabilizing selection on the time of development in *Drosophila melanogaster*. *Genetical Research* **3**, 364–382.

R Core Team (2023) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reinhold, K. (2011) Variation in acoustic signalling traits exhibits footprints of sexual selection. *Evolution* **65**, 738–745.

Reinhold, K. & Engqvist, L. (2013) The variability is in the sex chromosomes. *Evolution* **67**, 3662–3668.

Richardson, J., Lees, A.C., Miller, E.T. & Marsden, S.J. (2023) Avian diversity and function across the world's most populous cities. *Ecology Letters* **26**, 1301–1313.

Sánchez-Tójar, A., Moran, N.P., O’Dea, R.E., Reinhold, K. & Nakagawa, S. (2020) Illustrating the importance of meta-analysing variances alongside means in ecology and evolution. *Journal of Evolutionary Biology* **33**, 1216–1223.

Santostefano, F., Moiron, M., Sánchez-Tójar, A. & Fisher, D.N. (2025) Indirect genetic effects increase the heritable variation available to selection and are largest for behaviors: a meta-analysis. *Evolution Letters* **9**, qrae051.

Selander, R.K. (ed) (2017) Sexual Selection and Dimorphism in Birds. In *Sexual Selection and the Descent of Man* p. Routledge.

Sharma, M.D., Minder, A.M. & Hosken, D.J. (2013) No association between sperm competition and sperm length variation across dung flies (Scathophagidae). *Journal of Evolutionary Biology* **26**, 2341–2349.

Stearns, S.C., Kaiser, M. & Kawecki, T.J. (1995) The differential genetic and environmental canalization of fitness components in *Drosophila melanogaster*. *Journal of Evolutionary Biology* **8**, 539–557.

Subasinghe, K., Symonds, M.R.E., Vidal-García, M., Bonnet, T., Prober, S.M., Williams, K.J. & Gardner, J.L. (2021) Repeatability and Validity of Phenotypic Trait Measurements in Birds. *Evolutionary Biology* **48**, 100–114.

Swaddle, J.P., Karubian, J. & Pruett-Jones, S. (2000) A novel evolutionary pattern of reversed sexual dimorphism in fairy wrens: implications for sexual selection. *Behavioral Ecology* **11**, 345–349.

Tarka, M., Åkesson, M., Hasselquist, D. & Hansson, B. (2014) Intralocus Sexual Conflict over Wing Length in a Wild Migratory Bird. *The American Naturalist* **183**, 62–73.

Viechtbauer, W. (2010) Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* **36**, 1–48.

Willmore, K.E., Young, N.M. & Richtsmeier, J.T. (2007) Phenotypic Variability: Its Components, Measurement and Underlying Developmental Processes. *Evolutionary Biology* **34**, 99–120.

Yang, Y., Lagisz, M. & Nakagawa, S. (2023) Enriching meta-analyses through scoping review, bibliometrics, and alternative impact metrics: Visualizing study characteristics, hidden risk of bias, societal influence, and research translation. arXiv. http://arxiv.org/abs/2312.01643 [accessed 18 November 2024].

Yang, Y., Noble, D.W.A., Spake, R., Senior, A.M., Lagisz, M. & Nakagawa, S. (2024) A pluralistic framework for measuring and stratifying heterogeneity in meta-analyses. EcoEvoRxiv. https://ecoevorxiv.org/repository/view/6299/ [accessed 18 November 2024].

Yom-Tov, Y., Yom-Tov, S., Wright, J., J. R. Thorne, C. & Du Feu, R. (2006) Recent changes in body weight and wing length among some British passerine birds. *Oikos* **112**, 91–101.

Young, N.M. (2006) Function, ontogeny and canalization of shape variance in the primate scapula. *Journal of Anatomy* **209**, 623–636.

Yu, G., Smith, D.K., Zhu, H., Guan, Y. & Lam, T.T.-Y. (2017) ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8**, 28–36.

Supplementary material

S1. Male-Female wing length correlations

Overall and as expected, mean wing length was strongly correlated between males and females both at the original (Pearson’s r = 0.995, 95% CI = 0.993 to 0.996, df = 303, n = 305 estimates; Figure S1) and log-transformed scale (Pearson’s r = 0.997, 95% CI = 0.997 to 0.998, df = 303, n = 305 estimates; Figure S1), with the Great Bustard (*Otis tarda*) showing the strongest wing length size sexual dimorphism – male Great Bustards showed ca. 27% longer wings than females (males: mean = 566.9 mm, SD = 14.7, k = 50 individuals; females: mean = 444.7 mm, SD = 14.7, k = 19 individuals; source: Alonso et al. 2009). Note that none of those two correlations account for sources of nonindependence such as those accounted for in our models (i.e., study ID, pair ID, population ID, phylogeny, species ID, within-study variance).

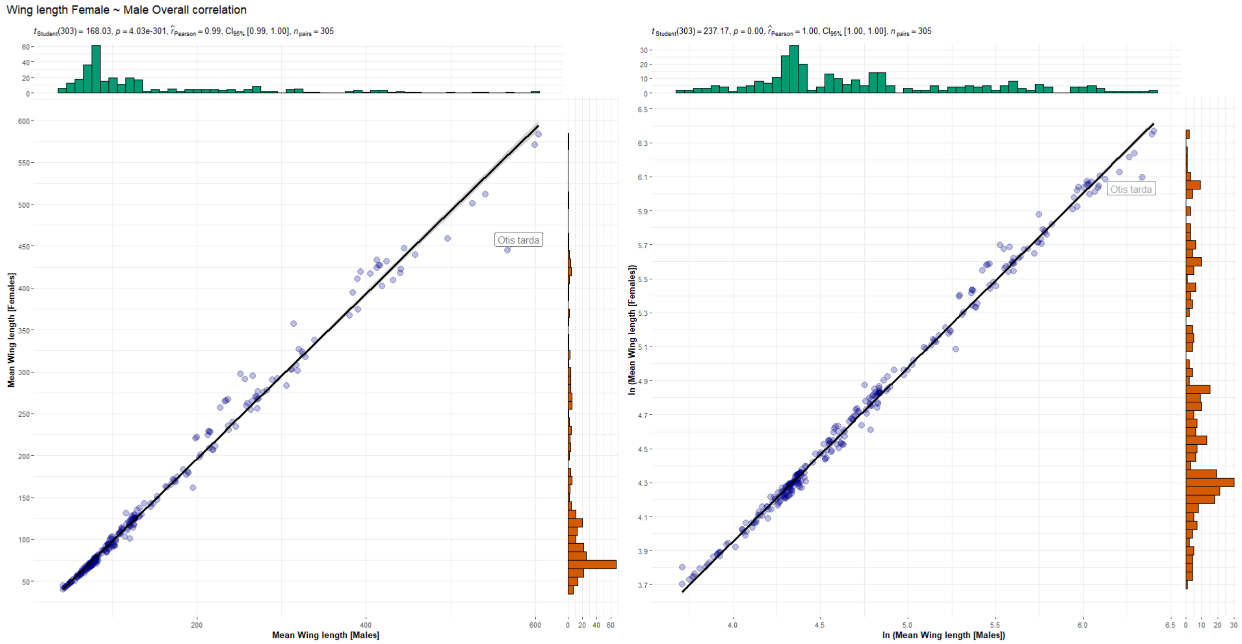

Figure S1. Mean wing length is strongly correlated between males and females at the original (left panel) and log scale (right panel). Circles represent mean values, with wing length originally expressed in mm.

S2. Mean-Variance relationship and CV summary statistics

Our dataset shows evidence for a mean-variance relationship (i.e., Taylor's Law; Cohen & Xu, 2015) both in males (Pearson's *r* = 0.894, 95% CI = 0.870 to 0.913, df = 338, n = 340 estimates; Figure S2A) and females (Pearson's *r* = 0.873, 95% CI = 0.844 to 0.897, df = 308, n = 310 estimates; Figure S2B). Note that none of those two correlations account for sources of nonindependence such as those accounted for in our models (i.e., study ID, pair ID, population ID, phylogeny, species ID, within-study variance).

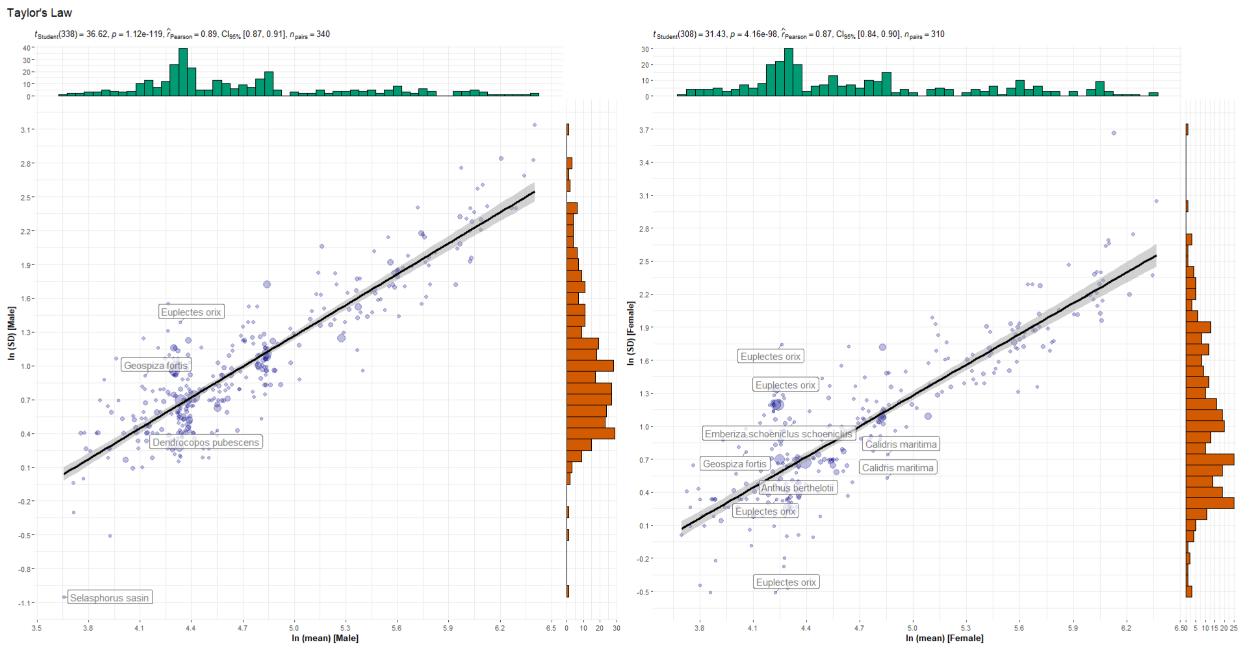

Figure S2. The regression lines show evidence of a mean-variance relationship in males (left panel:) and females. Circles represent log-transformed mean and SD values, with dot size being scaled by sample size. Wing length was originally expressed in mm. Labelled data points correspond with estimates based on less than 4 individuals.

Overall, after back-transforming them from lnCV, wing length *CV* values ranged from 1% to 13% across our dataset (median = 2.42%, mean = 2.67%, SD = 1.01) with males showing only slightly smaller values (median = 2.40, range: 1.28-6.81%; mean = 2.62, SD = 0.84; n = 340 estimates) than females (median = 2.42, range: 0.97-13.24%; mean = 2.73, SD = 1.17; n = 310 estimates; Figure S3). Note that, contrary to the estimates presented in the main manuscript, these do not account for the sources of nonindependence identified and accounted for in all our meta-analytic models (i.e., study ID, pair ID, population ID, phylogeny, species ID, within-study variance).

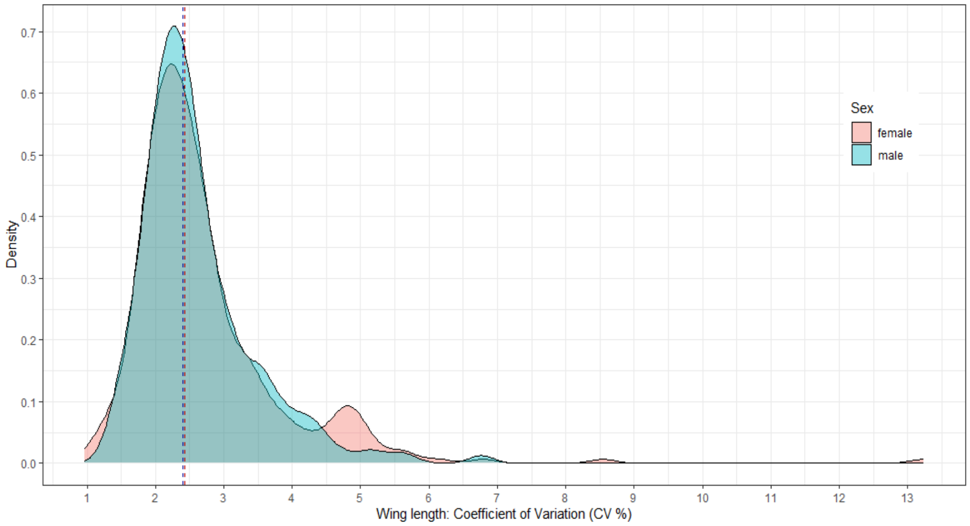

Figure S3. Density plot showing the distribution of wing length coefficient of variation (*CV*) values (%) across bird species. Dashed lines show median *CV* values for males (blue; 172 species, n = 340 estimates) and females (red; 163 species, n = 310 estimates).

S3. Supplementary tables and figures

Table S1. Heterogeneity metrics for a phylogenetic multilevel meta-analysis on the coefficient of variation of wing length values across 172 bird species (I^2: source of heterogeneity, *CVH2*: mean-standardized heterogeneity, *M2*: variance-mean-standardized heterogeneity, σ^2: total heterogeneity).

	I^2 (%)	***CVH2***	***M2***	σ^2
Total	94.2	0.005	0.005	0.068
Between-study (Study ID)	25.1	0.001	0.001	0.018
Paired estimates (Pair ID)	1.9	0	0	0.001
Population (Population ID)	14.4	0.001	0.001	0.010
Phylogeny	23.0	0.001	0.001	0.017
Species	8.3	0	0	0.006
Within-study (unit-level observation ID)	21.3	0.001	0.001	0.015

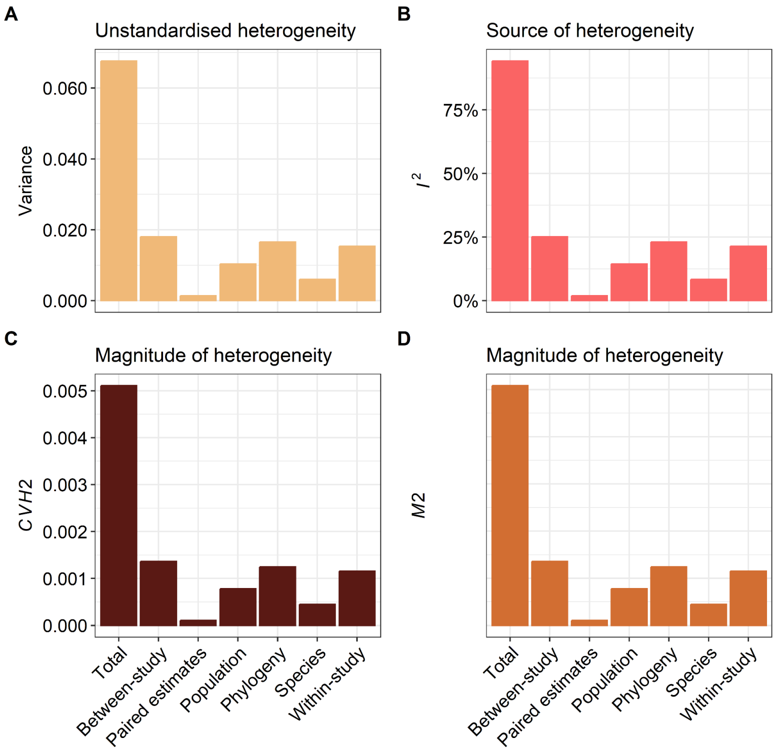

Figure S4. Heterogeneity quantification and stratification for a phylogenetic multilevel intercept-only meta-analysis on the coefficient of variation of wing length across 172 bird species. Heterogeneity was quantified using for metrics: raw variance (A), source measure I^2 (B), mean-standardized magnitude measure *CVH2* (C), and variance-mean-standardized magnitude measure *M2* (D), and stratified at all the random effect levels included in the meta-analysis.

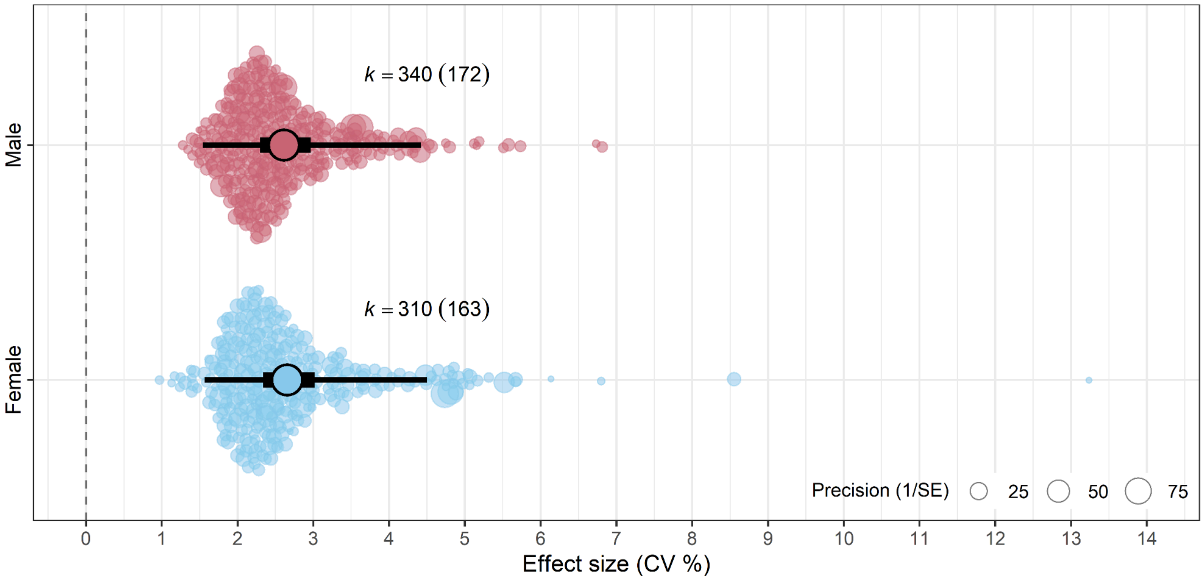

Figure S5. Males and females do not statistically differ in the coefficient of variation of wing length across 172 bird species. Orchard plots show the mean estimate, 95% confidence intervals (thick whisker) and 95% prediction intervals (thin whisker), with dot size being scaled by effect size precision (i.e., 1/SE). k corresponds to the number of estimates, with the number of species shown in parentheses.

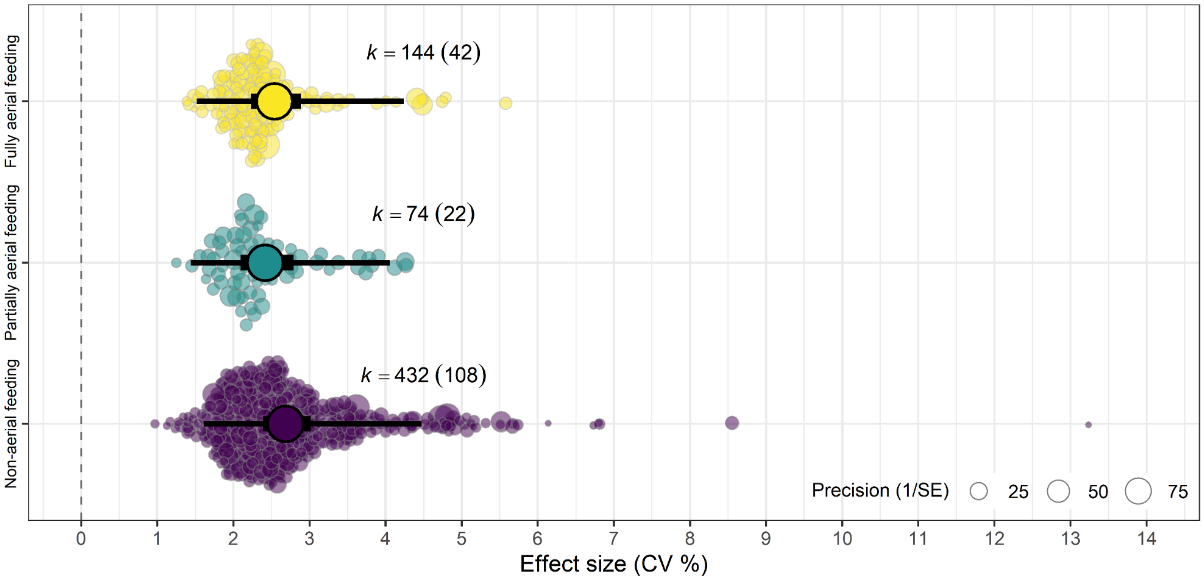

Figure S6. Bird species showing aerial or partially aerial feeding tend to show higher canalisation (i.e., smaller values of coefficient of variation, *CV*) on wing length, but those differences were only statistically significant between non-aerial and partially aerial feeding. Orchard plots show the mean estimate, 95% confidence intervals (thick whisker) and 95% prediction intervals (thin whisker), with dot size being scaled by effect size precision (i.e., 1/SE). k corresponds to the number of estimates, with the number of species shown in parentheses.

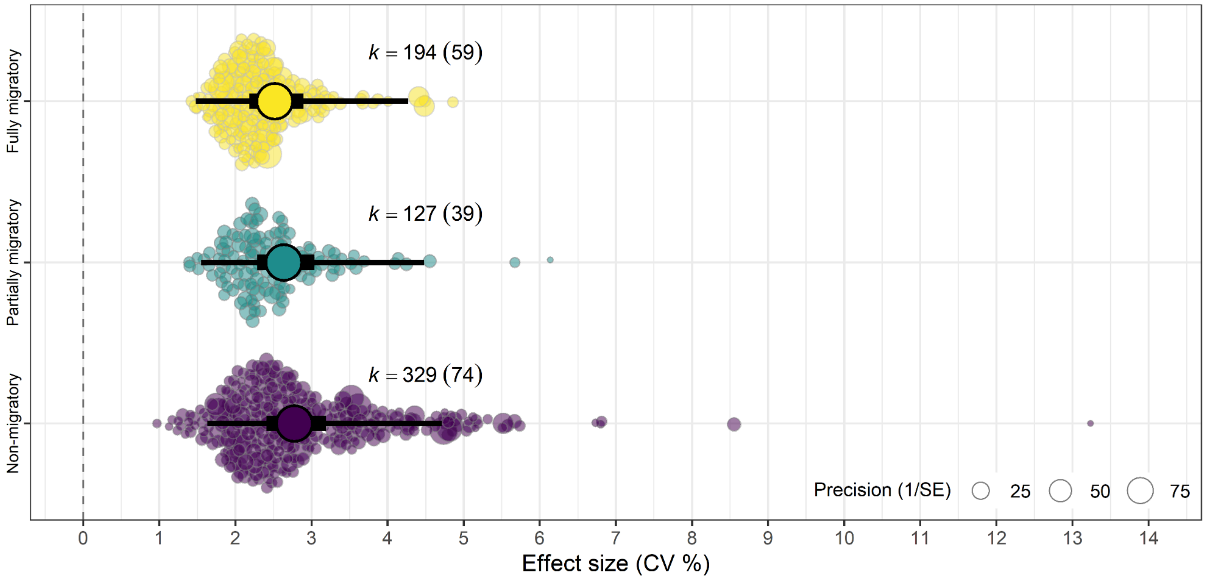

Figure S7. Bird species showing full or partial migration tend to show higher canalisation (i.e., smaller values of coefficient of variation, *CV*) on wing length, but those differences were only statistically significant between non-migratory and fully migratory. Orchard plots show the mean estimate, 95% CI (thick whisker) and 95% PI (thin whisker), with dot size being scaled by effect size precision (i.e., 1/SE). k corresponds to the number of estimates, with the number of species shown in parentheses.

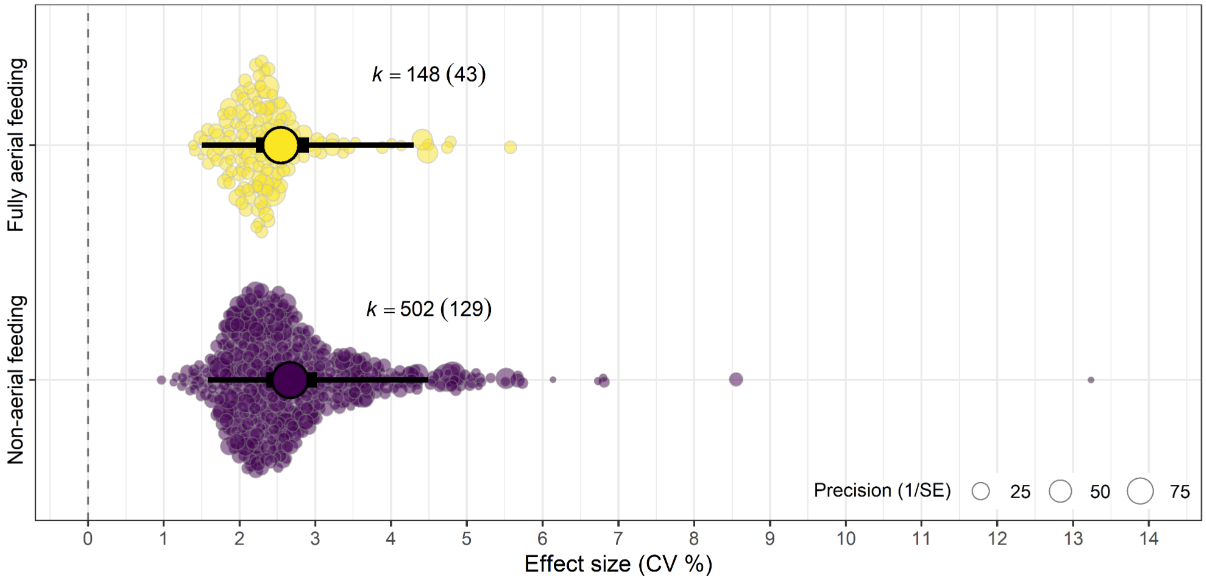

Figure S8. Although bird species populations showing aerial feeding seemingly show higher canalisation (i.e., smaller values of coefficient of variation, *CV*) on wing length than non-aerial ones, that difference was not statistically significant. Orchard plots show the mean estimate, 95% confidence intervals (thick whisker) and 95% prediction intervals (thin whisker), with dot size being scaled by effect size precision (i.e., 1/SE). k corresponds to the number of estimates, with the number of species shown in parentheses.

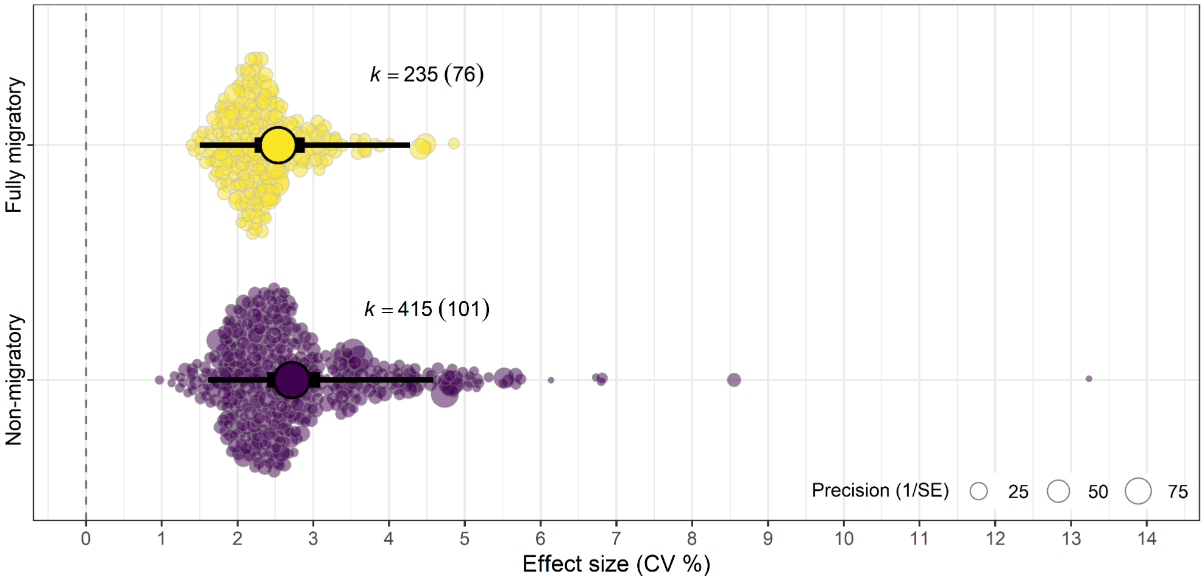

Figure S9. Bird species populations showing full migration show higher canalisation (i.e., smaller values of coefficient of variation, *CV*) on wing length than non-migratory ones. Orchard plots show the mean estimate, 95% CI (thick whisker) and 95% PI (thin whisker), with dot size being scaled by effect size precision (i.e., 1/SE). k corresponds to the number of estimates, with the number of species shown in parentheses.

S4. Calculating total heterogeneity explained

Table S2. Results from a phylogenetic multilevel meta-regression that included both feeding and migratory type as moderators categorised at the species level. Estimates are shown as *lnCV* (k = 650 estimates, N = 181 studies, 172 species).

Estimate	**Mean**	**Lower and upper 95% CI**	**t-value**	**p-value**
Intercept	-3.577	[-3.709,-3.445]	-53.21	<0.001
Partially migratory	-0.050	[-0.140,0.040]	-1.09	0.277
Fully migratory	-0.089	[-0.173,-0.005]	-2.08	0.038
Partially aerial feeding	-0.093	[-0.196,0.009]	-1.78	0.075
Fully aerial feeding	-0.027	[-0.124,0.071]	-0.54	0.588

Table S3. Results from a phylogenetic multilevel meta-regression that included both feeding and migratory type as moderators categorised at the population level. Estimates are shown as *lnCV* (k = 650 estimates, N = 181 studies, 172 species).

Estimate	**Mean**	**Lower and upper**	**t-value**	**p-value**

		95% CI		
Intercept	-3.601	[-3.731,-3.470]	-54.33	<0.001
Fully migratory	-0.065	[-0.133,0.004]	-1.86	0.064
Fully aerial feeding	-0.024	[-0.118,0.069]	-0.51	0.607

S5. Sensitivity analyses: Treating vagrants as migrants

Species-specific level

We reran the uni-moderator phylogenetic multilevel meta-regressions testing whether migratory type explained among-species differences in wing length canalisation after classifying vagrants as fully migratory instead of non-migratory. This model showed that non-migratory behaviour was associated with larger wing length CV values (mean *CV* = 2.82%, 95% CI = 2.46 to 3.23) than those found in partially migratory (mean *CV* = 2.64%, 95% CI = 2.29 to 3.03) and fully migratory species (mean *CV* = 2.49%, 95% CI = 2.17 to 2.85), with the difference being statistically significant for fully migratory species (p-value = 0.004) but not for partially migratory species (p-value = 0.147; Figure S10). The percentage of variance explained by migratory type alone was higher than when not considering vagrants as fully migratory ($R^2_{marginal}$ = 4.38%), and both moderators combined (i.e., feeding and migration type) explained up to 5.58% of the heterogeneity (Table S3).

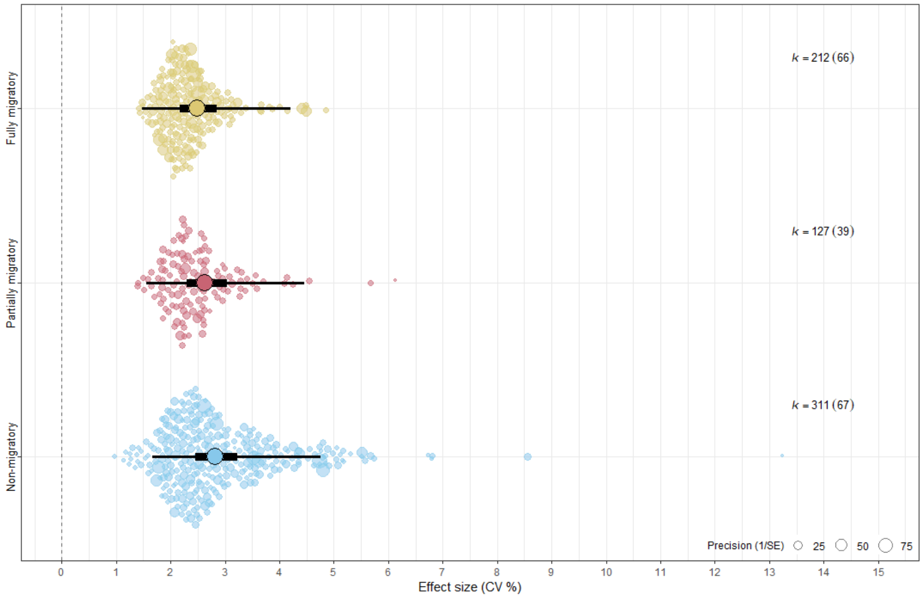

Figure S10. When considering vagrants as fully migratory, bird species showing full or partial migration tend to show higher canalisation (i.e., smaller values of coefficient of variation, *CV*) on wing length, but those differences were only statistically significant between non-migratory and fully migratory. Orchard plots show the mean estimate, 95% CI (thick whisker) and 95% PI (thin whisker), with dot size being scaled by effect size precision (i.e., 1/SE). k corresponds to the number of estimates, with the number of species shown in parentheses.

Table S4. Results from a phylogenetic multilevel meta-regression that considered vagrants as fully migratory and included both feeding and migratory type as moderators categorised at the species level. Estimates are shown as lnCV (k = 650 estimates, N = 181 studies, 172 species).

Estimate	**Mean**	**Lower and upper 95% CI**	**t-value**	**p-value**
Intercept	-3.563	[-3.695,-3.430]	-52.74	<0.001
Partially migratory	-0.070	[-0.162,0.022]	-1.49	0.138
Fully migratory	-0.122	[-0.212,-0.032]	-2.66	0.008
Partially aerial feeding	-0.090	[-0.192,0.012]	-1.73	0.084
Fully aerial feeding	-0.005	[-0.105,0.095]	-0.10	0.922

Population-specific level

We reran the uni-moderator phylogenetic multilevel meta-regressions testing whether migratory type categorised at the population level explained among-species differences in wing length canalisation after classifying vagrants as fully migratory instead of non-migratory. This model showed that non-migratory behaviour was associated with statistically significantly larger wing length *CV* values (mean *CV* = 2.74%, 95% CI = 2.42 to 3.11) than those found for fully migratory populations (mean *CV* = 2.52%, 95% CI = 2.22 to 2.85; p-value = 0.014; Figure S11). The percentage of variance explained by migratory type alone was again higher than when not considering vagrants as fully migratory (R^2_{marginal} = 2.61%), and both moderators combined (i.e., feeding and migration type) explained up to 2.72% of the heterogeneity (Table S4).

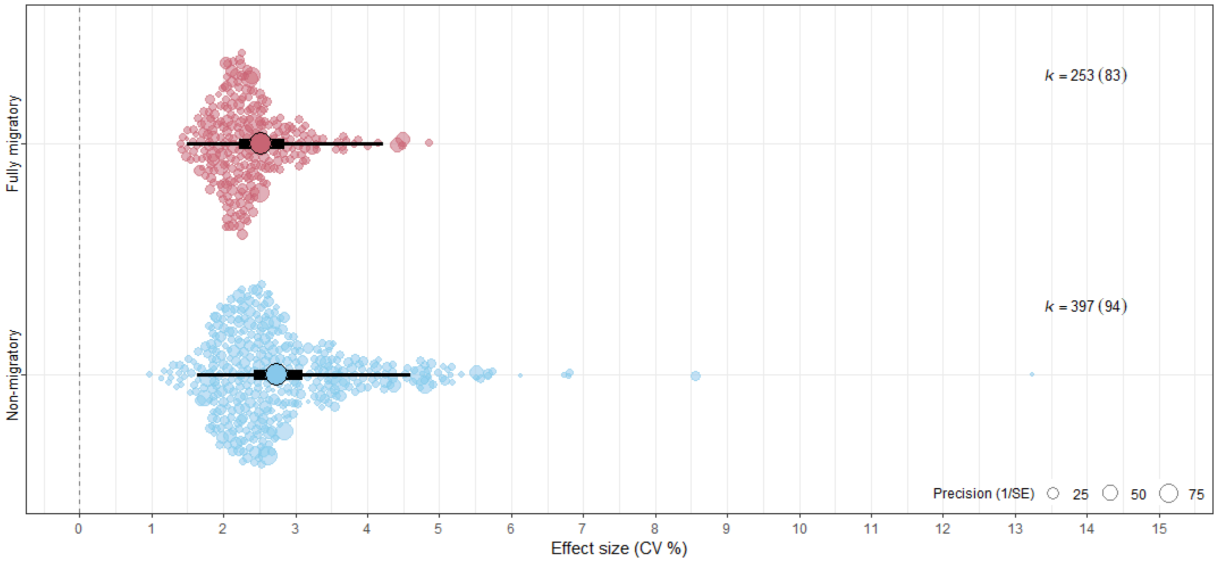

Figure S11. When considering vagrants as fully migratory, bird species populations showing full migration show higher canalisation (i.e., smaller values of coefficient of variation, *CV*) on wing length than non-migratory ones.

Orchard plots show the mean estimate, 95% CI (thick whisker) and 95% PI (thin whisker), with dot size being scaled by effect size precision (i.e., 1/SE). k corresponds to the number of estimates, with the number of species shown in parentheses.

Table S5. Results from a phylogenetic multilevel meta-regression that considered vagrants as fully migratory and included both feeding and migratory type as moderators categorised at the population level. Estimates are shown as lnCV (k = 650 estimates, N = 181 studies, 172 species).

Estimate	**Mean**	**Lower and upper 95% CI**	**t-value**	**p-value**
Intercept	-3.595	[-3.723,-3.466]	-54.94	<0.001
Fully migratory	-0.083	[-0.155,-0.011]	-2.25	0.025
Fully aerial feeding	-0.011	[-0.106,0.084]	-0.23	0.820

S6. Re-analysis of Nakagawa's et al. 2015 Example 2 dataset

The final dataset contained 250 male and 250 female estimates of wing length (k = 266), body mass (k = 172) and size (head: k = 9, beak: k = 6, body: k = 4) obtained from 86 studies and 108 bird species. The phylogenetic multilevel meta-analysis showed that the mean *CV* value across species was 4.35% (95% CI = 3.44 to 5.51, 95% PI = 1.28 to 14.80) with most differences among effect sizes coming from sources other than sampling error (I^2_{total} = 98.6%) despite the magnitude of the heterogeneity being very small ($CVH2_{total}$ = 0.038; $M2_{total}$ = 0.037; Table S6, Figure S12). Such an apparent disagreement between I^2_{total} and both $CVH2_{total}$ = and $M2_{total}$ was due to the small meta-analytic mean and typical sampling variance ($\overline{\nu}$ = 0.005) in our data.

The phylogenetic multilevel meta-regression showed that the mean *CV* value across traits differed statistically (p-value = 0.001) between males (mean *CV* = 4.25%, 95% CI = 3.35 to 5.38) and females (mean *CV* = 4.47%, 95% CI = 3.53 to 5.67), but the variance explained by the moderator sex was negligible (R^2_{marginal} = 0.18%; Figure S13). That difference corresponds to, on average, 4.9% lower *CV* values for males compared to females.

Table S6. Heterogeneity metrics for a phylogenetic multilevel meta-analysis on the coefficient of variation of wing length values across 108 bird species (I^2: source of heterogeneity, *CVH2*: mean-standardized heterogeneity, *M2*: variance-mean-standardized heterogeneity, σ^2: total heterogeneity). The typical sampling variance found in our dataset was $\overline{\nu}$ = 0.005.

	I^2 (%)	***CVH2***	***M2***	σ^2
Total	98.6	0.038	0.037	0.374
Between-study (Study ID)	17.2	0.007	0.006	0.065

Paired estimates (Pair ID)	62.2	0.024	0.023	0.236
Population (Population ID)	NA	NA	NA	NA
Phylogeny	0	0	0	0
Species	11.8	0.005	0.004	0.045
Within-study (unit-level observation ID)	7.3	0.003	0.003	0.028

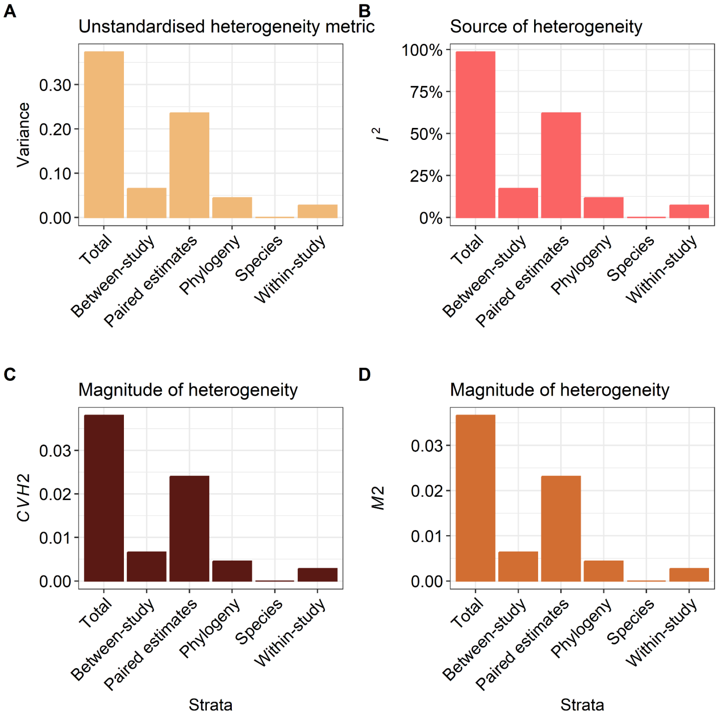

Figure S12. Quantification and stratification for four heterogeneity metrics for a phylogenetic intercept-only multilevel meta-analysis on the coefficient of variation of general body size values across 108 bird species. Heterogeneity was quantified using raw variance (A), source measure I^2 (B), mean-standardized magnitude measure *CVH2* (C), and variance-mean-standardized magnitude measure *M2* (D), and stratified at all the random effect levels.

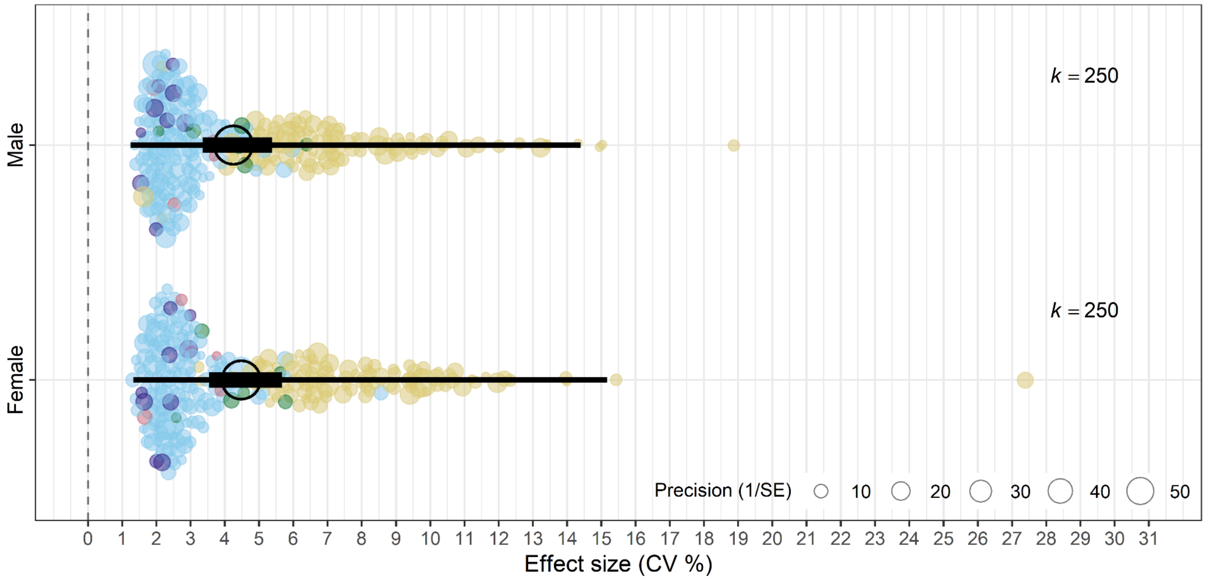

Figure S13. Males and females differ statistically in the coefficient of variation of size across 108 bird species. Orchard plots show the mean estimate, 95% confidence intervals (thick whisker) and 95% prediction intervals (thin whisker). Dot size is scaled by effect size precision (i.e., 1/SE) and dots are coloured by the type of trait: wing length (blue), body mass (wing) and size (head: purple, beak: green, body: red) . k corresponds to the number of estimates.

When running the same analysis but for the subset of wing length estimates (i.e., blue dots in Figure S13), the final dataset contained 133 male estimates and 133 female estimates obtained from 61 studies and 90 bird species. The phylogenetic multilevel meta-analysis showed that the mean *CV* value across species was 2.64% (95% CI = 2.20 to 3.18, 95% prediction intervals = 1.38 to 5.07). As for the full dataset, most differences among effect sizes coming from sources other than sampling error (I^2_{total} = 94.9%) despite the magnitude of the heterogeneity being very small ($CVH2_{total}$ = 0.008; $M2_{total}$ = 0.008; Table S7, Figure S14). Such an apparent disagreement between I^2_{total} and both $CVH2_{total}$ = and M2_{total} was due to the small meta-analytic mean and typical sampling variance ($\overline{\nu}$ = 0.005) in our data.

The phylogenetic multilevel meta-regression showed that the mean *CV* value did not differ statistically (p-value = 0.4125) between males (mean *CV* = 2.63%, 95% CI = 2.18 to 3.16) and females (mean *CV* = 2.67%, 95% CI = 2.22 to 3.21) and, indeed the variance explained by the moderator sex was negligible (R^2_{marginal} = 0.06%; Figure S15). That difference corresponds to, on average, 1.5% lower *CV* values for males compared to females.

Table S7. Heterogeneity metrics for a phylogenetic multilevel meta-analysis on the coefficient of variation of wing length values across 90 bird species (I^2: source of heterogeneity, *CVH2*: mean-standardized heterogeneity, *M2*: variance-mean-standardized heterogeneity, σ^2: total heterogeneity). The typical sampling variance found in our dataset was $\overline{\nu}$ = 0.005.

	I^2 (%)	***CVH2* (%)**	***M2***	**σ^2**
Total	94.9	0.008	0.008	0.100

Between-study (Study ID)	29.7	0.002	0.002	0.031
Paired estimates (Pair ID)	15.2	0.001	0.001	0.016
Population (Population ID)	NA	NA	NA	NA
Phylogeny	6.2	0.000	0.000	0.007
Species	26.6	0.002	0.002	0.028
Within-study (unit-level observation ID)	17.1	0.001	0.001	0.018

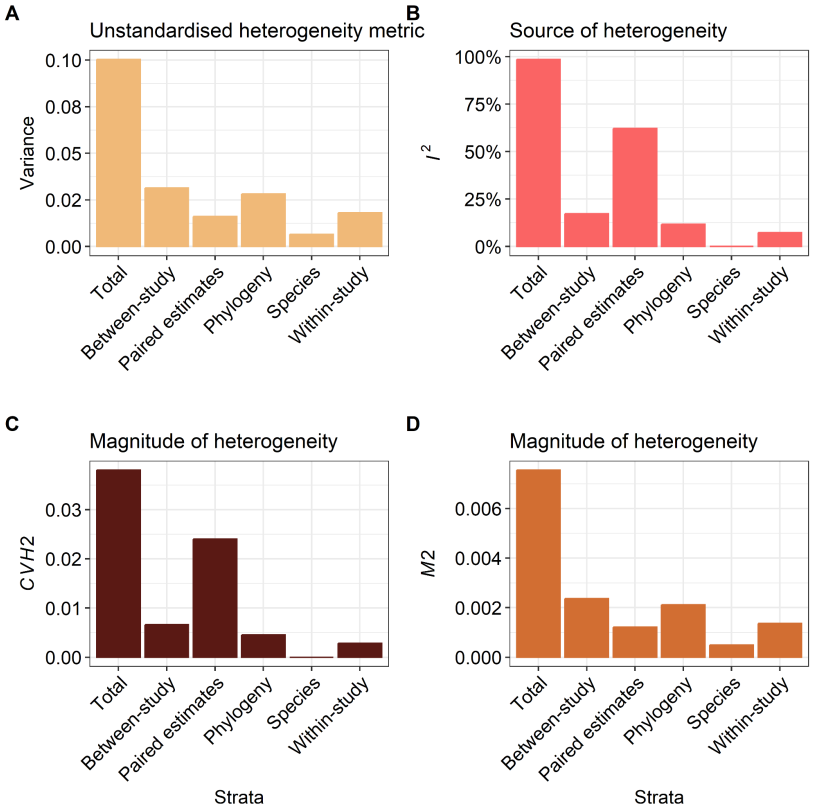

Figure S14. Quantification and stratification for four heterogeneity metrics for a phylogenetic intercept-only multilevel meta-analysis on the coefficient of variation of wing length values across 90 bird species. Heterogeneity was quantified using raw variance (A), source measure I^2 (B), mean-standardized magnitude measure *CVH2* (C), and variance-mean-standardized magnitude measure *M2* (D), and stratified at all the random effect levels.

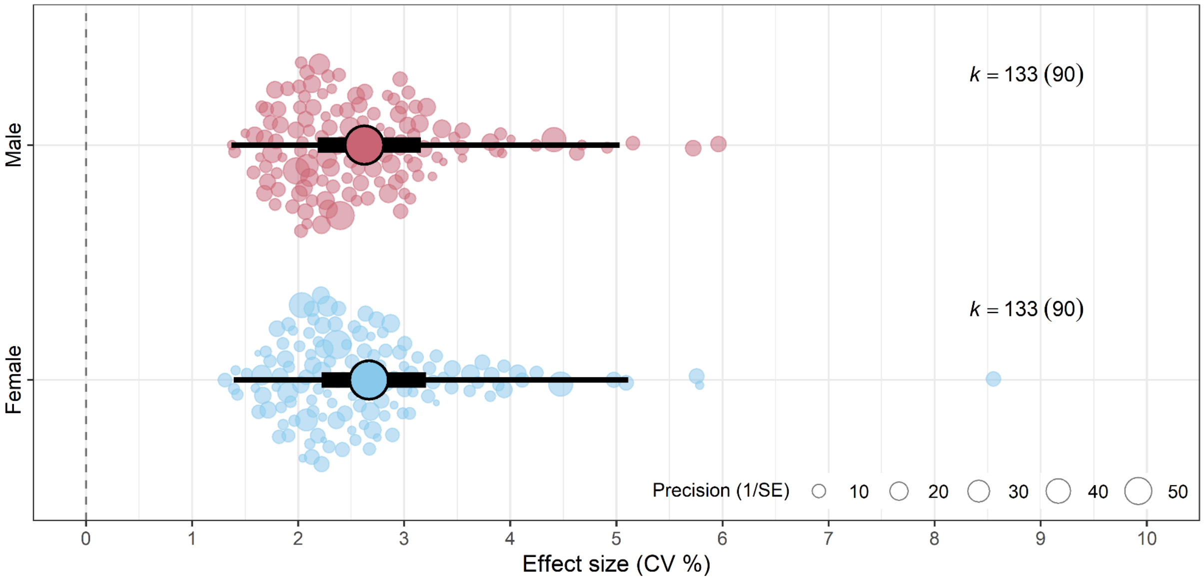

Figure S15. Males and females do not statistically differ in the coefficient of variation of wing length across 90 bird species. Orchard plots show the mean estimate, 95% confidence intervals (thick whisker) and 95% prediction intervals (thin whisker), with dot size being scaled by effect size precision (i.e., 1/SE). k corresponds to the number of estimates, with the number of species shown in parentheses.

References (Supplementary Material)

Cohen, J.E. & Xu, M. (2015) Random sampling of skewed distributions implies Taylor's power law of fluctuation scaling. *Proceedings of the National Academy of Sciences* **112**, 7749–7754.

Nakagawa, S., Poulin, R., Mengersen, K., Reinhold, K., Engqvist, L., Lagisz, M. & Senior, A.M. (2015) Meta-analysis of variation: ecological and evolutionary applications and beyond. *Methods in Ecology and Evolution* **6**, 143–152.