

# 1 **Wing length canalisation and behaviour across birds: a phylogenetic meta-analysis of variance**

2

3 **Authors:** Klaus Reinhold, Alfredo Sánchez-Tójar

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

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

6

## 7 **Abstract**

8 We examined the expectation that stronger stabilising selection leads to a decrease in trait variation  
9 across species by investigating individual variation in wing length. We hypothesised that species that  
10 heavily rely on aerial feeding, as well as long-distance migratory species, show higher canalisation  
11 (lower coefficients of variation, i.e., CV values) in wing length than non-aerial feeders and non-  
12 migratory species. We collected summary statistics on wing length for males and females from the  
13 literature and analysed them using recently developed meta-analytic metrics for integrating  
14 phenotypic variance estimates. Our phylogenetic multilevel meta-analysis showed relatively low  
15 heterogeneity among CV values, indicating generalisability of the overall CV value (2.6%). Although  
16 not all pairwise comparisons were statistically significant, all our analyses consistently showed higher  
17 canalisation in aerial compared to non-aerial feeders, and in migratory compared to non-migratory  
18 species. We conclude that wing length in bird species relying on their wings more heavily is likely  
19 under stronger (stabilising) selection, which in turn would have led to the observed higher  
20 canalisation on this trait for those species. Our study showcases how to combine already available  
21 descriptive statistics for phenotypic traits with underused meta-analysis of variance approaches to  
22 test often-neglected evolutionary predictions.

23 **Keywords:** meta-analysis of variance; evidence synthesis; avian; comparative analysis; coefficient of  
24 variation; canalisation; replicability; generalizability; life-history evolution

25

## 26 **Introduction**

27 Much progress has been made in understanding phenotypic traits in recent years; however, most  
28 research has focused on understanding mean phenotypic differences (e.g., Fattorini et al. 2023,  
29 Richardson et al. 2023, Nokelainen et al. 2024) and less so on understanding differences in  
30 phenotypic variance (but see Kleven et al. 2008, Lifjeld et al. 2010, Sánchez-Tójar et al. 2020, Moran  
31 et al. 2021). This scarcity of research on phenotypic variance in evolutionary biology is surprising  
32 given the key role that it plays in evolution and its importance in the study of selection, evolvability

33 and canalisation (Mitchell-Olds et al. 2007, Willmore et al. 2007, Geiler-Samerotte et al. 2013, Bolnick  
34 et al. 2011, see also Des Roches et al. 2018, Thompson et al. 2022, for research on the topic from an  
35 ecological view). Without phenotypic variation, selection and adaptation cannot take place. The  
36 amount of phenotypic variation associated with additive genetic differences will determine the speed  
37 of evolutionary change, together with the strength of selection, regardless of whether directional or  
38 stabilising selection is at work (Price 1972). Nonetheless, the amount of variance in a quantitative  
39 trait depends on both the type of selection occurring and the effects of the environment on the trait  
40 (i.e., phenotypic plasticity). First, both stabilising and directional selection are expected to decrease  
41 additive genetic variance, whereas disruptive and fluctuating selection should increase it. Second,  
42 phenotypic plasticity can potentially increase trait variance if individuals experience different  
43 environments and if canalisation is imperfect. Here, we use the term canalization in line with the  
44 review from Flatt (2005) and see the coefficient of variation (CV) of a trait as an estimate for the  
45 extent of phenotypic trait variation. Similarly to phenotypic plasticity, bet-hedging strategies,  
46 independent of whether these refer to within- or among-individual developmental variance, should  
47 also lead to an increase in phenotypic variance (reviewed by Levis & Pfennig 2019). Although trait  
48 variance is expected to be mostly driven by selection and affected by phenotypic plasticity, trait  
49 variance can increase due to other processes. This includes sources such as developmental  
50 stochasticity, but also measurement error. Indeed, trait variance might be increased due to random  
51 deviations occurring during development, even when strong stabilising selection is expected to  
52 reduce this developmental load, leading to stronger canalisation (i.e., less variance) during trait  
53 development. Lastly, although measurement error might be estimated and accounted for via  
54 repeated measurement designs, it is an unavoidable source of trait variance.

55 From the processes listed above, canalisation has received comparatively little attention. For  
56 example, variation in sperm length in birds seems negatively correlated with the rate of extra-pair  
57 paternity (Kleven et al. 2008, Lifjeld et al. 2010), whereas variation in conspicuous plumage traits  
58 under strong selection is surprisingly large (Delhey et al. 2017). In insects and frogs, traits used in  
59 acoustic communication that are under strong directional selection via female choice have been  
60 shown to be less variable than acoustic traits under weaker directional selection (Reinhold 2011).  
61 Nonetheless, studies on directional selection have shown compelling evidence for both a decrease in  
62 genetic trait variance (and, thus, also in phenotype trait variance) and an increase in phenotypic  
63 variance via disrupting developmental canalisation (Groth et al. 2020; but see Hayden et al. 2014).  
64 Whereas disruptive selection has been shown to generally lead to lower canalisation (Pelabon et al.  
65 2010), several experimental studies suggest that trait canalisation increases with stabilising selection  
66 (e.g., Prout 1962), and that traits closely correlated with fitness are usually strongly canalised (Stearns  
67 et al. 1995; but see Young 2006). Overall, few studies have examined the extent of trait canalisation

68 among species and for those that have, results remain contradictory. Consequently, the importance  
69 of this variance-reduction process remains unclear.

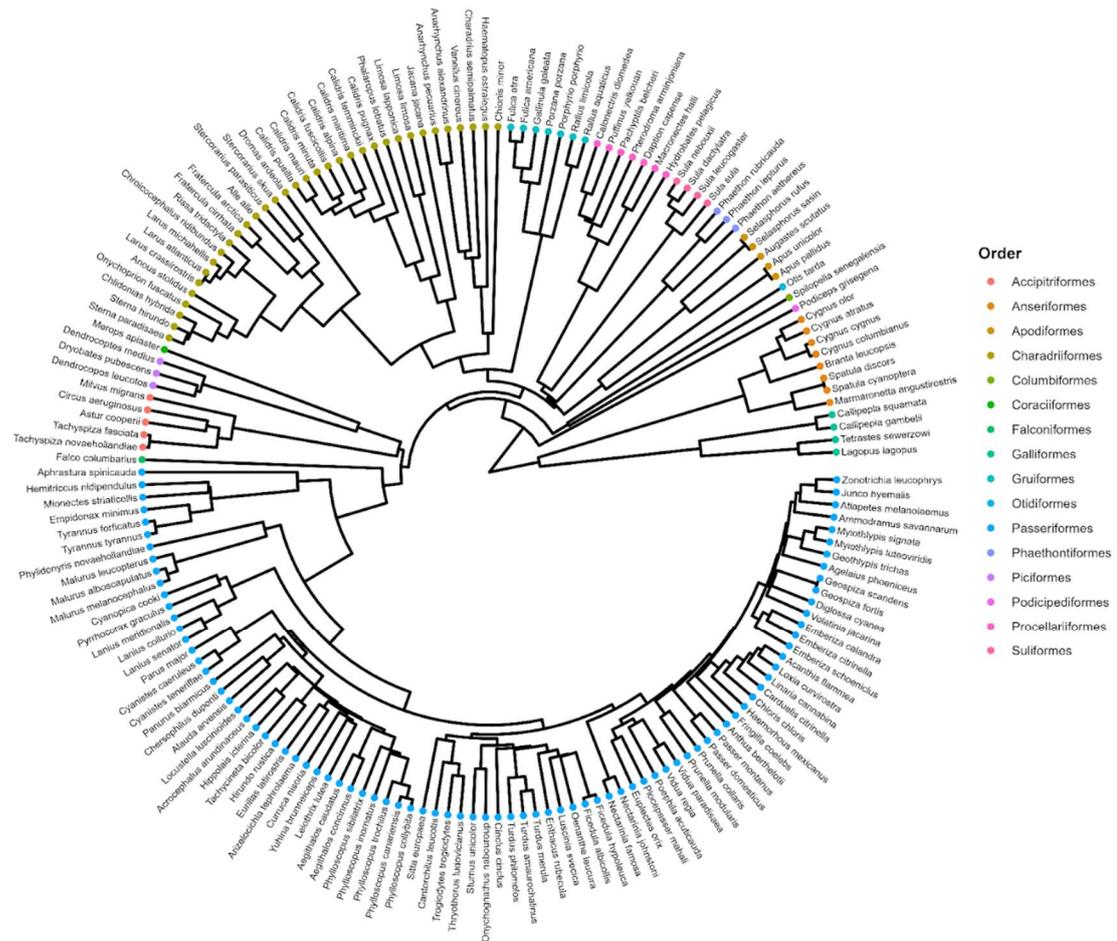
70 Here, we make use of recently developed meta-analytic metrics for integrating variance estimates  
71 across studies (Nakagawa et al. 2015, Noble et al. 2021) to study trait canalisation on an essential bird  
72 trait: wing length. There are several reasons that make bird wing length a great model trait to study  
73 canalisation. First, the heritability of bird wing length is relatively high, ranging from 0.4 to 0.6 (Cava  
74 2019, Jensen 2003, Tarka et al. 2014), and this trait has been shown to be under stabilising selection  
75 (Hall et al. 2004) as well as to respond to environmental effects (Yom-Tov et al. 2006). Second, wing  
76 length is comparatively little affected by measurement error (Subasinghe et al. 2021) and often  
77 measured and reported in articles studying questions other than canalisation (e.g., sexual  
78 dimorphism or population differentiation; Selander 2017, James 1970), which reduces the likelihood  
79 that our conclusions would be affected by publication bias. Our main hypothesis was that bird wing  
80 length variance is smaller in species that are more likely to experience strong stabilising selection on  
81 this trait. Specifically, we tested whether phenotypic variance in wing length is smaller in species  
82 showing aerial foraging and in long-distance migratory species compared to non-aerial feeding and  
83 resident species, respectively. We chose those two flight-related behaviours because deviations from  
84 an optimal wing length are expected to be more costly for species that rely on their wings more  
85 heavily, consequently leading to strong stabilising selection on wing length. Besides our main  
86 hypothesis, we used the collected data to assess the replicability of the sex-chromosome hypothesis, i.e.  
87 that the heterogametic sex shows higher phenotypic variation related to its haploid sex chromosomes  
88 (Reinhold and Engqvist 2013, Nakagawa et al. 2015), specifically for wing length in birds.

## 89 **Methods**

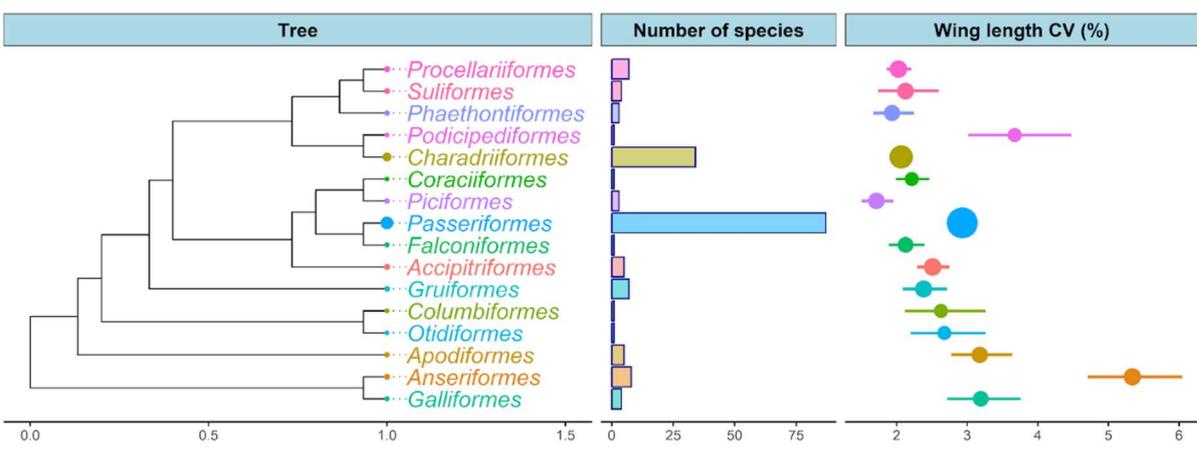
### 90 *Search strategy*

91 We collected wing length measurements (i.e., mean, standard deviation [SD], sample size [n]) for  
92 male and female birds from two sources: (1) the dataset from Reinhold and Engqvist (2013), which  
93 systematically searched literature up to 2011 to test the sex-chromosome hypothesis, and (2) an  
94 update of this dataset for the period 2011-2019, which we performed using similar keyword search  
95 terms, specifically: (“coefficient of variation” OR “CV?” OR “variation in size”) AND (“wing size” OR  
96 “wing length”) AND (“bird\$” OR [a list of all bird genera]) [for the complete list of bird genera see  
97 Capilla-Lasheras et al. 2022]. We searched for studies published between 2011 and 2019 to  
98 complement the existing Reinhold and Engqvist (2013) dataset without duplication. Our main  
99 searches were performed in the Web of Science Core Collection (databases: Science Citation Index  
100 Expanded (SCIE); Social Sciences Citation Index (SSCI); Arts & Humanities Citation Index (AHCI);  
101 Emerging Sources Citation Index (ESCI)), and performed additional searches in Biological Abstracts

102 and Google Scholar. To reduce a potential bias towards species with strong sexual dimorphism, we  
 103 also performed additional searches in the same platforms specifically targeting studies using  
 104 molecular sex determination by adding the keywords ("molecular sexing" OR "molecular sex  
 105 determination") to the search string detailed above. Importantly, our search was not intended to be  
 106 fully systematic nor comprehensive, but representative and unbiased across the bird phylogeny. We  
 107 believe that the full search strategy detailed above, which led to a representative sample of 172 bird  
 108 species (Figure 1), accomplished those goals.



109  
 110



111

112 **Figure 1.** The panel on top shows the full phylogenetic tree of all 172 bird species included in the  
113 phylogenetic multilevel meta-analyses performed to understand canalisation in wing length. The  
114 three panels below show the corresponding phylogenetic tree at the order level (16 orders) with its  
115 corresponding number of species, and the mean and 95% confidence interval for wing length CV (%).  
116 Dot size is relative to the number of CV values for each order.

117

118 *Data description*

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

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

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

172 To classify feeding type, we referred to the same sources as for migratory behaviour (i.e., IUCN and  
173 Wikipedia, using the latter only as supplementary), and if insufficient information was available from  
174 those sources, we also searched for information published in the scientific literature by performing  
175 searches in Google Scholar. We classified bird species as showing fully aerial feeding if they mainly  
176 collected their food by using their wings, which included species that mainly catch flying prey (e.g.,  
177 *Hirundo rustica* or *Tyrannus tyrannus*), plunge dive (e.g., gannets), are flying surface skimmers (e.g.,  
178 *Puffinus yelkouan*), fly underwater (e.g., *Fratercula cirrhata*) or hover in front of flowers (e.g.,  
179 hummingbirds). Bird species showing such feeding behaviours frequently but also frequently feeding  
180 while climbing, walking, hopping, or swimming with their feet (e.g., *Phylloscopus canariensis* or  
181 *Cinclus cinclus*) were classified as showing partially aerial feeding, whereas the remaining bird  
182 species, which may only sometimes or never feed ‘on the wing’, were classified as showing non-aerial  
183 feeding. To enhance comparability, we excluded entirely flightless species (e.g., kiwis, penguins, etc).

184

185 *Statistical analyses*

186 For all analyses, we calculated the log coefficient of variation (hereafter *lnCV*; equation 1) and its  
187 associated sampling variance (equation 2) using the function ‘escalc()’ from the R package ‘metafor’  
188 v.4.6-0 (Viechtbauer 2010) in R v.4.3.1 (R Core Team 2023), which includes a small-size bias correction  
189 (more in Nakagawa et al. 2015). Although we used *lnCV* for all the analyses, we present back-  
190 transformed percentage values throughout (i.e.,  $\exp(\ln CV) * 100$ ) to aid biological interpretation. For  
191 example, for a mean of 10 and an SD of 1, *lnCV* (i.e.,  $\ln(1/10)$ ) would be -2.30, whereas its back-  
192 transformed value would be 10%, which is more easily interpreted (i.e., SD is 10% of the mean) and  
193 comparable to previous studies on CVs. Our visualisations were performed following general  
194 recommendations in Yang et al. (2023) and primarily using the R packages ‘ggtree’ (Yu et al. 2017)  
195 and ‘orchaRd’ v.2.0 (Nakagawa et al. 2023).

196  $\ln CV = \log(SD/mean) + 1/2 * (n-1)$ , (eqn. 1)

197  $V \ln CV = SD^2 / n * mean^2 + 1/2 * (n-1)$ , (eqn. 2)

198 We ran phylogenetic multilevel meta-analytic and meta-regression models with *lnCV* as the response  
199 variable and where sampling variances were fit as a variance-covariance matrix that assumed a  
200 correlation of 0.5 between sampling variances obtained from the same study following  
201 recommendations in Noble et al. (2017). To further model nonindependence, we included the  
202 following six random effects in all models: (1) study identity, which encompassed estimates extracted  
203 from the same primary study, (2) pair identity, which encompassed male-female pair estimates  
204 obtained from the same population and sampling event, (3) population identity, which encompassed  
205 estimates obtained from the same geographic location, (4) phylogenetic correlation matrix, which  
206 modelled the phylogenetic relationships among the species included in our dataset, (5) species,  
207 which accounted for among-species variation additional to phylogeny (Cinar et al. 2022), and (6) unit-  
208 level observation identity, which modelled within-study variance. Decomposing species-level  
209 variation in two components (phylogenetic and non-phylogenetic) is necessary to obtain unbiased  
210 estimates of means and variance components (Cinar et al. 2022). For the phylogenetic correlation  
211 matrix among species, we used v1.5 of the new global tree of birds (McTavish et al. 2025), output in  
212 the 2024 taxonomy, which we obtained using the functions ‘taxonomyGet()’ and ‘extractTree()’ from  
213 the R package ‘cloutl’ v.0.1.2 (Miller et al. 2025). Branch lengths in that tree are in millions of years  
214 and generated as described in McTavish et al. (2025). To generate a phylogenetic tree at the order  
215 level, we extracted the phylogenetic information from the Open Tree of Life using the R package ‘rotl’  
216 v.3.1.0 (Michonneau et al. 2016), computed branch lengths using the Grafen method as implemented  
217 in the R package ‘ape’ v.5.7-1 (Paradis and Schliep 2019) and computed average effects per order to  
218 aid in visualising general differences in wing length CV among orders (Figure 1). We performed

219 several sensitivity analyses that modelled the variance-covariance matrix assuming a 0.25, 0.5 and  
220 0.75 correlation between sampling variances with the same study identity, pair identity and  
221 population identity, and results remained (virtually) the same. We provide all those sensitivity models  
222 in our data and code repository ("models/sensitivity\_analysis/").

223 We first ran a phylogenetic intercept-only multilevel meta-analytic model to (i) estimate the mean  
224 wing length CV across estimates while accounting for all the sources of nonindependence detailed  
225 above and (ii) explore heterogeneity among effect sizes. We then ran five phylogenetic multilevel uni-  
226 moderator meta-regressions (i.e., only one moderator included at a time) with the goal of testing our  
227 predictions regarding canalisation and the replicability of the sex-chromosome hypothesis. First, we  
228 tested whether wing length CV values differed between sexes (levels: male, female; "sex-  
229 chromosome hypothesis"; see section "S6. Sex-chromosome hypothesis"). Then, we tested whether  
230 some of the heterogeneity among effect sizes could be explained by differences among species and  
231 populations in feeding type or migration status. Last, we also ran two additional models including  
232 both moderators (i.e., feeding and migratory type), one at the species-specific level (Table S2) and  
233 another at the population-specific level (Table S3), to estimate the total amount of heterogeneity  
234 explained when both moderators are included in the same model. All models were run using the  
235 function 'rma.mv()' (method = "REML") from the R package 'metafor' v.4.6-0 (Viechtbauer 2010).

236 For the intercept-only models, we estimated heterogeneity across estimates using the pluralistic  
237 approach suggested by Yang et al. (2025) and implemented in the R package 'orchaRd' v.2.0  
238 (Nakagawa et al. 2023). This pluralistic approach involves estimating total heterogeneity ( $\sigma^2$ ), its  
239 source ( $I^2$ ) and two different metrics for its magnitude – a mean-standardised metric ( $CVH2$ ) and a  
240 variance-mean-standardised metric ( $M2$ ). For the meta-regressions, we report the percentage of  
241 variance explained by the moderator(s) as  $R^2_{\text{marginal}}$  (Nakagawa & Schielzeth, 2017), which we also  
242 calculated using the R package 'orchaRd'. Throughout, we present mean estimates with their  
243 associated 95% confidence intervals (CI) and prediction intervals (PI).

## 244 **Results**

245 Our final dataset contained 340 male estimates and 310 female estimates obtained from 181 studies  
246 from 314 different geographic locations. We obtained data from a total of 172 bird species for males  
247 (median = 1 estimate/species, range: 1-23; mean = 1.98, SD = 2.44; median = 42 individuals/estimate,  
248 range: 2-3179) and 163 species for females (median = 1 estimate/species, range: 1-25; mean = 1.90,  
249 SD = 2.47; median = 36 individuals/estimate, range: 2-4116; Figure 1). Our dataset covered 119  
250 genera from 63 families belonging to 16 orders of birds, with the most common order being  
251 Passeriformes (87 species), and the two most common families being Scolopacidae (11 species) and  
252 Laridae (10 species). In all, 407 estimates came from a previous compilation of bird wing-size

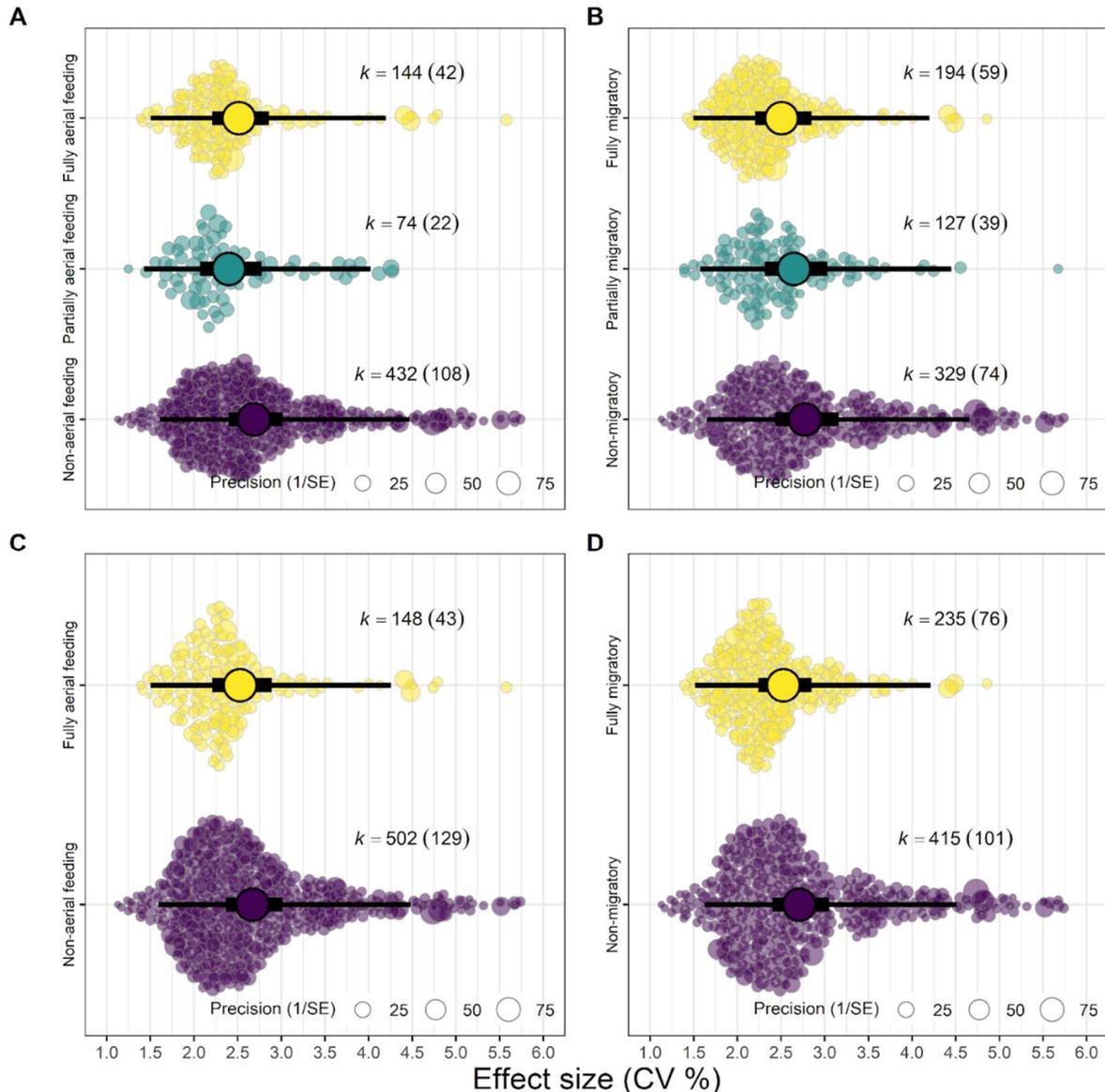
253 variation (Reinhold and Engqvist 2013) to which we added 243 additional estimates through our  
254 searches (see Supplementary Material S1-2 for additional descriptive statistics).

### 255 *Overall canalisation and heterogeneity*

256 Our phylogenetic multilevel meta-analysis showed that the mean wing length CV value across species  
257 was 2.63% (95% CI = 2.33 to 2.97, 95% PI = 1.56 to 4.43) with most differences among effect sizes  
258 coming from sources other than sampling error ( $I^2_{total} = 94.1\%$ ). Although  $I^2_{total}$  was high and indicated  
259 that heterogeneity in our dataset is, on average, around 16 times larger than statistical noise,  
260 absolute heterogeneity was relatively small ( $\sigma^2 = 0.067$ ). Indeed, the other two heterogeneity metrics  
261 showed that the magnitude of the heterogeneity in our data set is small ( $CVH2_{total} = 0.005$ ;  $M2_{total} =$   
262  $0.005$ ; Table S1, Figure S4), suggesting that the meta-analytic mean obtained is generalisable and  
263 replicable across different contexts. The discrepancy between  $I^2$  with both CVH2 and M2 is likely the  
264 consequence of the small typical sampling variance found in our dataset ( $\bar{\nu} = 0.004$ ), which can make  
265 the interpretation of  $I^2$  challenging – something often encountered with similar effect size measures  
266 such as the log coefficient of variation ratio ( $lnCVR$ ; more in Yang et al. 2025) as well as the much  
267 larger meta-analytic mean observed ( $lnCV = -3.69$ ) compared to  $\sigma^2$ .

### 268 *Species-specific level*

269 We ran two uni-moderator phylogenetic multilevel meta-regressions to understand among-species  
270 differences in wing length canalisation. First, non-aerial feeding was associated with larger wing  
271 length CV values (mean CV = 2.69%, 95% CI = 2.39 to 3.02) than those found in species showing  
272 partially (mean CV = 2.40%, 95% CI = 2.07 to 2.77) and fully aerial feeding (mean CV = 2.51%, 95% CI =  
273 2.21 to 2.86), but the difference was only statistically significant for partially aerial feeding (p-value =  
274 0.034) and not for fully aerial feeding (p-value = 0.176; Figure 2A). Second, non-migratory behaviour  
275 was associated with larger wing length CV values (mean CV = 2.77%, 95% CI = 2.43 to 3.16) than those  
276 found in partially migratory (mean CV = 2.65%, 95% CI = 2.31 to 3.03) and fully migratory species  
277 (mean CV = 2.50%, 95% CI = 2.20 to 2.85), but the difference was only statistically significant for fully  
278 migratory species (p-value = 0.017) and not for partially migratory species (p-value = 0.315; Figure  
279 2B). The percentage of variance explained by feeding ( $R^2_{marginal} = 2.57\%$ ) and migratory type ( $R^2_{marginal} =$   
280  $2.91\%$ ) separately were relatively small, but both moderators combined explained up to 4.50% of the  
281 heterogeneity (Table S2).



282

283 **Figure 2.** Flight-related behaviours such as partial or aerial feeding (A, C) and migration (B, D) are  
 284 generally associated with higher wing length canalisation (i.e., smaller values of coefficient of  
 285 variation, CV) across 172 bird species, but only some of the pairwise comparisons were statistically  
 286 significant (see text). Orchard plots show the mean estimate, 95% confidence intervals (thick whisker)  
 287 and 95% prediction intervals (thin whisker), with dot size being scaled by effect size precision (i.e.,  
 288 1/SE). The x-axes are cropped between CV values of 1 and 6 to aid visualisation, see section “S3.  
 289 Supplementary tables and figures” for the full version.

290

291 *Population-specific level*

292 We ran two additional uni-moderator phylogenetic multilevel meta-regressions to understand if  
 293 differences in wing length canalisation remained when species’ feeding and migratory type were

294 explored at the population level (e.g., different populations of a partially migratory species may differ  
295 in whether they are full or non-migratory). First, non-aerial feeding was associated with larger wing  
296 length CV values (mean CV = 2.67%, 95% CI = 2.36 to 3.02) than those found for fully aerial feeding  
297 (mean CV = 2.53%, 95% CI = 2.21 to 2.89), but the difference was not statistically significant (p-value =  
298 0.258; Figure 2C). Second, non-migratory behaviour was associated with statistically significantly  
299 larger wing length CV values (mean CV = 2.71%, 95% CI = 2.40 to 3.05) than those found for fully  
300 migratory populations (mean CV = 2.53%, 95% CI = 2.24 to 2.85; p-value = 0.043; Figure 2D). The  
301 percentage of variance explained by feeding ( $R^2_{\text{marginal}} = 0.80\%$ ) and migratory type ( $R^2_{\text{marginal}} = 1.69\%$ )  
302 separately were small and both moderators combined only explained up to 2.16% of the  
303 heterogeneity (Table S3).

## 304 Discussion

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

313 Indeed, our results showed evidence indicating that among-species differences in wing use explain  
314 some of the observed among-species differences in wing length CV values, where species that rely on  
315 their wings more heavily show higher canalisation on wing length. Despite that not all differences  
316 were statistically significant, bird species that mostly or partially feed on the wing showed, on  
317 average, around 5 to 10% lower wing length CV values than species not feeding on the wing. Along  
318 the same lines, and again despite not all differences being statistically significant, bird species that  
319 perform full or partial migration showed, on average, around 7 to 9% lower wing length CV values  
320 than resident species. Importantly, the results regarding migratory behaviour became clearer when  
321 we categorised vagrant species such as *Puffinus yelkouan*, *Phaethon aethereus* and *Sula sula* that fly  
322 over the ocean for extended periods of time as migratory rather than resident (Supplementary  
323 Material S5), further confirming that species heavily relying on their wings show lower wing length CV  
324 values.

325 Feeding and migratory type combined explained 4.5 to 6.4% of the observed differences in CV values,  
326 and these results provide preliminary support for our hypothesis that wing length should be under  
327 stronger stabilising selection in bird species strongly relying on their wings, leading to higher

328 canalisation on this trait. These results would, therefore, agree with previous among-species studies  
329 on other animal groups showing that traits under strong selection show low variation (i.e., higher  
330 canalisation). For example, previous among-species studies have shown that acoustic courtship traits  
331 under strong selection in amphibians and insects show lower variation (Reinhold 2011), and that  
332 canalisation in sperm length is stronger in bird species that show higher rates of extra-pair matings  
333 (Lifjeld et al. 2010). In line with these results, a recent meta-analysis of variance (Noble et al. 2021)  
334 has shown small between-individual variation in body size of hatchlings in snakes, lizards and turtles.  
335 Strong canalisation seems to be present, as no effect of incubation temperature could be observed  
336 for this trait that likely is under strong selection.

337 Finally, using our dataset we assessed the replicability and generalizability of the “sex-chromosome  
338 hypothesis”, which predicts higher variability in body size in the heterogametic sex compared to the  
339 homogametic sex, and for which evidence has been found in mammals, insects and birds (Reinhold &  
340 Engqvist 2013). Our results did not show clear evidence for the “sex-chromosome hypothesis” for  
341 bird wing length, as the difference between males and females was small and not statistically  
342 significant, and the moderator explained a negligible amount of heterogeneity. The observed effect  
343 was, nonetheless, in the expected direction, with males ( $CV_{males} = 2.61$ ) showing, on average, about  
344 1.5% lower wing length CV values than females ( $CV_{females} = 2.65$ ), which is lower than the mean effect  
345 size previously found for general body size in birds (Reinhold & Engqvist 2013: 4.0%). Our estimate for  
346 the difference in CV between males and females was also smaller than what was found in a recent re-  
347 analysis of Reinhold & Engqvist (2013)’s data using a phylogenetic multilevel meta-analysis similar to  
348 the one found in an earlier reanalysis (Nakagawa et al. 2015). The main difference between our data  
349 and that of Reinhold & Engqvist (2013) and its reanalysis by Nakagawa et al. (2015) is that ours  
350 focuses on wing length only, whereas Reinhold & Engqvist’s (2013) focused on general body size (i.e.,  
351 body mass and size, wing, head and beak size). To confirm that the difference in trait coverage likely  
352 led to the difference in results, we re-analysed the subset of Nakagawa et al. (2015) corresponding to  
353 bird wing length. This re-analysis, which included only a third of the studies and about half of the bird  
354 species included in our study, confirmed little support for the “sex-chromosome hypothesis”: males  
355 showed, on average, a non-statistically significant 1.5% lower wing length CV values than females  
356 (Supplementary Material S6). Our replication attempt showing that the sex-chromosome hypothesis  
357 explains negligible variation among wing length CV values suggests that the previous evidence for the  
358 sex-chromosome hypothesis has likely been underestimated by including this trait. Therefore, we call  
359 for future research reassessing the sex-chromosome hypothesis across animals with a special focus  
360 on more variable traits such as body mass.

361

362 **Conclusions**

363 Our analyses indicate that coefficients of variation in wing length are about 7-9% smaller in migratory  
364 bird species compared to non-migratory bird species, and that birds feeding on the wing have about  
365 5-10% smaller values than species that do not usually feed on the wing. Although not all comparisons  
366 were statistically significant, their directions generally supported the expectation that stabilising  
367 selection on wing length is stronger for species using their wings more intensely during foraging and  
368 migration. The observed differences between these groups only explained a small proportion of the  
369 total variation, likely because a large proportion of the remaining variance is attributed to  
370 phylogenetic differences. This suggests that different levels of canalisation have evolved within the  
371 different bird taxa. We suggest that future research should make use of the meta-analytic techniques  
372 showcased in our study together with large-scale datasets to study trait canalisation across taxonomic  
373 groups.

374

375 **Conflict of interest statement**

376 The authors declare no competing interests.

377 **Authors contributions**

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

382 **Acknowledgements**

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

387 **Data and code availability statement**

388 All data and code are available at the following Zenodo repository  
389 (<http://doi.org/10.5281/zenodo.18152048>), as well as in the connected GitHub repository  
390 ([https://github.com/ASanchez-Tojar/meta-analysis\\_canalisation\\_bird\\_wing](https://github.com/ASanchez-Tojar/meta-analysis_canalisation_bird_wing)).

391

392

393 **References**

- 394 Alonso, J.C., Magaña, M., Alonso, J.A., Palacín, C., Martín, C.A. & Martín, B. (2009) The most extreme  
395 sexual size dimorphism among birds: allometry, selection, and early juvenile development in the  
396 great bustard (*Otis tarda*). *The Auk* **126**, 657–665.
- 397 Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M., Novak, M., Rudolf, V.H.W.,  
398 Schreiber, S.J., Urban, M.C. & Vasseur, D.A. (2011) Why intraspecific trait variation matters in  
399 community ecology. *Trends in Ecology & Evolution* **26**, 183–192.
- 400 Capilla-Lasheras, P., Thompson, M.J., Sánchez-Tójar, A., Haddou, Y., Branston, C.J., Réale, D.,  
401 Charmantier, A. & Dominoni, D.M. (2022) A global meta-analysis reveals higher variation in  
402 breeding phenology in urban birds than in their non-urban neighbours. *Ecology Letters* **25**,  
403 2552–2570.
- 404 Cava, J.A., Perlut, N.G. & Travis, S.E. (2019) Heritability and evolvability of morphological traits of  
405 Savannah sparrows (*Passerculus sandwichensis*) breeding in agricultural grasslands. *PLOS ONE*  
406 **14**, e0210472.
- 407 Cinar, O., Nakagawa, S. & Viechtbauer, W. (2022) Phylogenetic multilevel meta-analysis: A simulation  
408 study on the importance of modelling the phylogeny. *Methods in Ecology and Evolution* **13**, 383–  
409 395.
- 410 Delhey, K., Szecsenyi, B., Nakagawa, S. & Peters, A. (2017) Conspicuous plumage colours are highly  
411 variable. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20162593.
- 412 Des Roches, S., Post, D.M., Turley, N.E., Bailey, J.K., Hendry, A.P., Kinnison, M.T., Schweitzer, J.A. &  
413 Palkovacs, E.P. (2018) The ecological importance of intraspecific variation. *Nat Ecol Evol* **2**, 57–64.
- 414 Fattorini, N., Costanzo, A., Romano, A., Rubolini, D., Baillie, S., Bairlein, F., Spina, F. & Ambrosini, R.  
415 (2023) Eco-evolutionary drivers of avian migratory connectivity. *Ecology Letters* **26**, 1095–1107.
- 416 Flatt, T. (2005) The evolutionary genetics of canalization. *The Quarterly review of Biology* **80**, 287–  
417 316.
- 418 Geiler-Samerotte, K., Bauer, C., Li, S., Ziv, N., Gresham, D. & Siegal, M. (2013) The details in the  
419 distributions: why and how to study phenotypic variability. *Current Opinion in Biotechnology* **24**,  
420 752–759.
- 421 Groth, B.R., Huang, Y., Monette, M.J. & Pool, J.E. (2018) Directional selection reduces developmental  
422 canalization against genetic and environmental perturbations in *Drosophila* wings. *Evolution* **72**,  
423 1708–1715.

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

426 Hayden, E.J., Bratulic, S., Koenig, I., Ferrada, E. & Wagner, A. (2014) The effects of stabilizing and  
427 directional selection on phenotypic and genotypic variation in a population of RNA enzymes.  
428 *Journal of Molecular Evolution* **78**, 101–108.

429 James, F.C. (1970) Geographic size variation in birds and its relationship to climate. *Ecology* **51**, 365–  
430 390.

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

434 Kleven, O., Laskemoen, T., Fossøy, F., Robertson, R.J. & Lifjeld, J.T. (2008) Intraspecific variation in  
435 sperm length is negatively related to sperm competition in passerine birds. *Evolution* **62**, 494–  
436 499.

437 Levis, N.A. & Pfennig, D.W. (2020) Plasticity-led evolution: A survey of developmental mechanisms  
438 and empirical tests. *Evolution & Development* **22**, 71–87.

439 Lifjeld, J.T., Laskemoen, T., Kleven, O., Albrecht, T. & Robertson, R.J. (2010) Sperm length variation as a  
440 predictor of extrapair paternity in passerine birds. *PLOS ONE* **5**, e13456.

441 McTavish, E.J., Gerbracht, J.A., Holder, M.-T., Iliff, M.J., Lepage, D., Rasmussen, P.C., Redelings, B.D.,  
442 Sánchez Reyes, L.L. & Miller, E.T. (2025) A complete and dynamic tree of birds. *Proc. Natl. Acad.*  
443 *Sci. U.S.A.* **122**, e2409658122.

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

446 Miller E.T., McTavish E.J. & Sánchez Reyes L.L. (2025). “cloodl: Fetch and Explore the Cornell Lab of  
447 Ornithology.” Open Tree of Life Avian Phylogeny.” <<https://github.com/eliotmiller/cloodl>>.

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

450 Moran, N.P., Sánchez-Tójar, A., Schielzeth, H. & Reinhold, K. (2021) Poor nutritional condition  
451 promotes high-risk behaviours: a systematic review and meta-analysis. *Biological Reviews* **96**,  
452 269–288.

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

455 Nakagawa, S., Johnson, P.C.D. & Schielzeth, H. (2017) The coefficient of determination  $R^2$  and intra-  
456 class correlation coefficient from generalized linear mixed-effects models revisited and  
457 expanded. *Journal of The Royal Society Interface* **14**, 20170213.

458 Nakagawa, S., Lagisz, M., O’Dea, R.E., Pottier, P., Rutkowska, J., Senior, A.M., Yang, Y. & Noble, D.W.A.  
459 (2023) orchaRd 2.0: An R package for visualising meta-analyses with orchard plots. *Methods in*  
460 *Ecology and Evolution* **14**, 2003–2010.

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

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

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

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

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

473 Pélabon, C., Hansen, T.F., Carter, A.J.R. & Houle, D. (2010) Evolution of variation and variability under  
474 fluctuating, stabilizing, and disruptive selection. *Evolution* **64**, 1912–1925.

475 Price, G.R. (1972) Fisher’s ‘fundamental theorem’ made clear. *Annals of Human Genetics* **36**, 129–  
476 140.

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

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

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

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

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

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

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

491 Sharma, M.D., Minder, A.M. & Hosken, D.J. (2013) No association between sperm competition and  
492 sperm length variation across dung flies (Scathophagidae). *Journal of Evolutionary Biology* **26**,  
493 2341–2349.

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

496 Subasinghe, K., Symonds, M.R.E., Vidal-García, M., Bonnet, T., Prober, S.M., Williams, K.J. & Gardner,  
497 J.L. (2021) Repeatability and validity of phenotypic trait measurements in birds. *Evolutionary*  
498 *Biology* **48**, 100–114.

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

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

503 Thompson, M.J., Capilla-Lasheras, P., Dominoni, D.M., Réale, D. & Charmantier, A. (2022) Phenotypic  
504 variation in urban environments: mechanisms and implications. *Trends in Ecology & Evolution*  
505 **37**, 171-182 <https://doi.org/10.1016/j.tree.2021.09.009>

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

508 Willmore, K.E., Young, N.M. & Richtsmeier, J.T. (2007) Phenotypic variability: its components,  
509 measurement and underlying developmental processes. *Evolutionary Biology* **34**, 99–120.

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

514 Yang, Y., Noble, D.W.A., Spake, R., Senior, A.M., Lagisz, M. & Nakagawa, S. (2025) A pluralistic  
515 framework for measuring, interpreting and decomposing heterogeneity in meta-analysis.  
516 *Methods in Ecology and Evolution* **16**, 2710-2725.

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

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

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

524