

1 **Automatic identification of bird females using egg phenotype**

2 Michal Šulc¹, Anna E. Hughes², Jolyon Troscianko³, Gabriela Štětková^{1,4}, Petr Procházka¹, Milica
3 Požgayová¹, Lubomír Piálek^{1,5}, Radka Piálková^{1,5}, Vojtěch Brlík^{1,6} and Marcel Honza¹

4
5 ¹ *Czech Academy of Sciences, Institute of Vertebrate Biology, Brno, Czech Republic*

6 ² *Department of Psychology, University of Essex, Colchester, U.K.*

7 ³ *Centre for Life and Environmental Sciences, University of Exeter, Penryn, U.K.*

8 ⁴ *Department of Botany and Zoology, Faculty of Sciences, Masaryk University, Brno, Czech Republic*

9 ⁵ *Faculty of Natural Sciences, University of South Bohemia, České Budějovice, Czech Republic*

10 ⁶ *Department of Ecology, Faculty of Science, Charles University, Prague, Czech Republic*

11 *Corresponding author: sulc@ivb.cz*

12

13 **Abstract**

14 1. Identification of individuals greatly contributes to understanding animal ecology and
15 evolution, and in many cases can only be achieved using expensive and invasive
16 techniques. Advances in computing technology offer alternative cost-effective techniques
17 which are less invasive and can discriminate between individuals based on visual and/or
18 acoustic cues. Here, we employ human assessment and an automatic analytical approach
19 to predict the identity of common cuckoo (*Cuculus canorus*) females based on the
20 appearance of their eggs. The cuckoo's secretive brood parasitic strategy makes studying
21 its life history very challenging. Eggs were analysed using calibrated digital photography
22 for quantifying spotting pattern, size and shape, and spectrometry for measuring colour.
23 Cuckoo females were identified from genetic sampling of their nestlings, allowing us to
24 determine the accuracy of human and automatic female assignment. Finally, we used a
25 novel 'same-different' approach that uses both genetic and phenotypic information to
26 assign eggs that were not genetically analysed.

- 27 2. Our results supported the ‘constant egg-type hypothesis’, showing that individual cuckoo
28 females lay eggs with a relatively constant appearance and that eggs laid by different
29 females differ more than eggs laid by the same female. The accuracy of unsupervised
30 hierarchical clustering was comparable to assessments of experienced human observers.
31 Supervised random forest analysis showed better results, with higher cluster accuracy.
32 Same-different analysis was able to assign 22 of 87 unidentified cuckoo eggs to seven
33 already known females.
- 34 3. Our study showed that egg appearance on its own is not sufficient for identification of
35 individual cuckoo females. We therefore advocate genetic analysis to be used for this
36 purpose. However, supervised analytical methods reliably assigned a relatively high
37 number of eggs without genetic data to their mothers which can be used in conjunction
38 with genetic testing as a cost-effective method for increasing sample sizes for eggs where
39 genetic samples could not be obtained.

40

41

42 **Introduction**

43 Identification of individuals is important in zoological research, particularly when investigating
44 variation among or within individuals in a population. Traditionally, capture-mark-recapture
45 techniques have been used to monitor individuals during their lifetime (Lindberg, 2012; Jung,
46 Boonstra, & Krebs, 2020). This method has been improved by employing more sophisticated
47 methods such as attaching GPS (global positioning system) and radio transmitters or RFID (radio
48 frequency identification) tags (Krause et al., 2013) that allow researchers to investigate the spatial
49 activity of animals in more detail. However, these methods still require capturing and tagging that
50 is usually time-consuming, expensive (depending on the method used), and may reduce animal
51 welfare (Weinstein, 2018). Therefore, cost-effective indirect approaches have been developed to
52 identify and monitor individuals within the same species.

53

54 These indirect approaches rely on the fact that individuals differ from each other visually or
55 acoustically and this variation may be used for their identification. Indeed, it has been shown that
56 e.g. face (Deb et al., 2018; Hansen et al., 2018; Hou et al., 2020) and body pattern data (Hiby et
57 al., 2009; Bolger, Morrison, Vance, Lee, & Farid, 2012; Crall, Stewart, Berger-Wolf, Rubenstein,
58 & Sundaresan, 2013; Ferreira et al., 2020) captured from photographs may allow discrimination
59 of individuals. Similarly, sounds produced by animals (especially bird song) also seem to serve
60 as a good individual fingerprint (Blumstein et al., 2011; Petrusková, Pišvejcová, Kinštová, Brinke,
61 & Petrušek, 2016; Ptacek, Machlica, Linhart, Jaska, & Muller, 2016). Recently, applying modern
62 computer technology and artificial intelligence techniques (such as convolutional neural networks)
63 that automate the analysis of various types of data from different sources such as pictures or
64 audio recordings has made these methods reliable and applicable for various animal taxa
65 (Hansen et al., 2018; Christin et al., 2019; Ferreira et al., 2020; Hou et al., 2020).

66

67 However, for many species, identification of all individuals in a population is still not
68 straightforward e.g. due to their secretive behaviour or due to the fact that it is difficult to catch
69 them. Here, we focus on one group of animals that are especially challenging to study – avian
70 brood parasites. There are more than a hundred obligate brood parasites that never build their
71 own nests and instead lay their eggs into nests of other species (Davies, 2010). Even more
72 species belong to conspecific brood parasites that only occasionally lay eggs into nests of the
73 unrelated conspecifics (Bruce E. Lyon & Eadie, 2008). Brood parasites and their hosts have been
74 the focus of considerable research into co-evolutionary arms races (Soler, 2017). But since brood
75 parasites only lay eggs and then usually do not return to host nests (but see Šulc et al. 2020),
76 and because egg laying is (especially in obligate brood parasites) very quick (McMaster, Neudorf,
77 Sealy, & Pitcher, 2004; Gloag, Fiorini, Reboreda, & Kacelnik, 2013; Jelínek, Šulc, Štetková, &
78 Honza, 2020), direct observation of parasitism in nature is difficult and identification of parasitic

79 females is problematic. As a consequence, many important aspects of females' life history
80 strategy are still poorly understood; in obligate brood parasites, we e.g. still know little about
81 spatio-temporal distribution of their egg laying, consistency in host selection and the total number
82 of eggs they lay during a breeding season. Conspecific brood parasitism is even less understood
83 and we do not even know why some females adopt this strategy (Bruce E. Lyon & Eadie, 2008).
84

85 The idea of identifying bird females according to the appearance of the eggs they laid depends
86 on the presumption that within-clutch variation in egg appearance is lower than between-clutch
87 variation which has been confirmed for several species including brood parasites (Øien, Moksnes,
88 & Røskaft, 1995; McRae & Burke, 1996; Paillisson, Latraube, Marion, & Bretagnolle, 2008; Höltje,
89 Mewes, Haase, & Ornés, 2016). This approach has therefore been applied (although with
90 caveats) for the identification of parasitic eggs in some conspecific brood parasite species
91 (Gibbons, 1986; Møller, 1987; Jackson, 1992; Petersen, 1992; Bruce E. Lyon, 1993; McRae &
92 Burke, 1996; B. E Lyon, 2003). However, some studies that estimated accuracy of parasitic egg
93 identification showed ambiguous results for some species (Ådahl, Lindström, Ruxton, Arnold, &
94 Begg, 2004; Pöysä, Lindblom, Rutila, & Sorjonen, 2009; Eadie, Smith, Zadworny, Kühnlein, &
95 Cheng, 2010; Lemons, Sedinger, & Randle, 2011; Petrželková, Pöysä, Klvaňa, Albrecht, & Hořák,
96 2017) and for others this method did not work at all (Brown & Sherman, 1989; Cariello, Lima,
97 Schwabl, & Macedo, 2004; Grønstøl, Blomqvist, & Wagner, 2006; Griffith, Barr, Sheldon, Rowe,
98 & Burke, 2009; Roy, Parker, & Gates, 2009). One of the reasons why many studies found low
99 accuracy of identification might be that closely related females lay indistinctive eggs. Several
100 studies showed that egg appearance, namely egg color (Wei, Bitgood, & Dentine, 1992; Collias,
101 1993; Morales et al., 2010), spotting pattern (Gosler, Barnett, & James Reynolds, 2000) and egg
102 size (Christians, 2002) are highly heritable traits which might complicate such analyses especially
103 in inbred populations. Another explanation might be that previous studies did not use the full

104 potential of egg variability (e.g. none of the presented studies measured egg colour in the
105 ultraviolet range of spectrum).

106

107 Identification of parasitic females using egg appearance has also been attempted in the common
108 cuckoo (hereafter cuckoo), but was unsuccessful (Moksnes et al., 2008). However, the study
109 assessed cuckoo eggs from a human perspective, with people sorting the eggs based on
110 photographs. To date, there have been no attempts to use more objective quantification methods
111 for egg classification in the cuckoo. These objective methods, such as spectrophotometry for
112 measuring colours (including the ultraviolet part of the spectrum), and image analysis of
113 photographs for quantifying spotting pattern, size and shape of eggs are now available, and may
114 allow more accurate classification that can be carried out in an automated manner.

115

116 In this study, we employ a novel automatic analytical approach to analyse phenotypic features of
117 cuckoo eggs such as dimensions (size, shape), colour and spotting pattern to predict maternal
118 identity. If successful, this low-cost and minimally invasive female identification method would
119 greatly facilitate studies into a range of key questions regarding this secretive brood parasitic
120 species. We also performed human assessment based on sorting of photos with cuckoo eggs to
121 compare the reliability of both methods with the true identity acquired from molecular analyses.
122 Moreover, we believe this automatic technique might be also used in other brood parasitic
123 systems or in species where females are difficult to catch (see e.g. Höltje et al. 2016). Finally, it
124 has been suggested that similarly looking eggs laid by different cuckoo females may belong to
125 closely related females, e.g. mother and daughter (Moksnes et al., 2008). Therefore, we will for
126 the first time investigate the relationship between the genetic distance of individual cuckoo
127 females and the phenotypic distance of their eggs.

128

129

130 **Material and Methods**

131 *Study system and data collection*

132 All data were collected in the fishpond area between Mutěnice (48°54'N, 17°02'E) and Hodonín
133 (48°51'N, 17°07'E) in South Moravia, Czech Republic from May to July 2017. We systematically
134 searched the littoral vegetation for the great reed warbler (*Acrocephalus arundinaceus*) and
135 Eurasian reed warbler (*Acrocephalus scirpaceus*) nests. Most great reed warbler (hereafter GRW)
136 nests were found during the building stage when mapping male territories and mating status
137 (Bensch, 1996). The rest of the GRW and all Eurasian reed warbler (hereafter RW) nests were
138 found in different stages of breeding by systematic searching. If possible, all GRW nests were
139 checked every day from the nest building stage until clutch completion and approximately every
140 third day during incubation. All RW nests were checked approximately every second day during
141 laying stage and extensively during incubation. GRWs experienced 92 % (59 out of 64 nests) and
142 RWs 20 % (91 out of 456 nests) cuckoo parasitism rate. Multiple parasitism was also common;
143 37 of 59 and two of 91 parasitized GRW and RW nests, respectively, were parasitized by more
144 than one cuckoo egg.

145
146 When a cuckoo egg was found in a host nest, we immediately measured its colour and took a
147 photo (see below) to avoid colour change during the incubation period (Hanley et al., 2016). In
148 the cases of multiply parasitized nests, we removed the newly laid cuckoo egg(s), transferred
149 them to an incubator (HEKA-Kongo; HEKA-Brutgeräte, Rietberg, Germany) and incubated them
150 artificially to prevent sample losses caused by the cuckoo nestlings (Honza, Vošlajerová, &
151 Moskát, 2007). The removed cuckoo eggs were either incubated until hatching and chicks placed
152 into non-parasitized host nests (for other experiments) or we froze the eggs before hatching for
153 the future genetic analysis (see *Genotyping and kinship analysis* section). We took a blood
154 sample from all 10-days old cuckoo nestlings from their ulnar or medial tarsometatarsal vein
155 (approx. 25 µl). Finally, we mist-netted 29 and 16 adult cuckoo males and females, respectively,

156 and collected their blood samples from the ulnar vein (approx. 25 μ l). All blood samples were
157 stored in 96% ethanol until later genetic analyses.

158

159 Altogether we found 203 cuckoo eggs (121 and 82 in the GRW and RW nests, respectively). We
160 photographed and measured the colour of 192 of them. Among these photographed cuckoo eggs,
161 genetic samples were collected from 105 nestlings or embryos.

162

163 *Measurements of egg appearance*

164 To obtain background colour we measured reflectance using JAZ Spectrometer (Ocean Optics,
165 Dunedin, FL, U.S.A.) in the range 300–700 nm, as that is the wavelengths range birds can
166 perceive (Cuthill 2006). We took nine measurements (each covering approximately 1 mm²) at
167 three different parts of the egg (sharp pole, middle part and blunt pole). Since we focused on
168 background colour, we avoided measuring dark spots (Šulc et al. 2016). For each egg, we used
169 the measurement with the highest reflectance that best corresponded to the colour of the
170 background (Šulc et al. 2019).

171

172 Spotting pattern was calculated from digital images taken by a Canon EOS 700D with prime
173 Canon EF 40 mm lens. All photos were taken under standardized diffuse sunlight conditions
174 (using a photography light tent), at the same angle and from the same distance and were referred
175 to a grey standard (X-Rite Colour Checker Grey Scale Chart) with known reflectance. Image
176 calibration, pattern analysis, analysis of shape and measurements of size were performed in
177 ImageJ software (Schneider et al. 2012) using the Multispectral Image Calibration and Analysis
178 (MICA) Toolbox (van den Berg et al. 2020). A scale bar was included in each photo, meaning that
179 all images were equally rescaled to the scale of the smallest image (30 pixels/mm), and egg
180 dimensions were obtained from the photos. For pattern investigation we applied a granularity
181 analysis approach (Troschianko and Stevens 2015) that creates a bandpass ‘energy’ spectrum

182 across a range of spatial frequencies, and then the pattern energy at each frequency band was
183 measured as the standard deviation of the filtered image (for details see Šulc et al. 2019 and van
184 den Berg et al. 2020). Since pattern energy does not discriminate the direction of the pattern (it
185 cannot distinguish between dark spots on light background and light spots on dark background),
186 we also calculated the 'skew' of the pattern, which quantifies the asymmetry of the pattern
187 luminance distribution. A negative value of skew implies there are more spots than background
188 colour, while a positive skew implies there is more background colour than spots. Skew was
189 measured independently at each granularity band. All colour measurements and photos were
190 taken by a single person (M.Š.) to ensure high consistency of the data.

191

192 *Genotyping and kinship analysis*

193 DNA was isolated from the blood of adults and nestlings or tissues of embryos. (65 nestlings, 41
194 embryos, 29 adult males and 16 adult females). We estimated kinship relationships from nuclear
195 SNPs and mitochondrial DNA haplotypes enabling us to exclude highly implausible maternal (or
196 maternal-sibling) relationships in the inferred genealogy. As an additional criterion we also used
197 the laying date of cuckoo eggs (data from daily checks) because it is known that a cuckoo female
198 cannot lay her eggs more often than every second day (Chance, 1922; Payne, 1973; Wyllie, 1981).

199

200 To acquire the SNP dataset, we genotyped all samples with the ddRAD (double digest restriction-
201 site associated DNA) technique (Petersen, 1992) following the protocol of (Piálek et al., 2019).
202 Two prepared libraries were sequenced on an Illumina HiSeq4000 system (2 lanes, 150 cycles
203 P/E) in the EMBL Genomic Core Facility, Heidelberg, Germany. The obtained RAD-tags were
204 processed in Stacks v2.4 (Catchen et al. 2011, Rochette et al. 2019) and mapped on the *Cuculus*
205 *canorus* genome GCA000709325.1 (<https://www.ncbi.nlm.nih.gov>) with Bowtie2 assembler
206 v2.2.4 (Langmead & Salzberg, 2012). Only loci with 95% or higher presence of individuals were
207 scored and further filtered based on Hardy–Weinberg equilibrium, linkage disequilibrium and

208 minimum minor allele frequency (0.4) in PLINK v1.9 (Purcell et al., 2007) which resulted in a
209 dataset with 1620 markers. Kinship relationships were estimated using Colony (Jones & Wang,
210 2010) based on >1000 nuclear SNPs.

211
212 For the mitochondrial haplotype analysis, we sequenced a 411-bp portion of the left-hand
213 hypervariable control region (Gibbs et al., 2000; Fossøy et al., 2011, 2012). Mitochondrial
214 sequence data were assembled and manually checked in Geneious v10.2.6 (Kearse et al., 2012)
215 and haplotypes were estimated based on a distance matrix with up to 1% tolerance (approx. 4
216 mutations) for genotyping errors.

217
218 Kinship analysis assigned the offspring (n = 105) to 30 clusters containing 1–10 eggs each (Fig.
219 1–5 in Supplementary material). Among these 30 clusters, nine corresponded to females that
220 were caught and genotyped as described above. Thus, we were able to calculate genetic
221 distances among these females.

222
223 For subsequent analysis dealing with egg phenotype (human and automatic assessment, see
224 below), we removed females to which only one egg has been assigned (n = 10), meaning that we
225 used a final dataset of 95 eggs laid by 20 females.

226
227 *Human assessment*

228 We printed 95 photographs of cuckoo eggs that were standardized in their colour and size using
229 the MICA Toolbox (van den Berg et al. 2020; Fig.1–7 in supplementary material). We then asked
230 twelve people to sort these photographs and create groups of pictures representing individual
231 females according to similarity in egg appearance. Firstly, we asked them to sort these pictures
232 into an unknown number of groups and, secondly, we asked them to sort these pictures into 20
233 groups corresponding to the real number of females identified by genetic assessment of identity.

234 For the assessments, we asked 1) five people with no experience with egg appearance from wild
235 animals, 2) three students of avian ecology that had experience with egg appearance from wild
236 birds but had never seen cuckoo eggs and 3) four people (mostly co-authors of this manuscript)
237 that had years of experience with cuckoo eggs. All participants received no other information
238 about the eggs. Cluster similarity in egg classification was assessed using the adjusted Rand
239 index, which provides a corrected-for-chance measure of the similarity between two data
240 clusterings, implemented using the 'cluster_similarity' function from the R package *clustereval*
241 (Ramey, 2012).

242

243 *Automatic assessment*

244 We developed an automatic method based on the similarities/differences of cuckoo egg
245 phenotypes. In the first step, we collected colour, pattern and dimension data from calibrated
246 photographs and spectrophotometry data (for details, see Šulc et al., 2019) for all cuckoo eggs.
247 Initially, we conducted Principal component analysis (PCA) on different aspects of the egg
248 photographs, in order to avoid the use of correlated variables in the models.

249

250 *Spectral data:* a spectrophotometer was used to assess the background colour of cuckoo eggs
251 (for details, see *Measurements of egg appearance*). PCA was carried out using binned, scaled
252 spectral data created in the R package *pavo* (Maia, Gruson, Endler, & White, 2019), and two
253 spectral PCA components were used in the final dataset (based on scree plot inspection). We
254 also used two other spectral measures extracted from *pavo*: the mean brightness (B2; mean
255 relative reflectance over the entire spectral range) (Hill & McGraw, 2006) and the position of the
256 ultraviolet (UV) peak (defined as a wavelength within the range of 300–360nm where reflectance
257 reached the highest point).

258

259 *Shape data:* the variables entered into the PCA were length, maximum width, volume, ellipse
260 deviation and surface area (Troscianko, 2014). Three shape PCA components were selected for
261 inclusion into the final dataset based on scree plot inspection.

262

263 *Pattern data:* the variables entered into the PCA were 12 pattern energies measured at a range
264 of scales (from 1 to 0.0221 in steps of $1/\text{square root of } 2$) across the whole egg (Troscianko &
265 Stevens, 2015), and 12 pattern energy skew values measured at the same range of scales across
266 the whole egg. We also included a measure of total pattern energy across the whole egg. Finally,
267 we divided up each egg into three segments and measured the total pattern energy in each
268 segment as well as the standard deviation between segments, to get a measure of how variable
269 the patterning was across the egg. Three pattern PCA components were selected for inclusion
270 into the final dataset based on scree plot inspection.

271

272 *Luminance data:* we analysed luminance from photographs, including both the spots and
273 background areas of the eggs. We divided the egg up into three segments and took the average
274 luminance and the standard deviation of luminance across each segment, as well as the standard
275 deviation of luminance across all three segments. One luminance PCA component was selected
276 for inclusion into the final dataset based on scree plot inspection.

277

278 In total, the final dataset contained 11 egg phenotypic traits that were used for clustering analysis.

279

280 *Within- and between-female variability in egg appearance*

281 To create a metric of within-female variance, we calculated the standard deviation for each
282 phenotypic trait within one female, and then took a mean value across all traits, giving an average
283 variability value for each female.

284 To create a metric of between-female variance, we calculated the average value of each
285 phenotypic trait for each female (i.e. created an “average” egg) and then calculated the standard
286 deviation for each phenotypic trait across all females. We then averaged these standard
287 deviations to create a measure of between-female variability across all traits. All trait values were
288 scaled to ensure comparability across different traits.

289 To test whether within-female variance is lower than between-female variance, we conducted a
290 one-sample t-test where the within-female variance metric (n=20) is compared with the test value
291 (the between-female variance value).

292

293 *Unsupervised learning*

294 Firstly, we carried out hierarchical clustering to attempt to cluster the eggs via visual similarity
295 without any training or further information (e.g. number of females present). All variables were
296 scaled for this analysis. To assess the accuracy of this method, we cut the tree by specifying the
297 real number of groups (20) and assessed the cluster similarity between the predictions of the
298 hierarchical model compared to the real data using the adjusted Rand index, as before.

299

300 *Supervised learning*

301 *Female clustering:* We used a random forest model with a ‘leave-one-out’ cross-validation
302 approach (Stone, 1974). For each egg in the dataset, the model was trained using a dataset
303 consisting of all other eggs, and then tested using the focal egg. The model attempted to classify
304 each egg to a given female, and the accuracy of the model was assessed using the classification
305 accuracy value, and through cluster similarity values, as before (taking the average of 1000 runs,
306 as random forest modelling is a stochastic process). We also fitted a random forest model to the
307 full dataset to allow us to assess the importance of the different variables included in the model
308 (using the mean decrease in accuracy).

309

310 *Same/different analysis:* We used an approach where a random forest model was trained to label
311 pairs of eggs as 'same' or 'different'. The training set used 3000 'same' rows, where the two eggs
312 were from the same female (but are not identical to each other) and 3000 'different' rows, where
313 the two eggs were from different females.

314

315 To test our models, we tested each egg in the labelled dataset on all eggs sequentially, including
316 itself. We first tested whether the model recognised the identical eggs as being the same. We
317 then tested whether each egg was only paired with other eggs from the same female i.e. whether
318 the model could uniquely identify clusters of eggs that belonged together. The entire process
319 (creating a training set, training the random forest model and testing the model) was repeated
320 1000 times.

321

322 For the unlabelled dataset, we calculated how many times in each of these 1000 runs the target
323 egg was matched with a cluster of eggs that were from the same female. If the percentage was
324 greater than 95%, we considered this egg as a candidate for being from this female. To
325 corroborate this conclusion, we used non-phenotypic data e.g. laying dates, laying locality and
326 host species.

327

328 *Phenotypic-genotypic similarity*

329 We had genetic samples for 9 of the adult females, allowing us to create a genetic distance matrix.
330 To compare the phenotypic-genotypic similarity between these females and their eggs, we
331 created a trait distance matrix by taking means of the phenotypic parameters from their egg data,
332 and then using Euclidean distance as the distance metric. We compared the genetic distance
333 matrix with the trait distance matrix using a Mantel test, a statistical test of the correlation between
334 two matrices, implemented in the vegan package in R using the Kendall method (as this is most
335 appropriate for a small dataset). We also split the phenotype data into different components

336 (spectral, pattern and shape) and calculated the phenotype-genotype similarities for each of these
337 components separately, to test whether different aspects of the egg phenotype are differentially
338 correlated with the female genotypes.

339

340 **Results**

341 *Within- and between-female variability in egg appearance*

342 Within-female variance was variable, with some females having very similar eggs (e.g. female 13
343 - within-female variance = 0.33, Fig. 2 in Supplementary material) and others having relatively
344 high variance (e.g. female 29 - within-female variance = 1.31, Fig. 4 in Supplementary material).
345 The between-female variance (comparing across females, using an “average egg” for each
346 female) was 1.832 (mean of trait standard deviations, $n = 11$ traits; $SD = 1.021$). Overall, within-
347 female variance (mean = 0.850, $SD = 0.295$) was smaller than between-female variance (one
348 sample t test, $t = 14.87$, $df = 19$, $p < 0.001$). Variability in the egg appearance is also visible in Fig.
349 1 where the two most informative variables in the random forest analysis (PC2 for pattern and
350 PC2 for spectral data), are plotted (see below and Table 2).

351

352

353

354

355

356

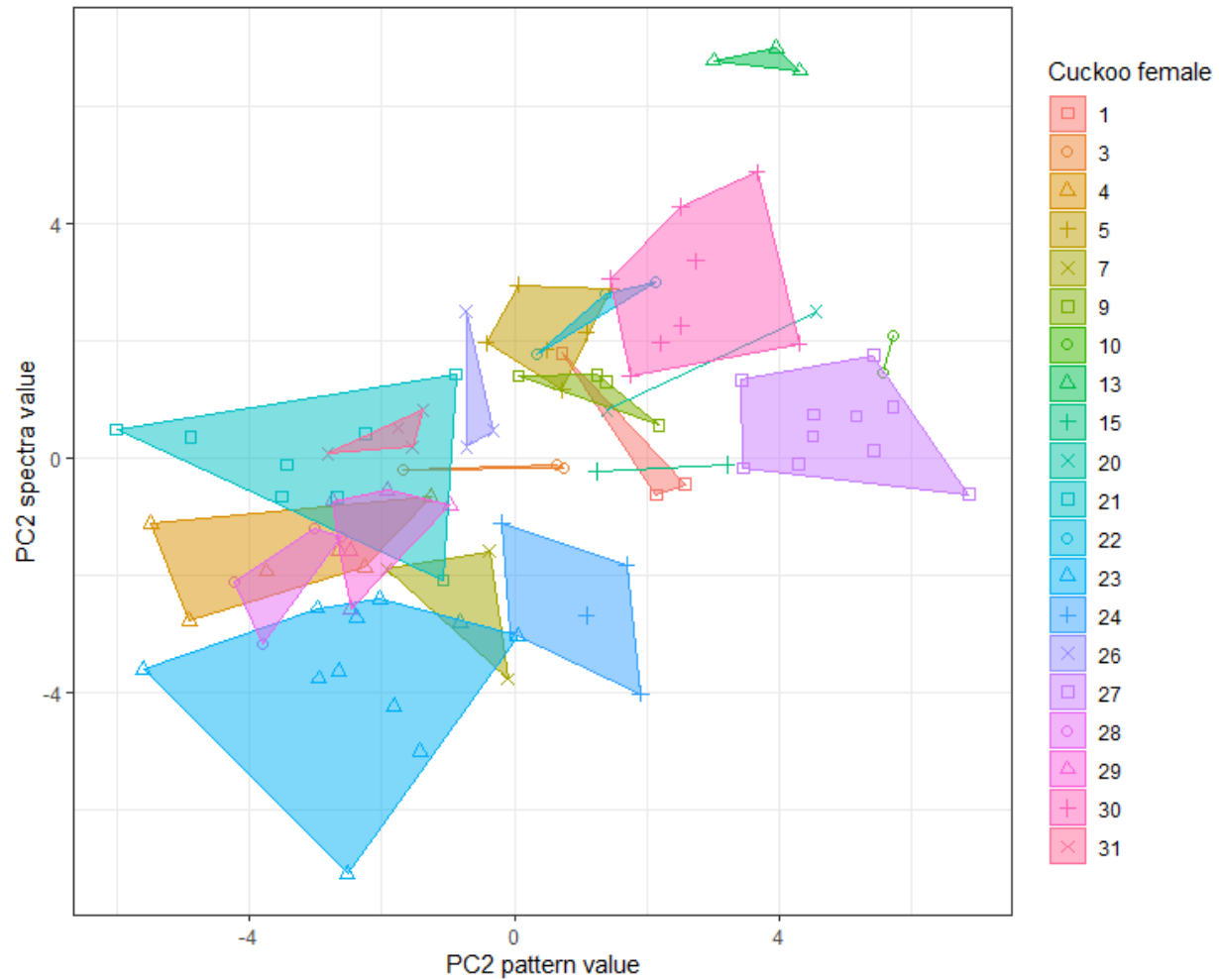
357

358

359

360

361



362

363 *Figure 1 Values for individual eggs on the two most important PC variables (according to the*
364 *random forest model), grouped by cuckoo female. PCA2 pattern variable indicates egg skew and*
365 *PC2 spectra variable indicates blueness/greenness of eggs (for details, see Table 2).*

366

367

368 *Human assessment*

369 Participants with some experience of working with biological data performed better at the grouping
370 task than those with no experience, though there is no clear evidence that specific experience of
371 working with cuckoo eggs is beneficial (Table 1).

372 *Table 1. Cluster similarities of egg sorting performed by humans both without knowledge (when*
373 *they did not know the number of females) and with a known number of females.*

Group	No knowledge	Known number of females
No experience (n = 5)	0.225 (0.066)	0.241 (0.041)
Non-specific experience (n = 3)	0.502 (0.170)	0.496 (0.057)
Specific experience (n = 4)	0.417 (0.050)	0.456 (0.158)

374 *Mean cluster similarity (and SD in brackets) is presented for each category.*

375

376 *Unsupervised learning*

377 Clustering using unsupervised hierarchical learning gave a cluster similarity value of 0.452; similar
378 to that of experienced human observers, but better than inexperienced observers.

379

380 *Supervised learning (random forest analysis)*

381 *Female clustering*

382 Clustering using supervised random forest analysis (with a leave-one-out protocol) led to good
383 classification, with a mean of 77.08/95 (81.1%) of eggs correctly assigned to their genetic parent.
384 The cluster similarity had a mean of 0.61 (SD = 0.03), higher than both experienced human
385 assessment and unsupervised learning.

386

387 We assessed variable importance (Table 2) using a full model including all data. PC2 for pattern
388 was the most important variable for classification, and the variables loading onto this PC were
389 predominantly those for the 'skew' of the pattern. PC2 for spectra was also important, with this
390 variable being influenced by the 'blueness/greenness' of the egg. Finally, PC3 and PC1 for shape
391 were also informative. The variables loading onto these PCs were the length, width, volume and
392 surface area of the egg.

393

394 *Table 2 The importance of individual variables in egg clustering using random forest analysis.*

Variable	Mean decrease in accuracy	Main PCA loadings
PC2_pattern	28.42	Skew values at pattern energy scales 1, 0.707, 0.5, 0.3536, 0.25, 0.1768, 0.125, 0.08839, 0.0625, 0.04419
PC2_spectra	26.8	426, 447, 468, 489, 510, 531nm
PC3_shape	23.81	Length, max width
PC1_shape	21.37	Length, max width, volume, surface area
PC1_spectra	19.79	342, 552, 573, 594, 636, 678, 699nm
UV_shape	19.36	-
PC2_shape	16.91	Ellipse deviation
PC1_luminance	15.42	Luminance sections 1, 2 and 3, standard deviation sections 1, 2 and 3
PC3_pattern	15.18	Pattern energy scales 1, 0.7071, 0.5, 0.3536, 0.04419, 0.03125, 0.0221
Brightness	12.9	-
PC1_pattern	11.23	Pattern energy scales 0.3536, 0.25, 0.1768, 0.125, 0.08839, 0.0625, total pattern energy, total pattern energy in segment 2

395 *Variables with larger mean decrease in accuracy are more important for classifying the data*
 396 *(mean decrease in accuracy is a measure of how much the accuracy of the random forest*
 397 *decreases due to the exclusion/permutation of a single variable). The main PCA loadings are*
 398 *those that were greater than +/- 0.25.*

399

400

401 *Same/different analysis*

402 For our labelled data, on average, 77.46/95 (81.5%) of eggs were matched with themselves.

403 During the training phase, the model was always trained with 'same' pairs that consisted of two

404 different eggs from the same female, so this was done as a test of whether the model was able

405 to recognise two truly identical eggs as coming from the same female.

406 In addition, on average 73.50/95 (77.4%) eggs were uniquely matched with other eggs laid by the

407 same female.

408 Out of 87 unlabelled eggs, the model was able to reliably (on 95% of runs) identify 22 as belonging

409 to a labelled female (9 eggs assigned to female 5, 5 eggs to female 27, 3 eggs to female 13, and

410 1 egg to each of females 4, 21, 28, 29, 30).

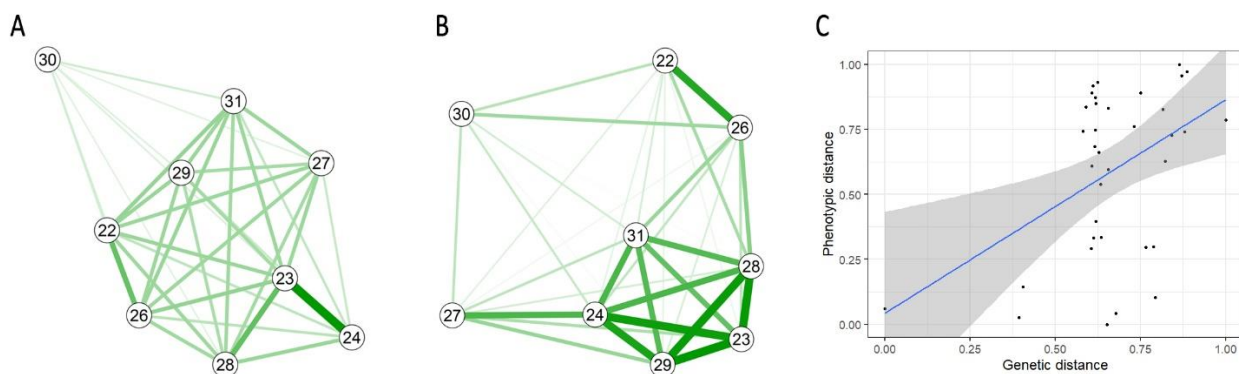
411

412 *Phenotypic-genotypic similarity*

413 There was no significant relationship between overall female egg phenotype distance and female

414 genetic distance (Mantel test $r = 0.1968$, $p = 0.098$, Fig. 2).

415



416

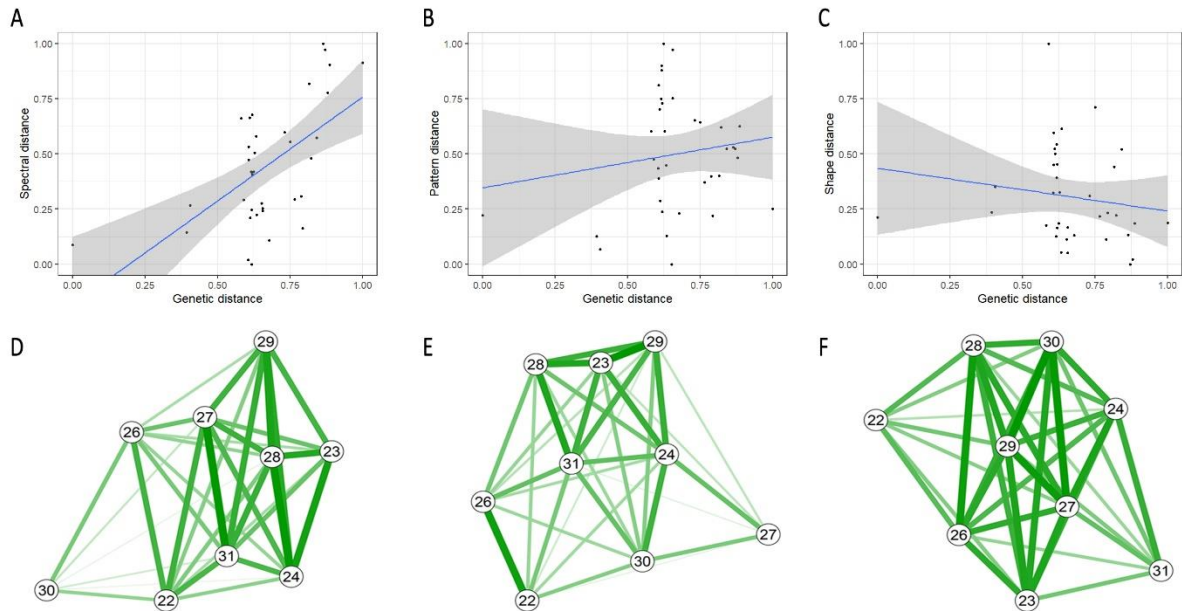
417 *Figure 2 Phenotypic distances of nine average eggs laid by nine genotyped common cuckoo*

418 *females (A) and their genetic distances (B). Thicker green lines denote higher phenotypic and*

419 *genetic similarity. Correlation between phenotypic and genetic distances (C).*

420

421 When considering each aspect of phenotypic similarity separately, both pattern/luminance and
422 shape distance metrics did not correlate with genotypic similarity ($r = 0.0254$, $p = 0.3861$ and $r =$
423 -0.2317 , $p = 0.9256$ respectively, Fig. 3A-D). However, spectral similarity did correlate with genetic
424 similarity ($r = 0.356$, $p = 0.037$, Fig. 3E-F).



426
427 *Figure 3 Correlation between spectral (A), pattern/lumiance (B) and shape (C) distances,*
428 *respectively and genetic distances. Individual phenotypic distances of average eggs laid by nine*
429 *genotyped common cuckoo females: spectral (D), pattern/luminance (E) and shape (F) distances.*

430

431

432 Discussion

433 The results of our study support the ‘constant egg-type hypothesis’ predicting that individual
434 cuckoo females lay eggs with a relatively constant appearance (Latter 1902, Chance 1940, Baker
435 1942, Lack 1968, Wyllie 1981, Moksnes et al. 2008). We confirmed that the within-clutch variation
436 of cuckoo egg appearance is significantly lower than between-clutch variation. This has also been

437 observed in other bird species and several adaptive explanations have been proposed for this
438 phenomenon, such as easier recognition of the parasitic egg by hosts (Øien et al., 1995),
439 recognition of an individual's own clutch in colonially-breeding birds (Hauber et al., 2019) or
440 signalling the female quality (Moreno & Osorno, 2003). Moreover, it has been shown that diet has
441 only a small effect on eggshell colour and that this trait is more affected by maternal identity,
442 suggesting that egg colour may carry information about intrinsic properties of the female
443 (Dearborn, Hanley, Ballantine, Cullum, & Reeder, 2012). Therefore, it seems that there is the
444 potential to use egg appearance to identify individual bird females.

445

446 Here, we demonstrated that the unsupervised computer-vision based classifier can outperform
447 human participants (especially inexperienced ones) in assigning cuckoo eggs to correct female
448 clusters. Human egg classifiers with experience of handling natural eggs (either cuckoo or of other
449 species) were able to more accurately sort them compared to people without this experience
450 (Table 1). This may indicate that experienced observers have some knowledge of which cues are
451 likely to be useful. They may also be more likely to be practiced at this type of fine discrimination
452 task, or be more motivated to take part, given their interest in biology. Interestingly, knowing the
453 number of groups (here: cuckoo females that laid eggs) did not increase sorting accuracy
454 substantially.

455

456 The automatic hierarchical clustering method showed very similar results to experienced human
457 classifiers, while supervised random forest analysis showed considerably better results and 81%
458 of cuckoo eggs were assigned correctly. This suggests that in some cases, automatic assessment
459 methods may be preferable to human assessment. Detailed consideration of the clusters created
460 by humans and the automatic methods showed that the same females were problematic for both
461 clustering methods, probably reflecting phenotypic overlap between some individuals (all sorting
462 results can be found in Supplementary Material). Our results showed that one of the pattern

463 characteristics (skew), blueness of colour and finally egg size were the most important parameters
464 that helped the automatic method to cluster eggs the most accurately. The slight improvement in
465 clustering accuracy for the automatic methods over human assessment may reflect the use of
466 features that humans are not able to see (e.g. the UV peak of the spectra). Therefore, we would
467 recommend automatic assessment over human assessment where possible.

468
469 A previous study suggested that closely related cuckoo females may lay eggs that are
470 indistinguishable from each other (Moksnes et al., 2008). Our limited dataset of nine caught
471 females for which we calculated genetic distances partially supports this idea as we found that
472 the background colour of eggs was more similar between closely related females. However,
473 pattern and dimension distances between females did not correlate with their genetic distances.
474 As an interesting example, we caught two closely related females, mother (female 23) and
475 daughter (female 24), and we photographed and measured the colour of their eggs. From the
476 photographs, we can see that their eggs look very similar in the colour and spotting pattern;
477 however, they differ considerably in size (Fig 3 in Supplementary material). Similarly, we can also
478 see the resemblance in egg colour and pattern (but not in size) between other closely related
479 females (females 23 vs 28 and females 22 vs 26; Fig. 3 and 4 in Supplementary material).
480 Previous studies showed that all investigated egg features - colour, spotting pattern and also size
481 have high heritability (Wei et al., 1992; Collias, 1993; Gosler et al., 2000; Christians, 2002;
482 Morales et al., 2010). Our results indicate that the background colour might be a more heritable
483 trait than spotting pattern and egg size or shape, which supports the idea that egg colour is of
484 vital importance for egg recognition (Avilés et al., 2010; Spottiswoode & Stevens, 2010; Michal
485 Šulc, Procházka, Capek, & Honza, 2016). Since several studies reported that hosts and even
486 parasites themselves also use spotting pattern (López-de-Hierro & Moreno-Rueda, 2010;
487 Spottiswoode & Stevens, 2010; de la Colina, Pompilio, Hauber, Reboreda, & Mahler, 2012) or
488 egg size (Marchetti, 2000; Spottiswoode, 2013) when recognizing and eliminating parasitic eggs,

489 we still expect relatively high heritability of these egg traits in brood parasites. We suspect that
490 the insignificant relationship between genetic distance and phenotypic distance in spotting pattern
491 and size is only a matter of our limited sample size. A Larger sample size, including more mother-
492 daughter pairs, is needed to truly estimate heritability values of individual egg traits (de
493 Villemereuil, Gimenez, & Doligez, 2013). The lack of significant correlation between egg shape
494 and genetic similarity may also be explained by the fact that egg size often reflects the size of
495 laying females (Larsson & Forslund, 1992; Nager & Zandt, 1994), which depends on the genetic
496 contribution of both parents and therefore might differ more even in closely related females.
497 Moreover, even within one host population, cuckoos are raised by host parents that vary in their
498 provisioning care (Požgayová, Piálková, Honza, & Procházka, 2018), which may also influence
499 the body size of cuckoo females in adulthood. Finally, some studies showed there is a positive
500 relationship between food availability and egg size (reviewed in Christians 2002). Consequently,
501 since egg size and shape may even differ in such closely related females, these traits may be
502 used for the separation of their eggs. And indeed, this is what we observed in several of our
503 human participants and also in computer analysis where separation of mother and daughter eggs
504 was relatively straightforward presumably because of the apparent difference in the size and
505 shape (see Supplementary material).

506

507 Although our results are in concordance with a previous study showing that the visual appearance
508 of cuckoo eggs cannot be used to assign them reliably to individual females without genetic data
509 (Moksnes et al. 2008), here we present a new approach that uses both genetic and phenotypic
510 information. We used this method for assigning cuckoo eggs for which we did not have genetic
511 data (because they were ejected by hosts or predated), allowing us to assign 22 eggs (out of 87)
512 to eight known females. This method seemed to work well for females that laid very distinctive
513 eggs and therefore results will strongly depend on within- and between-clutch variation. We may
514 expect better results of the method in species where between-clutch variation substantially

515 exceeds the within-clutch variation. It must also be noted that the accuracy of the assignment will
516 increase with the relative number of (genetically and phenotypically) analysed samples in the
517 study area that are used for the training dataset, because broad sampling will reduce the chance
518 that an unsampled egg that has been laid by a completely new female will be assigned to an
519 existing (incorrect) female. Finally, it is important to apply other information (laying date and laying
520 area) to eliminate potential incorrect assignments. However, in our dataset, we did not find any
521 such discrepancies for the 22 eggs that were automatically assigned based on their phenotypes.
522

523 We conclude that egg appearance alone cannot be used to identify individual cuckoo females.
524 Clusters created either by people using egg photographs or by the computer using spectral and
525 image data did not fully correspond to the true female identity acquired from molecular analyses,
526 though the supervised automatic assessment was the most accurate for classification. Our results
527 support the idea that more closely related females lay eggs more similar in their colour. However,
528 the size and shape of the eggs seems to be the least heritable trait, which may substantially help
529 to distinguish even between eggs of closely related cuckoo females. We advocate genetic
530 analysis to be used for determination of maternity in this species. However, in sufficiently sampled
531 systems, supervised analytical methods that use egg visual features might additionally help to
532 broaden sample sizes, which may be very desirable for studying biological questions (see e.g.
533 Koleček et al. 2020 in prep). We encourage researchers investigating inter- and intra-specific
534 brood parasitism to use this low-cost and ethically more appropriate method of individual
535 identification. Since a similar technique has been successfully used in non-parasitic species
536 (Höltje et al., 2016), identification of laying females using egg appearance therefore has the
537 potential to be of widespread use.

538

539

540

541 **Acknowledgements**

542 We would like to thank Lisandrina Mari, Kamila Bendová, Kristýna Míčková, Jana Hodanová,
543 Klára Stehlíková, Martina Fridrichová, Marta Potůčková, David Tesař, Václav Jelínek and Petr
544 Potůček for their sorting of cuckoo eggs. Lisandrina Mari also substantially helped with writing of
545 the genetic part of methods. We also thank Lukáš Kulísek, Boris Prudík, Kateřina Sosnovcová,
546 Jaroslav Koleček and Miroslav Čapek for help with fieldwork. We thank Micha Elsner for his help
547 with data analyses. We are also very grateful to Vladimír Beneš and the European Molecular
548 Biology Laboratory Genomic Core Facility in Heidelberg (Germany) for their kind advice and
549 technical support regarding Illumina sequencing. The managers of the Hodonín Fish Farm kindly
550 permitted us conducting the fieldwork on their grounds.

551

552 **Ethical note**

553 This study was carried out with the permission of the regional nature conservation authorities
554 (permit numbers JMK: 115874/2013 and 38506/2016; MUHOCJ: 41433/2012/OŽP,
555 34437/2014/OŽP, and 14306/2016/OŽP). The fieldwork adhered to the animal care protocol
556 (experimental project numbers 039/2011 AV ČR and 3030/ENV/17-169/630/17) and to the Czech
557 Law on the Protection of Animals against Mistreatment (license numbers CZ 01272 and CZ
558 01284). This study was carried out with the permission of the regional nature conservation
559 authorities (permit numbers JMK: 115874/2013 and 38506/2016; MUHOCJ: 41433/2012/OŽP,
560 34437/2014/OŽP, and 14306/2016/OŽP).

561

562 **Funding**

563 This work was supported by the Czech Science Foundation (grant number 17-12262S) and by
564 the Programme for research and mobility of young researchers of the Czech Academy of
565 Sciences (MSM200931801, awarded to M.Š.).

566

567 **References**

- 568 Ådahl, E., Lindström, J., Ruxton, G. D., Arnold, E. K., & Begg, T. (2004). Can intraspecific brood
569 parasitism be detected using egg morphology only? *Journal of Avian Biology*, 35(4),
570 360–364. doi:10.1111/j.0908-8857.2004.03224.x
- 571 Avilés, J. M., Vikan, J. R., Fossøy, F., Antonov, A., Moksnes, A., Røskaft, E., & Stokke, B. G.
572 (2010). Avian colour perception predicts behavioural responses to experimental brood
573 parasitism in chaffinches. *Journal of Evolutionary Biology*, 23(2), 293–301.
574 doi:10.1111/j.1420-9101.2009.01898.x
- 575 Bensch, S. (1996). Female Mating Status and Reproductive Success in the Great Reed
576 Warbler: Is there a Potential Cost of Polygyny that Requires Compensation. *Journal of*
577 *Animal Ecology*, 65(3), 283–296. doi:10.2307/5875
- 578 Berg, C. P. van den, Troscianko, J., Endler, J. A., Marshall, N. J., & Cheney, K. L. (2020).
579 Quantitative Colour Pattern Analysis (QCPA): A comprehensive framework for the
580 analysis of colour patterns in nature. *Methods in Ecology and Evolution*, 11(2), 316–332.
581 doi:10.1111/2041-210X.13328
- 582 Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., ... Kirschel, A. N.
583 G. (2011). Acoustic monitoring in terrestrial environments using microphone arrays:
584 applications, technological considerations and prospectus. *Journal of Applied Ecology*,
585 48(3), 758–767. doi:10.1111/j.1365-2664.2011.01993.x
- 586 Bolger, D. T., Morrison, T. A., Vance, B., Lee, D., & Farid, H. (2012). A computer-assisted
587 system for photographic mark–recapture analysis. *Methods in Ecology and Evolution*,
588 3(5), 813–822. doi:10.1111/j.2041-210X.2012.00212.x
- 589 Brown, C. R., & Sherman, L. C. (1989). Variation in the appearance of swallow eggs and the
590 detection of intraspecific brood parasitism. *The Condor*, 91(3), 620–627.
- 591 Cariello, M. O., Lima, M. R., Schwabl, H. G., & Macedo, R. H. (2004). Egg characteristics are
592 unreliable in determining maternity in communal clutches of guira cuckoos *Guira guira*.

- 593 *Journal of Avian Biology*, 35(2), 117–124. doi:10.1111/j.0908-8857.2004.03146.x
- 594 Chance, E. P. (1922). *The cuckoo's secret* (pp. 1–290). London : Sidgwick & Jackson, .
- 595 doi:10.5962/bhl.title.161352
- 596 Christians, J. K. (2002). Avian egg size: variation within species and inflexibility within
- 597 individuals. *Biological Reviews*, 77(1), 1–26.
- 598 doi:<https://doi.org/10.1017/S1464793101005784>
- 599 Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology.
- 600 *Methods in Ecology and Evolution*, 10(10), 1632–1644. doi:10.1111/2041-210X.13256
- 601 Collias, E. C. (1993). Inheritance of egg-color polymorphism in the village weaver (*Ploceus*
- 602 *cucullatus*). *The Auk*, 110(4), 683–692.
- 603 Crall, J. P., Stewart, C. V., Berger-Wolf, T. Y., Rubenstein, D. I., & Sundaresan, S. R. (2013).
- 604 HotSpotter — Patterned species instance recognition. In *2013 IEEE Workshop on*
- 605 *Applications of Computer Vision (WACV)* (pp. 230–237).
- 606 doi:10.1109/WACV.2013.6475023
- 607 Davies, N. B. (2010). *Cuckoos, Cowbirds and Other Cheats*. A&C Black.
- 608 de la Colina, M. A., Pompilio, L., Hauber, M. E., Reboreda, J. C., & Mahler, B. (2012). Different
- 609 recognition cues reveal the decision rules used for egg rejection by hosts of a variably
- 610 mimetic avian brood parasite. *Animal Cognition*, 15(5), 881–889.
- 611 de Villemereuil, P., Gimenez, O., & Doligez, B. (2013). Comparing parent–offspring regression
- 612 with frequentist and Bayesian animal models to estimate heritability in wild populations:
- 613 a simulation study for Gaussian and binary traits. *Methods in Ecology and Evolution*,
- 614 4(3), 260–275.
- 615 Dearborn, D. C., Hanley, D., Ballantine, K., Cullum, J., & Reeder, D. M. (2012). Eggshell colour
- 616 is more strongly affected by maternal identity than by dietary antioxidants in a captive
- 617 poultry system. *Functional Ecology*, 26(4), 912–920. doi:[https://doi.org/10.1111/j.1365-](https://doi.org/10.1111/j.1365-2435.2012.02001.x)
- 618 [2435.2012.02001.x](https://doi.org/10.1111/j.1365-2435.2012.02001.x)

- 619 Deb, D., Wiper, S., Gong, S., Shi, Y., Tymoszek, C., Fletcher, A., & Jain, A. K. (2018). Face
620 Recognition: Primates in the Wild. In *2018 IEEE 9th International Conference on*
621 *Biometrics Theory, Applications and Systems (BTAS)* (pp. 1–10).
622 doi:10.1109/BTAS.2018.8698538
- 623 Eadie, J. M., Smith, J. N. M., Zadworny, D., Kühnlein, U., & Cheng, K. (2010). Probing
624 parentage in parasitic birds: an evaluation of methods to detect conspecific brood
625 parasitism using goldeneyes *Bucephala islandica* and *Bl. clangula* as a test case.
626 *Journal of Avian Biology*, *41*(2), 163–176. doi:10.1111/j.1600-048X.2009.04735.x
- 627 Ferreira, A. C., Silva, L. R., Renna, F., Brandl, H. B., Renoult, J. P., Farine, D. R., ... Doutrelant,
628 C. (2020). Deep learning-based methods for individual recognition in small birds.
629 *Methods in Ecology and Evolution*, *11*(9), 1072–1085. doi:10.1111/2041-210X.13436
- 630 Fossøy, F., Antonov, A., Moksnes, A., Røskaft, E., Vikan, J. R., Møller, A. P., ... Stokke, B. G.
631 (2011). Genetic differentiation among sympatric cuckoo host races: males matter.
632 *Proceedings of the Royal Society B: Biological Sciences*, *278*(1712), 1639–1645.
633 doi:10.1098/rspb.2010.2090
- 634 Fossøy, F., Moksnes, A., Røskaft, E., Antonov, A., Dyrzcz, A., Moskat, C., ... Stokke, B. G.
635 (2012). Sex Allocation in Relation to Host Races in the Brood-Parasitic Common Cuckoo
636 (*Cuculus canorus*). *PLOS ONE*, *7*(5), e36884. doi:10.1371/journal.pone.0036884
- 637 Gibbons, D. W. (1986). Brood parasitism and cooperative nesting in the moorhen, *Gallinula*
638 *chloropus*. *Behavioral Ecology and Sociobiology*, *19*(3), 221–232.
- 639 Gibbs, H. L., Sorenson, M. D., Marchetti, K., de L. Brooke, M., Davies, N. B., & Nakamura, H.
640 (2000). Genetic evidence for female host-specific races of the common cuckoo. *Nature*,
641 *407*(6801), 183–186. doi:10.1038/35025058
- 642 Gloag, R., Fiorini, V. D., Reboreda, J. C., & Kacelnik, A. (2013). The wages of violence:
643 mobbing by mockingbirds as a frontline defence against brood-parasitic cowbirds.
644 *Animal Behaviour*, *86*(5), 1023–1029. doi:10.1016/j.anbehav.2013.09.007

- 645 Gosler, A. G., Barnett, P. R., & James Reynolds, S. (2000). Inheritance and variation in eggshell
646 patterning in the great tit *Parus major*. *Proceedings of the Royal Society of London.*
647 *Series B: Biological Sciences*, 267(1461), 2469–2473.
- 648 Griffith, S. C., Barr, I., Sheldon, B. C., Rowe, L. V., & Burke, T. (2009). Egg patterning is not a
649 reliable indicator of intraspecific brood parasitism in the blue tit *Cyanistes caeruleus*.
650 *Journal of Avian Biology*, 40(3), 337–341. doi:10.1111/j.1600-048X.2009.04671.x
- 651 Grønstøl, G., Blomqvist, D., & Wagner, R. H. (2006). The Importance of Genetic Evidence for
652 Identifying Intra-Specific Brood Parasitism. *Journal of Avian Biology*, 37(2), 197–199.
- 653 Hanley, D., Šulc, M., Brennan, P. L., Hauber, M. E., Grim, T., & Honza, M. (2016). Dynamic egg
654 color mimicry. *Ecology and Evolution*, 6(12), 4192–4202.
- 655 Hansen, M. F., Smith, M. L., Smith, L. N., Salter, M. G., Baxter, E. M., Farish, M., & Grieve, B.
656 (2018). Towards on-farm pig face recognition using convolutional neural networks.
657 *Computers in Industry*, 98, 145–152. doi:10.1016/j.compind.2018.02.016
- 658 Hauber, M. E., Luro, A., McCarty, C. J., Barateli, K., Cassey, P., Hansen, E. S., & Dale, J.
659 (2019). Interannual repeatability of eggshell phenotype in individual female Common
660 Murres (*Uria aalge*). *Canadian Journal of Zoology*, 97(4), 385–391. doi:10.1139/cjz-
661 2018-0172
- 662 Hiby, L., Lovell, P., Patil, N., Kumar, N. S., Gopalaswamy, A. M., & Karanth, K. U. (2009). A
663 tiger cannot change its stripes: using a three-dimensional model to match images of
664 living tigers and tiger skins. *Biology Letters*, 5(3), 383–386. doi:10.1098/rsbl.2009.0028
- 665 Hill, G., & McGraw, K. (2006). *Bird Coloration, Volume I: Mechanisms and Measurements*.
- 666 Höltje, H., Mewes, W., Haase, M., & Ornés, A. S. (2016). Genetic evidence of female specific
667 eggshell colouration in the Common Crane (*Grus grus*). *Journal of Ornithology*, 157(2),
668 609–617. doi:10.1007/s10336-015-1311-5
- 669 Honza, M., Vošlajerová, K., & Moskát, C. (2007). Eviction behaviour of the common cuckoo
670 *Cuculus canorus* chicks. *Journal of Avian Biology*, 38(3), 385–389.

- 671 doi:10.1111/j.2007.0908-8857.03901.x
- 672 Hou, J., He, Y., Yang, H., Connor, T., Gao, J., Wang, Y., ... Zhou, S. (2020). Identification of
673 animal individuals using deep learning: A case study of giant panda. *Biological*
674 *Conservation*, 242, 108414. doi:10.1016/j.biocon.2020.108414
- 675 Jackson, W. M. (1992). Estimating Conspecific Nest Parasitism in the Northern Masked Weaver
676 Based on Within-Female Variability in Egg Appearance. *The Auk*, 109(3), 435–443.
677 doi:10.1093/auk/109.3.435
- 678 Jelínek, V., Šulc, M., Štetková, G., & Honza, M. (2020). Fast and furious: host aggression
679 modulates behaviour of brood parasites. *BioRxiv*, 2020.04.27.063933.
680 doi:10.1101/2020.04.27.063933
- 681 Jones, O. R., & Wang, J. (2010). COLONY: a program for parentage and sibship inference from
682 multilocus genotype data. *Molecular Ecology Resources*, 10(3), 551–555.
683 doi:10.1111/j.1755-0998.2009.02787.x
- 684 Jung, T. S., Boonstra, R., & Krebs, C. J. (2020). Mark my words: experts' choice of marking
685 methods used in capture-mark-recapture studies of small mammals. *Journal of*
686 *Mammalogy*, 101(1), 307–317.
- 687 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... Drummond, A.
688 (2012). Geneious Basic: an integrated and extendable desktop software platform for the
689 organization and analysis of sequence data. *Bioinformatics (Oxford, England)*, 28(12),
690 1647–1649. doi:10.1093/bioinformatics/bts199
- 691 Krause, J., Krause, S., Arlinghaus, R., Psorakis, I., Roberts, S., & Rutz, C. (2013). Reality
692 mining of animal social systems. *Trends in Ecology & Evolution*, 28(9), 541–551.
693 doi:10.1016/j.tree.2013.06.002
- 694 Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*
695 *Methods*, 9(4), 357–359. doi:10.1038/nmeth.1923
- 696 Larsson, K., & Forslund, P. (1992). Genetic and social inheritance of body and egg size in the

- 697 barnacle goose (*Branta leucopsis*). *Evolution*, 46(1), 235–244.
- 698 Lemons, P. R., Sedinger, J. S., & Randle, P. S. (2011). Detecting conspecific brood parasitism
699 using egg morphology in black brant *Branta bernicla nigricans*. *Journal of Avian Biology*,
700 42(4), 282–288. doi:10.1111/j.1600-048X.2011.05217.x
- 701 Lindberg, M. S. (2012). A review of designs for capture–mark–recapture studies in discrete time.
702 *Journal of Ornithology*, 152(2), 355–370.
- 703 López-de-Hierro, M. D. G., & Moreno-Rueda, G. (2010). Egg-spot pattern rather than egg colour
704 affects conspecific egg rejection in the house sparrow (*Passer domesticus*). *Behavioral*
705 *Ecology and Sociobiology*, 64(3), 317–324. doi:10.1007/s00265-009-0811-9
- 706 Lyon, B. E. (2003). Egg recognition and counting reduce costs of avian conspecific brood
707 parasitism. *Nature*, 422(6931), 495–499.
- 708 Lyon, Bruce E. (1993). Tactics of parasitic American coots: host choice and the pattern of egg
709 dispersion among host nests. *Behavioral Ecology and Sociobiology*, 33(2), 87–100.
- 710 Lyon, Bruce E., & Eadie, J. M. (2008). Conspecific brood parasitism in birds: a life-history
711 perspective. *Annual Review of Ecology, Evolution, and Systematics*, 39, 343–363.
- 712 Maia, R., Gruson, H., Endler, J. A., & White, T. E. (2019). pavo 2: New tools for the spectral and
713 spatial analysis of colour in r. *Methods in Ecology and Evolution*, 10(7), 1097–1107.
714 doi:10.1111/2041-210X.13174
- 715 Marchetti, K. (2000). Egg rejection in a passerine bird: size does matter. *Animal Behaviour*,
716 59(4), 877–883.
- 717 McMaster, D. G., Neudorf, D. L. H., Sealy, S. G., & Pitcher, T. E. (2004). A comparative analysis
718 of laying times in passerine birds. *Journal of Field Ornithology*, 75(2), 113–122.
719 doi:10.1648/0273-8570-75.2.113
- 720 McRae, S. B., & Burke, T. (1996). Intraspecific brood parasitism in the moorhen: parentage and
721 parasite-host relationships determined by DNA fingerprinting. *Behavioral Ecology and*
722 *Sociobiology*, 38(2), 115–129. doi:10.1007/s002650050224

- 723 Moksnes, A., Røskaft, E., Rudolfson, G., Skjelseth, S., Stokke, B. G., Kleven, O., ... Taborsky,
724 M. (2008). Individual female common cuckoos *Cuculus canorus* lay constant egg types
725 but egg appearance cannot be used to assign eggs to females. *Journal of Avian Biology*,
726 39(2), 238–241. doi:10.1111/j.2008.0908-8857.04158.x
- 727 Møller, A. P. (1987). Intraspecific nest parasitism and anti-parasite behaviour in swallows,
728 *Hirundo rustica*. *Animal Behaviour*, 35(1), 247–254. doi:10.1016/S0003-3472(87)80230-
729 0
- 730 Morales, J., Kim, S.-Y., Lobato, E., Merino, S., Tomás, G., MARTÍNEZ-de la PUENTE, J., &
731 Moreno, J. (2010). On the heritability of blue-green eggshell coloration. *Journal of*
732 *Evolutionary Biology*, 23(8), 1783–1791.
- 733 Moreno, J., & Osorno, J. L. (2003). Avian egg colour and sexual selection: does eggshell
734 pigmentation reflect female condition and genetic quality? *Ecology Letters*, 6(9), 803–
735 806.
- 736 Nager, R. G., & Zandt, H. S. (1994). Variation in egg size in great tits. *ARDEA-WAGENINGEN*,
737 82, 315–315.
- 738 Øien, I. J., Moksnes, A., & Røskaft, E. (1995). Evolution of variation in egg color and marking
739 pattern in European passerines: adaptations in a coevolutionary arms race with the
740 cuckoo, *Cuculus canorus*. *Behavioral Ecology*, 6(2), 166–174.
741 doi:10.1093/beheco/6.2.166
- 742 Paillisson, J.-M., Latraube, F., Marion, L., & Bretagnolle, V. (2008). Indirect evidence of
743 conspecific nest parasitism in the colonial whiskered tern (*Chlidonias hybrida*). *Comptes*
744 *Rendus Biologies*, 331(7), 559–567. doi:10.1016/j.crv.2008.04.010
- 745 Payne, R. B. (1973). Individual Laying Histories and the Clutch Size and Numbers of Eggs of
746 Parasitic Cuckoos. *The Condor*, 75(4), 414–438. doi:10.2307/1366563
- 747 Petersen, M. R. (1992). Intraspecific Variation in Egg Shape among Individual Emperor Geese
748 (Variación Intraespecífica en la Forma de los Huevos de *Chen canagicus*). *Journal of*

- 749 *Field Ornithology*, 63(3), 344–354.
- 750 Petrusková, T., Pišvejcová, I., Kinštová, A., Brinke, T., & Petrušek, A. (2016). Repertoire-based
751 individual acoustic monitoring of a migratory passerine bird with complex song as an
752 efficient tool for tracking territorial dynamics and annual return rates. *Methods in Ecology
753 and Evolution*, 7(3), 274–284. doi:10.1111/2041-210X.12496
- 754 Petrželková, A., Pöysä, H., Klvaňa, P., Albrecht, T., & Hořák, D. (2017). Egg morphology fails to
755 identify nests parasitized by conspecifics in common pochard: a test based on protein
756 fingerprinting and including female relatedness. *Journal of Avian Biology*, 48(2), 229–
757 234. doi:<https://doi.org/10.1111/jav.00865>
- 758 Piálek, L., Burrell, E., Dragová, K., Almirón, A., Casciotta, J., & Říčan, O. (2019).
759 Phylogenomics of pike cichlids (Cichlidae: Crenicichla) of the *C. mandelburgeri* species
760 complex: rapid ecological speciation in the Iguazú River and high endemism in the
761 Middle Paraná basin. *Hydrobiologia*, 832(1), 355–375. doi:10.1007/s10750-018-3733-6
- 762 Pöysä, H., Lindblom, K., Rutila, J., & Sorjonen, J. (2009). Reliability of egg morphology to detect
763 conspecific brood parasitism in goldeneyes *Bucephala clangula* examined using protein
764 fingerprinting. *Journal of Avian Biology*, 40(4), 453–456. doi:10.1111/j.1600-
765 048X.2008.04528.x
- 766 Požgayová, M., Piálková, R., Honza, M., & Procházka, P. (2018). Sex-specific nestling growth in
767 an obligate brood parasite: Common Cuckoo males grow larger than females. *The Auk:
768 Ornithological Advances*, 135(4), 1033–1042.
- 769 Ptacek, L., Machlica, L., Linhart, P., Jaska, P., & Muller, L. (2016). Automatic recognition of bird
770 individuals on an open set using as-is recordings. *Bioacoustics*, 25(1), 55–73.
771 doi:10.1080/09524622.2015.1089524
- 772 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P.
773 C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based
774 Linkage Analyses. *American Journal of Human Genetics*, 81(3), 559–575.

- 775 Ramey, J. A. (2012). *clusteval: Evaluation of Clustering Algorithms*. Retrieved from
776 <https://CRAN.R-project.org/package=clusteval>
- 777 Roy, C. L., Parker, P. G., & Gates, R. J. (2009). Egg Morphology is an Unreliable Indicator of
778 Intraspecific Nest Parasitism in Wood Ducks. *The Condor*, *111*(2), 377–381.
779 doi:10.1525/cond.2009.080117
- 780 Soler, M. (Ed.). (2017). *Avian Brood Parasitism*. Cham, Switzerland: Springer International
781 Publishing.
- 782 Spottiswoode, C. N. (2013). A brood parasite selects for its own egg traits. *Biology Letters*, *9*(5),
783 20130573.
- 784 Spottiswoode, C. N., & Stevens, M. (2010). Visual modeling shows that avian host parents use
785 multiple visual cues in rejecting parasitic eggs. *Proceedings of the National Academy of*
786 *Sciences*, *107*(19), 8672–8676.
- 787 Stone, M. (1974). Cross-validators choice and assessment of statistical predictions. *Journal of*
788 *the Royal Statistical Society: Series B (Methodological)*, *36*(2), 111–133.
- 789 Šulc, M., Štětková, G., Jelínek, V., Czyż, B., Dyrz, A., Karpińska, O., ... Honza, M. (2020).
790 Killing behaviour of adult brood parasites. *Behaviour*, *157*(12–13), 1099–1111.
791 doi:10.1163/1568539X-bja10033
- 792 Šulc, Michal, Procházka, P., Capek, M., & Honza, M. (2016). Birds use eggshell UV reflectance
793 when recognizing non-mimetic parasitic eggs. *Behavioral Ecology*, *27*(2), 677–684.
- 794 Šulc, Michal, Troscianko, J., Štětková, G., Hughes, A. E., Jelínek, V., Capek, M., & Honza, M.
795 (2019). Mimicry cannot explain rejection type in a host–brood parasite system. *Animal*
796 *Behaviour*, *155*, 111–118. doi:10.1016/j.anbehav.2019.05.021
- 797 Troscianko, J. (2014). A simple tool for calculating egg shape, volume and surface area from
798 digital images. *Ibis*, *156*(4), 874–878.
- 799 Troscianko, J., & Stevens, M. (2015). Image calibration and analysis toolbox—a free software
800 suite for objectively measuring reflectance, colour and pattern. *Methods in Ecology and*

- 801 *Evolution*, 6(11), 1320–1331.
- 802 Wei, R., Bitgood, J. J., & Dentine, M. R. (1992). Inheritance of Tinted Eggshell Colors in White-
- 803 Shell Stocks1. *Poultry Science*, 71(3), 406–418. doi:10.3382/ps.0710406
- 804 Weinstein, B. G. (2018). A computer vision for animal ecology. *Journal of Animal Ecology*,
- 805 87(3), 533–545. doi:10.1111/1365-2656.12780
- 806 Wyllie, I. (1981). *The Cuckoo*. Universe Books.
- 807