Automatic identification of bird females using egg phenotype

Running title**: Automatic identification of bird females**

Michal Šulc[1,*], Anna E. Hughes[2], Jolyon Troscianko[3], Gabriela Štětková[1,4], Petr Procházka[1], Milica Požgayová[1], Lubomír Piálek[1,5], Radka Piálková[1,5], Vojtěch Brlík[1,6] and Marcel Honza[1]

[1] *Czech Academy of Sciences, Institute of Vertebrate Biology, Brno, Czech Republic*
[2] *Department of Psychology, University of Essex, Colchester, UK*
[3] *Centre for Life and Environmental Sciences, University of Exeter, Penryn, UK*
[4] *Department of Botany and Zoology, Faculty of Sciences, Masaryk University, Brno, Czech Republic*
[5] *Faculty of Natural Sciences, University of South Bohemia, České Budějovice, Czech Republic*
[6] *Department of Ecology, Faculty of Science, Charles University, Prague, Czech Republic*
**Correspondence author: sulc@ivb.cz*

ABSTRACT

Individual identification is crucial for studying animal ecology and evolution, and in birds, is often achieved by capturing and tagging. However, these methods are insufficient for identifying individuals/species that are secretive or difficult to catch. Here, we employ an automatic analytical approach to predict the identity of bird females based on the appearance of their eggs, using the common cuckoo (*Cuculus canorus*) as a model species. We analysed 192 cuckoo eggs using digital photography and spectrometry. Cuckoo females were identified from genetic sampling of nestlings, allowing us to determine the accuracy of automatic (unsupervised and supervised) and human assignment. Finally, we used a novel analytical approach to identify eggs that were not genetically analysed. Our results showed that individual cuckoo females lay eggs with a relatively constant appearance and that eggs laid by more genetically distant females differed more in colour. Unsupervised clustering had similar cluster accuracy to experienced human observers, but supervised methods were able to outperform humans. Our novel method reliably assigned a relatively high number of eggs without genetic data to their mothers. This is therefore a cost-

effective and minimally invasive method for increasing sample sizes which may facilitate research on brood parasites and other avian species.

Keywords: machine learning, individual assignment, spotting pattern, colour, genotyping, parental analysis, brood parasitism, common cuckoo

INTRODUCTION

Identification of individuals is important in animal ecology and biology research, particularly when investigating variation among or within individuals in a population. Traditionally, capture-mark-recapture techniques have been used to monitor individuals during their lifetime (Lindberg, 2012). This method has been extended by employing more sophisticated methods such as attaching GPS (global positioning system) and radio transmitters or RFID (radio frequency identification) tags (Krause *et al.*, 2013) that allow researchers to investigate the spatial-temporal activity of animals in more detail. However, these methods still require capturing and tagging which is usually time-consuming, expensive, and may reduce animal welfare (Weinstein, 2018). Therefore, there have been efforts to develop cost-effective indirect approaches to identify and monitor individuals within a species.

These indirect approaches rely on the fact that individuals differ from each other visually and/or acoustically and this variation may be used for their identification. Indeed, it has been shown that e.g. face (Hou *et al.*, 2020) and body pattern data (Ferreira *et al.*, 2020) captured from photographs may allow discrimination of individuals. Similarly, sounds produced by animals also seem to serve as a good individual fingerprint (Petrusková *et al.*, 2016; Stowell *et al.*, 2019). Recently, applying artificial intelligence techniques that automate the analysis of various types of

data such as pictures or audio recordings has made these methods reliable and applicable for various animal taxa (Christin, Hervet, & Lecomte, 2019).

However, for many species, identification of all individuals in a population is still not straightforward e.g. because it is difficult to catch them or due to their secretive behaviour. Here, we focus on one group of animals that are especially challenging to study – avian brood parasites. There are more than a hundred obligate brood parasites that never build their own nests and instead lay their eggs into nests of other species (Davies, 2010). Brood parasites and their hosts have been the focus of considerable research into co-evolutionary arms races (Soler, 2017) but since they only lay eggs and then usually do not return to host nests (but see Šulc *et al.*, 2020), and because egg laying is fast (Jelínek *et al.*, 2021), direct observation of parasitism in nature is difficult which makes identification of parasitic females problematic. As a consequence, many important aspects of females' life history strategy are still poorly understood (Soler, 2017).

Since it has been demonstrated in several bird species (including brood parasites) that individual females lay eggs with a relatively constant appearance compared to other females (e.g. Øien, Moksnes, & Røskaft, 1995; Höltje *et al.*, 2016), there is the potential to use egg appearance to identify individual females. This method has already been applied for the identification of parasitic eggs in conspecific brood parasites (e.g. Lyon, 2003). However, later studies that estimated accuracy of parasitic egg identification showed ambiguous results for some species and for others this method did not work at all (reviewed in Petrželková *et al.*, 2017). One of the reasons why many studies found low accuracy of identification might be that closely related females lay similar eggs. Indeed, it has been shown that egg appearance, namely egg colour (Morales *et al.*, 2010), spotting pattern (Gosler, Barnett, & James Reynolds, 2000) and egg size (Christians, 2002) are highly heritable traits which might complicate female identification especially in inbred populations. Another explanation might be that previous studies did not use the most informative measures of egg variability.

In this study, we focus on eggs of the common cuckoo (*Cuculus canorus*, hereafter cuckoo) because we still have little information about the breeding biology and evolution of individual host-specific races (Gibbs *et al.*, 2000; Fossøy *et al.*, 2011) in this brood parasite. Moreover, there has been recent population decline (Hewson *et al.*, 2016) and a low-cost and minimally invasive method of female identification would greatly facilitate conservation of this enigmatic species. Using egg appearance to identify cuckoo females has already been attempted, but was unsuccessful (Moksnes *et al.*, 2008). However, this study assessed cuckoo eggs from a human perspective, with people sorting the eggs based on photographs. To date, there have been no attempts to use more objective quantification methods for egg classification in the cuckoo. These objective methods, such as spectrophotometry for measuring colours (including the ultraviolet part of the spectrum), and image analysis of photographs for quantifying spotting pattern, size and shape of eggs are now available, and may allow more accurate classification that can be carried out in an automated manner (Gómez, Gordo, & Minias, 2021).

Here, we employ a detailed egg examination and novel analytical approach to analyse a wide range of phenotypic features of cuckoo eggs to predict maternal identity. We also performed human assessment based on photograph sorting to compare the reliability of both methods with the true identity acquired from molecular analyses. Finally, since it has been suggested that similar looking eggs laid by different females may belong to closely related individuals, e.g. mother and daughter (Moksnes *et al.*, 2008), we will for the first time investigate the relationship between the genetic distance of individual cuckoo females and the phenotypic distance of their eggs.

MATERIALS AND METHODS

(a) Study system and data collection

All data were collected in the fishpond area between Mutěnice (48°54′N, 17°02′E) and Hodonín (48°51′N, 17°07′E) in South Moravia, Czech Republic from May to July 2016 and 2017. Here we searched for and regularly checked the great reed warbler (*Acrocephalus arundinaceus*, hereafter

GRW) and Eurasian reed warbler (*Acrocephalus scirpaceus*, hereafter RW) nests, two common hosts of the cuckoo. Most great reed warbler (hereafter GRW) nests were found during the building stage. The rest of the GRW and all Eurasian reed warbler (hereafter RW) nests were found in different stages of breeding by systematic searching. If possible, all GRW nests were checked every day from the nest building stage until clutch completion and approximately every third day during incubation. All RW nests were checked approximately every second day during laying stage and extensively during incubation.

When a cuckoo egg was found in a host nest, we immediately measured its colour and took a photo (see below) to avoid colour change during the incubation period (Hanley *et al.*, 2016). When the eggshell was dirtied (e.g. by faeces or vegetation), we cleaned it with a wet cloth before measuring and photographing. In the cases of multiply parasitized nests, we removed the newly laid cuckoo egg(s), transferred them to an incubator (HEKA-Kongo; HEKA-Brutgeräte, Rietberg, Germany) and incubated them artificially to prevent sample losses caused by the cuckoo nestlings (Honza, Vošlajerová, & Moskát, 2007). The removed cuckoo eggs were either incubated until hatching and chicks placed into non-parasitized host nests (for other experiments) or we froze the eggs before hatching for the future genetic analysis (see *Genotyping and kinship analysis* section). We took a blood sample from all 10-day old cuckoo nestlings from their ulnar or medial tarsometatarsal vein (approx. 25 µl). Finally, we mist-netted 36 and 17 adult cuckoo males and females, respectively, and collected their blood samples from the ulnar vein (approx. 25 µl). All blood samples were stored in 96% ethanol until later genetic analysis.

We performed genealogical analysis based on samples collected in 2016 and 2017 (GenBank project accession No. PRJNA733884). However, here we only analysed the appearance of cuckoo eggs laid in 2017 because we were able to take higher quality photographs in 2017. In 2017, we found 203 cuckoo eggs in total (121 and 82 in the GRW and RW nests, respectively).

We photographed and measured the colour of 192 of them. Among these photographed cuckoo eggs, genetic samples were collected from 109 nestlings or embryos.

(b) Measurements of egg appearance

To obtain background colour we measured reflectance using JAZ Spectrometer (Ocean Optics, Dunedin, FL, USA) in the range 300–700 nm. We took nine measurements (each covering approximately 1 mm^2) at three different parts of the egg (sharp pole, middle part and blunt pole). Since we focused on background colour, we avoided measuring dark spots. For each egg, we used the measurement with the highest reflectance that best corresponded to the colour of the background (Šulc *et al.*, 2019).

Spotting pattern was calculated from digital images taken by a Canon EOS 700D with prime Canon EF 40 mm lens. All photos were taken in RAW format under diffuse sunlight conditions, at the same angle and from the same distance and were referred to a grey standard (X-Rite Colour Checker Grey Scale Chart) with known reflectance. Exposure settings were adjusted accordingly with lighting conditions, yet the ISO value was set constant at 200 and aperture *f*/8. Image calibration, pattern analysis, analysis of shape and measurements of size were performed in ImageJ software (Schneider, Rasband, & Eliceiri, 2012) using the Multispectral Image Calibration and Analysis (MICA) Toolbox (van den Berg *et al.*, 2020). A scale bar was included in each photo, meaning that all images were equally rescaled to the scale of the smallest image (30 pixels/mm). For pattern investigation we applied a granularity analysis approach (van den Berg *et al.*, 2020) that creates a bandpass 'energy' spectrum across a range of spatial frequencies. The pattern energy at each frequency band was measured as the standard deviation of the filtered image (for details, see (Šulc *et al.*, 2019; van den Berg *et al.*, 2020). Since pattern energy cannot distinguish between dark spots on light background and light spots on dark background, we also calculated the 'skew' of the pattern, which quantifies the asymmetry of the pattern luminance distribution. A negative value of skew implies there are more spots than background colour, while a positive

skew implies there is more background colour than spots. Skew was also measured at each granularity band. Since the calculation of the skew is not implemented in the MICA Toolbox, we provide the code in the Supporting Information (Appendix S1). All colour measurements and photos were taken by a single person (M.Š.) to ensure high consistency of the data.

(c) Genotyping and kinship analysis

The genealogical analysis was performed on DNA samples isolated from the blood of adults (36 males and 17 females) and nestlings (n=165) or embryonic tissues (n=47) using a Tissue Genomic DNA mini kit (Geneaid Biotech Ltd, New Taipei, Taiwan) following the manufacturer's protocol. We estimated kinship relationships from nuclear SNPs and mitochondrial DNA haplotypes enabling us to exclude highly implausible maternal (or maternal-sibling) relationships in the inferred genealogy. Kinship relationships were estimated using Colony (Jones & Wang, 2010) based on >1000 nuclear SNPs. The input data file that went into the pedigree analysis in Colony can be found in Supporting Information (Appendix S2).

To acquire the SNP dataset, we genotyped all samples with the ddRAD (double digest restriction-site associated DNA) technique (Peterson *et al.*, 2012) following the protocol of (Piálek *et al.*, 2019). Two prepared libraries were sequenced on an Illumina HiSeq4000 system (2 lanes, 150 cycles P/E) in the EMBL Genomic Core Facility, Heidelberg, Germany. The obtained RAD-tags were processed in Stacks v2.4 (Rochette, Rivera-Colón, & Catchen, 2019) and mapped on the *Cuculus canorus* genome GCA000709325.1 (https://www.ncbi.nlm.nih.gov) with Bowtie2 assembler v2.2.4 (Langmead & Salzberg, 2012). Only loci with 95% or higher presence of individuals were scored and further filtered based on Hardy–Weinberg equilibrium, linkage disequilibrium and minimum minor allele frequency (0.4) in PLINK v1.9 (Purcell *et al.*, 2007) which resulted in a dataset with 1620 markers.

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

Kinship analysis assigned the offspring (n=109) to 31 clusters containing 1–12 eggs each. Since human errors might have created incorrect genetic assignments (e.g. due to confusion of samples), all assigned cuckoo eggs were checked against four additional criteria; 1) laying date – cuckoo females cannot lay eggs more often that every second day (Wyllie, 1981), 2) host species – cuckoos preferentially parasitize a single host (Nakamura, Miyazawa, & Kashiwagi, 2005), 3) laying area – cuckoos lay their eggs in a spatially restricted laying area (Nakamura *et al.*, 2005), and 4) visual check of cuckoo egg appearance – individual cuckoo females lay eggs with a constant egg appearance (Moksnes *et al.*, 2008). Four eggs violated two of these criteria and we suspected them to be assigned incorrectly (for details, see Fig. 8 in Supporting Information. Appendix S3). Therefore, we excluded them from the dataset of genetically assigned eggs (final n=105) and included them into the dataset of photographed eggs without genetic samples (unlabelled dataset, final n=87). For all subsequent analyses dealing with egg phenotype (see below) except the same-different analysis, we removed females to which only one egg has been genetically assigned (n=10), meaning that we used a final dataset of 95 eggs laid by 20 females (labelled dataset). Singleton females were removed as supervised random forest learning cannot be done without at least two eggs per female, and thus we kept the sample size the same across the other clustering methods to enable comparability.

(d) Human assessment

We printed 95 photographs of cuckoo eggs that were standardized in their colour and size (Fig.1–5 in Supporting Information, Appendix S3) using the MICA Toolbox (van den Berg *et al.*, 2020).

We then asked twelve people to sort these photographs and create groups of pictures representing individual females according to similarity in egg appearance. Firstly, we asked them to sort these pictures into an unknown number of groups and, secondly, we asked them to sort these pictures into 20 groups corresponding to the real number of females identified by genetic assessment. For the assessments, we asked 1) five people with no experience with egg appearance from wild animals, 2) three students of avian ecology that had experience with egg appearance from wild birds but not cuckoo eggs and 3) four people (co-authors of this manuscript) that had years of experience with cuckoo eggs. All participants received no other information about the eggs. Cluster similarity between the human assessments compared to the real data was determined using the adjusted Rand index, which provides a corrected-for-chance measure of the similarity between two data clusterings, implemented using the 'cluster_similarity' function from the R package clustereval (Ramey, 2012).

(e) Automatic assessment

We developed an automatic method based on the similarities/differences of cuckoo egg phenotypes. In the first step, we collected colour, pattern and dimension data from calibrated photographs and spectrophotometry data for all cuckoo eggs. Initially, we conducted Principal component analysis (PCA) on different aspects of the egg photographs, in order to avoid the use of correlated variables in the models. PCA components used in the final dataset were selected based on scree plot inspection.

Spectral data: PCA was carried out using binned, scaled spectral data created in the R package *pavo* (Maia *et al.*, 2019), and two spectral PCA components were used in the final dataset. We also used two other spectral measures extracted from *pavo*: the mean brightness (B2 variable; mean relative reflectance over the entire spectral range) and the position of the ultraviolet peak

(UV variable; defined as a wavelength within the range of 300–360nm where reflectance reached the highest point).

Shape data: the variables entered into the PCA were length, maximum width, volume, ellipse deviation and surface area (Troscianko, 2014). Three shape PCA components were selected for inclusion into the final dataset.

Pattern data: the variables entered into the PCA were 12 pattern energies measured at a range of scales (from 1 to 0.0221 in steps of 1/square root of 2) across the whole egg (van den Berg *et al*., 2020), and 12 pattern energy skew values measured at the same range of scales across the whole egg. We also included a measure of total pattern energy across the whole egg. Finally, we divided up each egg into three segments and measured the total pattern energy in each segment as well as the standard deviation between segments, to get a measure of how variable the patterning was across the egg. Three pattern PCA components were selected for inclusion into the final dataset.

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

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

(f) Within- and between-female variability in egg appearance

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

To create a metric of between-female variance, we calculated the average value of each phenotypic trait (n=11) for each female (i.e. created an “average” egg) and then calculated the standard deviation for each phenotypic trait across all females. We then averaged these standard deviations to create a measure of between-female variability across all traits. All trait values were scaled to ensure comparability across different traits.

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

We also quantified individuality using Beecher’s information statistic which can enable comparison across different studies of individual identity signals (Linhart *et al.*, 2019), using the R package *IDmeasurer*. We compared the real data with a control statistic where the ID labels were shuffled.

(g) Unsupervised learning

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

(h) Supervised learning

Female clustering: We used a random forest model with a ‘leave-one-out’ cross-validation approach (Stone, 1974). For each egg in the dataset, the model was trained using a dataset consisting of all other eggs, and then tested using the focal egg. The model attempted to classify

each egg to a given female, and the accuracy of the model was assessed using the classification accuracy value, and through cluster similarity values, as before (taking the average of 1000 runs, as random forest modelling is a stochastic process). We also fitted a random forest model to the full dataset to allow us to assess the importance of the different variables included in the model (using the mean decrease in accuracy).

Same/different analysis: We used an approach where a random forest model was trained to label pairs of eggs as 'same' or 'different'. The training set used 4000 'same' rows, where the two eggs were from the same female and 4000 'different' rows, where the two eggs were from different females.

As above, we used a 'leave-one-out' cross-validation approach. For each egg in the dataset, the model was trained using a same/different training dataset generated from all other eggs. In the test phase, we compared the target egg on all other eggs. We calculated whether the target egg was successfully labelled (i.e. it was consistently matched to eggs from the same female) or whether it was erroneously labelled (i.e. it was consistently matched to eggs from another female). The entire process (i.e. the training and testing process on the full dataset) was repeated 100 times to allow us to calculate a reliability metric i.e. what percentage of the matches made were true positives.

For the unlabelled dataset, we ran the training component as above. For the testing phase, we tested each of the unlabelled eggs on all the other eggs, calculating how many times in each of 100 runs the target egg was matched with a cluster of eggs that were from the same female. If the percentage was greater than 95%, we considered this egg as a candidate for being from this female. To corroborate this conclusion, we used non-phenotypic data: laying dates, laying locality and host species.

(i) Phenotype-genotype similarity

Nine of the 30 labelled females were caught, and they were genotyped via blood sampling as described above. Thus, we were able to calculate genetic similarities among these females (Supporting Information, Appendix S4) which was done in Geneious 10.2.6 (https://www.geneious.com). To compare the genetic similarities between these females with phenotype similarities of their eggs, we created a trait distance matrix by taking means of the phenotypic parameters from their egg data, and then using Euclidean distance as the distance metric. We compared the genetic distance matrix with the trait distance matrix using a Mantel test, a statistical test of the correlation between two matrices, implemented in the vegan package in R using the Kendall method (as this is most appropriate for a small dataset). We also split the phenotype data into different components (spectral, pattern and shape) and calculated the phenotype-genotype similarities for each of these components separately, to test whether different aspects of the egg phenotype are differentially correlated with the female genotypes.

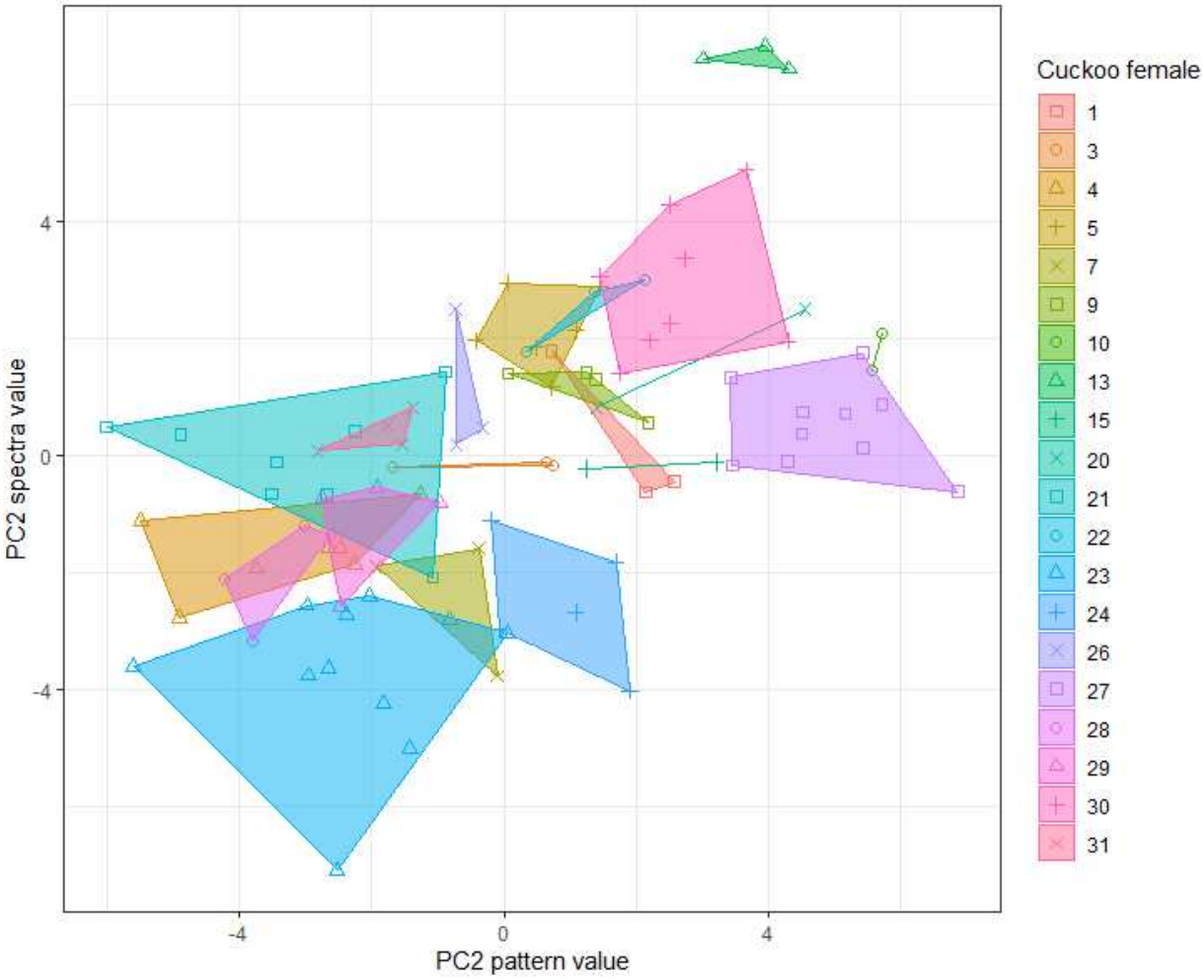
All code used for measuring egg appearance and carrying out analyses performed in R (R Development Core Team, 2018) is provided in Supporting Information (Appendix S5).

RESULTS

(a) Within- and between-female variability in egg appearance

Some females laid eggs with very low variability in their appearance (e.g. female 13 – within-female variance=0.33, Fig. 2 in Supporting Information, Appendix S3) and others, on the contrary, had relatively high variability (e.g. female 29 – within-female variance=1.31, Fig. 4 in Supporting Information, Appendix S3). The mean within-female variance was 0.85 (SD=0.30). Overall, between-female variance (mean of trait standard deviations=1.83, n=11 traits; SD=1.02) was higher than within-female variance (one sample t-test, $t=14.87$, $df=19$, $p<0.001$). Beecher's information statistic $H_s = 1.97$ for this dataset, considering only significant variables. (This compares to a control $H_s = 0.56$, where the ID labels were randomly shuffled). Variability in the egg appearance is also visible in Fig. 1 where the two most informative variables in the random

316 forest analysis (PC2 for pattern and PC2 for spectral data), are plotted (for more information about
317 the variables, see below and Table 2).



318

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

323 *(b) Human assessment*

324 Participants with some experience of working with bird eggs performed better at the grouping task
325 than those with no experience, though there is no clear evidence that specific experience of

working with cuckoo eggs is beneficial (Table 1; for all results of individual people, see Supporting Information, Appendix S4).

Table 1. Cluster similarities of egg sorting performed by humans both without knowledge (when 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)

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

(c) Unsupervised learning

Clustering using unsupervised hierarchical learning gave a cluster similarity value of 0.452; similar to that of experienced human observers, but better than inexperienced observers (Table 1).

(c) Supervised learning (random forest analysis)

Female clustering

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

We assessed variable importance (Table 2) using a full model including all data. PC2 for pattern was the most important variable for classification, and the variables loading onto this PC were

predominantly those for the ‘skew’ of the pattern. PC2 for spectra was also important, with this variable being influenced by the ‘blueness/greenness’ of the egg.

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	26.80	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	12.90	-
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

Variables with larger mean decrease in accuracy are more important for classifying the data (mean decrease in accuracy is a measure of how much the accuracy of the random forest

decreases due to the exclusion/permutation of a single variable). The main PCA loadings are those that were greater than +/- 0.25.

Same/different analysis

40 labelled eggs (out of 105) passed the reliability criterion, being assigned to a unique female on 95% or more of the 100 runs. 39 of these (97.5%) were assigned to the correct female; only one was consistently erroneously assigned to the incorrect female. In this case, an egg from female 29 (e92) was matched with eggs from female 23.

Out of 87 unlabelled eggs, the model was able to reliably (on 95% of runs) identify 25 as belonging to a labelled female (8 eggs assigned to female 5, 5 eggs to female 27, 3 eggs to female 13, 2 eggs to female 29, 21 and 23, and 1 egg to each of females 4, 28 and 30). For visual comparison, see figures 1–5 in Supporting Information (Appendix S3).

(d) Phenotype-genotype similarity

The average genetic similarity between 36 pairs of nine cuckoo females was 99.38% (± 0.03 SD). The most genetically similar were females 23 and 24 (genetic similarity=99.50%) where female 23 was the mother and female 24 her daughter. There was no significant relationship between female genetic distance and their overall egg phenotype distance (Mantel test r=0.1968, p=0.10, Fig. 2).

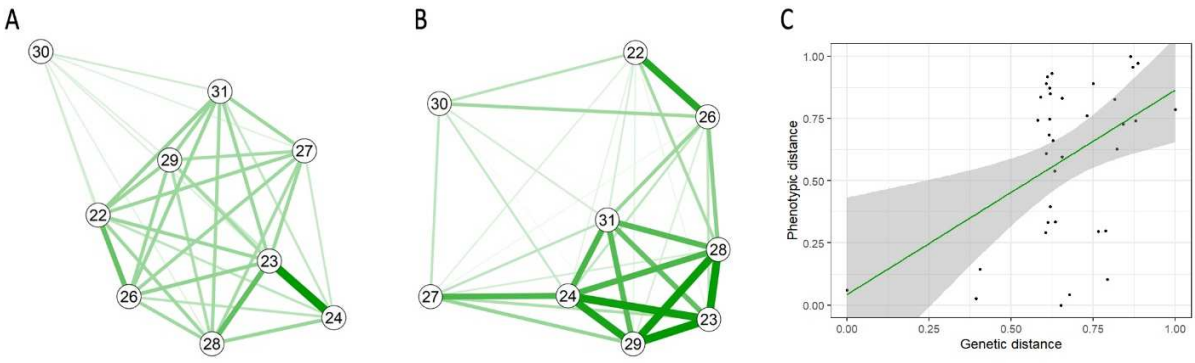


Figure 2 Phenotypic distances of nine average eggs laid by nine genotyped common cuckoo females (A) and their genetic distances (B). Thicker green lines denote higher phenotypic and genetic similarity. Correlation between phenotypic and genetic distances (C).

When considering each aspect of phenotype distance separately, both pattern/luminance and shape distance metrics did not correlate with genetic distance (r=0.03, p=0.39 and r=-0.23, p=0.93 respectively, Fig. 3). However, spectral distance did correlate with genetic distance (r=0.36, p=0.04, Fig. 3).

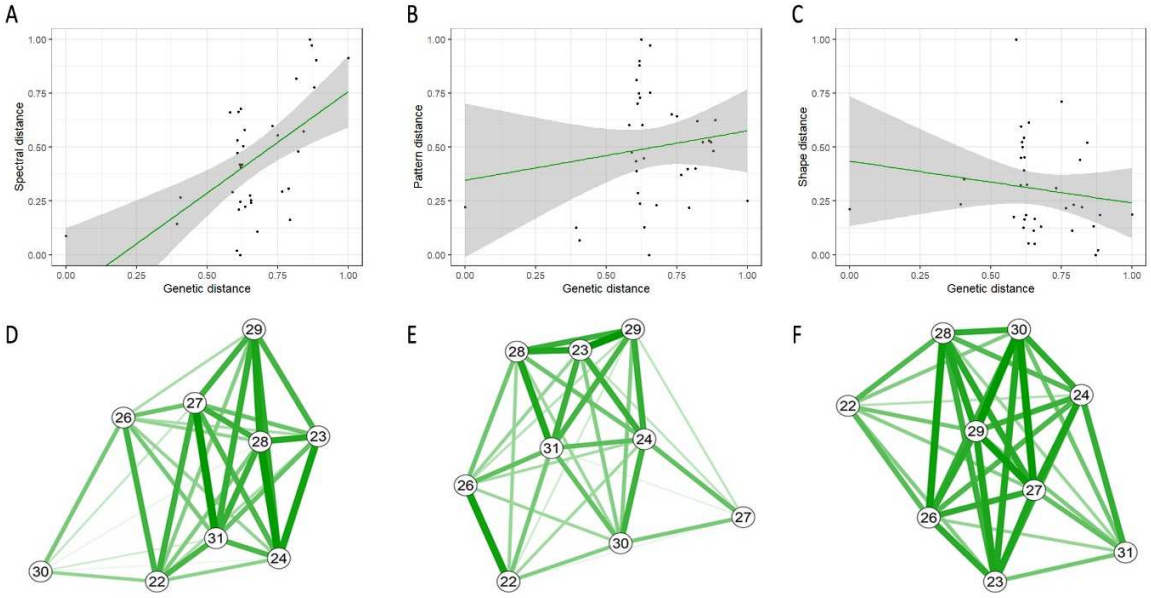


Figure 3 Correlation between spectral (A), pattern/luminance (B) and shape (C) distances, respectively and genetic distances. Individual phenotypic distances of average eggs laid by nine genotyped common cuckoo females: spectral (D), pattern/luminance (E) and shape (F) distances.

DISCUSSION

The results of our study support the 'constant egg-type hypothesis' predicting that individual cuckoo females lay eggs with a constant appearance (Moksnes *et al.*, 2008). This is apparent

382 from the photos of cuckoo eggs (Fig. 1–5 in Supporting Information, Appendix S3) and supported
383 by the fact that the within-clutch variation of cuckoo eggs is significantly lower than between-clutch
384 variation. This has also been observed in other bird species and several adaptive explanations
385 have been proposed for this phenomenon (reviewed in Gómez *et al.*, 2021), such as easier
386 recognition of the parasitic egg by hosts (Øien *et al.*, 1995), recognition of an individual's own
387 clutch in colonially-breeding birds (Hauber *et al.*, 2019) or signalling female quality (Moreno &
388 Osorno, 2003). Therefore, there is the potential to use egg appearance to identify individual bird
389 females and our study shows that automatic analyses may be a more accurate method than
390 human assessments.

391 The unsupervised hierarchical clustering method showed very similar results to experienced
392 human classifiers, while supervised random forest analysis showed considerably better results:
393 81% of cuckoo eggs were assigned correctly. This suggests that in some cases, automatic egg
394 assignment to females should be used rather than human assessment. Detailed consideration of
395 the clusters created by humans and the automatic methods showed that the same females were
396 problematic for both clustering methods (all sorting results can be found in Supporting Information,
397 Appendix S4), probably reflecting phenotypic overlap between some individuals (Fig. 1). Our
398 results showed that one of the pattern characteristics (skew), blueness of colour and finally egg
399 size were the most important parameters for improving clustering accuracy. The slight
400 improvement in clustering accuracy for the automatic methods over human assessment may
401 reflect the use of features that humans are not able to see (e.g. the reflected ultraviolet radiation).

402 The greatest benefit of the methods we present is the possibility to reliably assign unlabelled eggs
403 to individual females. Same-different analysis that uses both genetic and phenotypic information
404 of the labelled dataset showed 97.5% (39 of 40 cases) accuracy of egg assignment. Moreover,
405 the one wrongly assigned egg (although looking similar to the other eggs of the assigned female)
406 would be the only one posteriori suspected to be an incorrect assignment because it was laid into

407 the nest of another host species, in another locality and on the same day as another egg laid by
408 the same female (Supporting Information, Appendix S4).

409 Using this method, we were able to assign 25 eggs (out of 87) to nine known females. The
410 reliability is supported by the fact that all these 25 eggs meet all additional criteria and their
411 appearance, host species and locality where they were laid and laying date perfectly matches
412 with other eggs laid by the assigned cuckoo females (Supporting Information, Appendix S4). Our
413 method seemed to work well especially for females that laid very distinctive eggs and therefore
414 we may expect better results of the method in species where between-clutch variation
415 substantially exceeds the within-clutch variation. It must also be noted that the accuracy of the
416 assignment will increase with the relative number of (genetically and phenotypically) analysed
417 samples in the study area that are used for the training dataset, because broad sampling will
418 reduce the chance that an unsampled egg that has been laid by a completely new female will be
419 assigned to an existing (incorrect) female. Finally, we recommend applying other available
420 information (e.g. laying date and laying area) to eliminate potential incorrect assignments.

421 A previous study suggested that closely related cuckoo females may lay eggs that are
422 indistinguishable from each other (Moksnes *et al.*, 2008). Our results partially agree because
423 humans (even experienced ones) and the unsupervised automatic clustering method failed to
424 distinguish eggs of three most closely related pairs of cuckoo females (females 23 vs 24 – mother
425 and daughter, 23 vs 28 and 22 vs 26, respectively: Supporting Information, Appendix S3).
426 Moreover, detailed comparison between genetic distances of nine laying females and phenotypic
427 distances of their eggs showed the background colour of eggs was more similar between more
428 related females. However, genetic distances between females did not correlate with pattern and
429 shape distances of their eggs. Therefore, although it has been shown that all investigated egg
430 features – colour, spotting pattern and also size – have high heritability (Gosler *et al.*, 2000;
431 Christians, 2002; Morales *et al.*, 2010), our results indicate that the background colour of cuckoo

eggs might be the most heritable. This also supports the idea that egg colour seems to be vital for egg recognition in brood parasitic systems (Spottiswoode & Stevens, 2010; Honza *et al.*, 2014). However, since several studies reported that hosts use spotting pattern (de la Colina *et al.*, 2012) or egg size (Marchetti, 2000) when recognizing and eliminating parasitic eggs, we still expect relatively high heritability of these egg traits in brood parasites. We suspect that the insignificant relationship between genetic distance and phenotypic distance in spotting pattern and size reflects our limited sample size. A larger sample size, including more mother-daughter pairs, is needed to truly estimate heritability values of individual egg traits (de Villemereuil, Gimenez, & Doligez, 2013). The lack of significant correlation between egg shape and genetic similarity may also be explained by the fact that egg size often reflects the size of laying females (Larsson & Forslund, 1992), which depends on the genetic contribution of both parents and therefore might differ more even in closely related females. Moreover, cuckoos are raised by host parents that vary in their provisioning care (Požgayová *et al.*, 2018), which may also influence the body size of cuckoo females in adulthood. Finally, there is a positive relationship between food availability and egg size (reviewed in Christians, 2002). Consequently, since egg size and shape may differ even in closely related females, these traits may be very useful for identification. Indeed, some human participants (and also supervised clustering analysis) distinguished eggs of the three closely related females correctly, presumably because of differences in size and shape (see Supporting Information, Appendix S4).

CONCLUSION

We conclude that although individual cuckoo females laid eggs with constant appearance, egg phenotype alone cannot be used to identify individual cuckoo females. This might be caused by the fact that closely related females lay eggs similar to each other. However, here we present a novel supervised method that substantially increased our sample size which consequently helped us to precisely estimate laying areas of cuckoo females (Koleček *et al.*, 2021). In future, we plan

to use this method to reveal more about the ecology and evolution of cuckoos, e.g. to investigate the number of eggs laid by individual females or host selection. We encourage researchers investigating inter- and intra-specific brood parasitism to use this low-cost and ethically more appropriate method of individual identification. As it seems that the phenomenon of higher between-female variation and lower within-female variation in egg appearance is common in birds, identification of laying females using our method has the potential to be of widespread use, both for brood parasitic species and also for other species where e.g. females are difficult to catch.

Data accessibility

The dataset supporting this article has been uploaded as part of the electronic Supporting Information (Appendices S1–S5). All ddRAD reads in a form of alignments (BAM) were deposited into the GenBank SRA (Sequence Read Archive) under project accession No. PRJNA733884.

Acknowledgements

We thank K. Bendová, M. Fridrichová, J. Hodanová, L. Mari, K. Míčková, M. Potůčková, K. Stehlíková, P. Potůček and D. Tesař for their sorting of cuckoo eggs. L. Mari also helped with writing of the genetic part of methods. We also thank R. Poláková, K. Sosnovcová, M. Čapek, V. Jelínek, J. Koleček, L. Kulísek and B. Prudík for help with fieldwork. We thank V. Jelínek for many fruitful comments especially at the early stage of data collection and M. Elsner that brought many good ideas regarding data analyses. We are also very grateful to P. Linhart for his comments on Beecher's information statistic and V. Beneš and the European Molecular Biology Laboratory Genomic Core Facility in Heidelberg (Germany) for their kind advice and technical support regarding Illumina sequencing. The managers of the Hodonín Fish Farm kindly permitted us conducting the fieldwork on their grounds. Finally, many thanks also to my handy and unfortunately deceased grandfather, who made me an amazing egg holder for the spectral measurements.

Ethical note

This study was carried out with the permission of the regional nature conservation authorities (JMK: 38506/2016; MUHOCJ: 14306/2016/OŽP). The fieldwork adhered to the animal care protocol (039/2011 AVČR and 3030/ENV/17-169/630/17) and to the Czech Law on the Protection of Animals against Mistreatment (CZ 01284). This study was carried out with the permission of the regional nature conservation authorities (JMK: 38506/2016; MUHOCJ: 14306/2016/OŽP).

Funding

This work was supported by the Czech Science Foundation (17-12262S) and by the Programme for research and mobility of young researchers of the Czech Academy of Sciences (MSM200931801, awarded to MŠ).

Author contributions

M.Š. and A.E.H. conceived the ideas and designed methodology; M.Š., G.Š., P.P., M.P., V.B. and M.H. collected data; M.Š., A.E.H., J.T., L.P. and R.P. analysed data; M.Š. led the writing of the manuscript. All authors contributed to the drafts and gave final approval for publication.

Preprint version

This manuscript has been previously submitted to bioRxiv as a preprint (Sulc *et al.*, 2020).

REFERENCES

Sulc, M., A. E. Hughes, J. Troscianko, G. Stetkova, P. Prochazka, M. Pozgayova, L. Pialek, et al. 2020. Automatic identification of bird females using egg phenotype. bioRxiv.

van den Berg CP, Troscianko J, Endler JA, Marshall NJ, Cheney KL. 2020. Quantitative Colour Pattern Analysis (QCPA): A comprehensive framework for the analysis of colour patterns in nature. *Methods in Ecology and Evolution* 11: 316–332.

Christians JK. 2002. Avian egg size: variation within species and inflexibility within individuals. *Biological Reviews* 77: 1–26.

Christin S, Hervet E, Lecomte N. 2019. Applications for deep learning in ecology. *Methods in Ecology and Evolution* 10: 1632–1644.

de la Colina MA, Pompilio L, Hauber ME, Reboreda JC, Mahler B. 2012. Different recognition cues reveal the decision rules used for egg rejection by hosts of a variably mimetic avian brood parasite. *Animal cognition* 15: 881–889.

Davies N. 2010. *Cuckoos, cowbirds and other cheats*. A&C Black.

Ferreira AC, Silva LR, Renna F, Brandl HB, Renoult JP, Farine DR, Covas R, Doutrelant C. 2020. Deep learning-based methods for individual recognition in small birds. *Methods in Ecology and Evolution* 11: 1072–1085.

Fossøy F, Antonov A, Moksnes A, Røskaft E, Vikan JR, Møller AP, Shykoff JA, Stokke BG. 2011. Genetic differentiation among sympatric cuckoo host races: males matter. *Proceedings of the Royal Society B: Biological Sciences* 278: 1639–1645.

Fossøy F, Moksnes A, Røskaft E, Antonov A, Dyrcz A, Moskat C, Ranke PS, Rutila J, Vikan JR, Stokke BG. 2012. Sex allocation in relation to host races in the brood-parasitic common cuckoo (Cuculus canorus). *PloS one* 7.

Gibbs HL, Sorenson MD, Marchetti K, Brooke M de L, Davies NB, Nakamura H. 2000. Genetic evidence for female host-specific races of the common cuckoo. *Nature* 407: 183–186.

Gómez J, Gordo O, Minias P. 2021. Egg recognition: The importance of quantifying multiple repeatable features as visual identity signals. *Plos one* 16: e0248021.

Gosler AG, Barnett PR, James Reynolds S. 2000. Inheritance and variation in eggshell patterning in the great tit Parus major. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 267: 2469–2473.

Hanley D, Šulc M, Brennan PL, Hauber ME, Grim T, Honza M. 2016. Dynamic egg color mimicry. *Ecology and evolution* 6: 4192–4202.

Hauber ME, Luro A, McCarty CJ, Barateli K, Cassey P, Hansen ES, Dale J. 2019. Interannual repeatability of eggshell phenotype in individual female Common Murres (*Uria aalge*). *Canadian Journal of Zoology* 97: 385–391.

Hewson CM, Thorup K, Pearce-Higgins JW, Atkinson PW. 2016. Population decline is linked to migration route in the Common Cuckoo. *Nature Communications* 7: 1–8.

Höltje H, Mewes W, Haase M, Ornés AS. 2016. Genetic evidence of female specific eggshell colouration in the Common Crane (Grus grus). *Journal of Ornithology* 157: 609–617.

539 Honza M, Šulc M, Jelínek V, Požgayová M, Procházka P. 2014. Brood parasites lay eggs
540 matching the appearance of host clutches. *Proceedings of the Royal Society B: Biological*
541 *Sciences* 281: 20132665.

542 Honza M, Vošlajerová K, Moskát C. 2007. Eviction behaviour of the common cuckoo Cuculus
543 canorus chicks. *Journal of Avian biology* 38: 385–389.

544 Hou J, He Y, Yang H, Connor T, Gao J, Wang Y, Zeng Y, Zhang J, Huang J, Zheng B. 2020.
545 Identification of animal individuals using deep learning: A case study of giant panda. *Biological*
546 *Conservation* 242: 108414.

547 Jelínek V, Šulc M, Štětková G, Honza M. 2021. Fast and furious: host aggression modulates
548 behaviour of brood parasites. *Ibis* 163: 824–833.

549 Jones OR, Wang J. 2010. COLONY: a program for parentage and sibship inference from
550 multilocus genotype data. *Molecular ecology resources* 10: 551–555.

551 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A,
552 Markowitz S, Duran C. 2012. Geneious Basic: an integrated and extendable desktop software
553 platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.

554 Koleček J, Piálková R, Piálek L, Šulc M, Hughes AE, Brlík V, Procházka P, Požgayová M,
555 Capek M, Sosnovcová K. 2021. Spatiotemporal patterns of egg laying in the common cuckoo.
556 *Animal Behaviour* 177: 107–116.

557 Krause J, Krause S, Arlinghaus R, Psorakis I, Roberts S, Rutz C. 2013. Reality mining of animal
558 social systems. *Trends in ecology & evolution* 28: 541–551.

559 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. 940. *Methods*
560 9: 541.

561 Larsson K, Forslund P. 1992. Genetic and social inheritance of body and egg size in the
562 barnacle goose (Branta leucopsis). *Evolution* 46: 235–244.

563 Lindberg MS. 2012. A review of designs for capture–mark–recapture studies in discrete time.
564 *Journal of Ornithology* 152: 355–370.

565 Linhart P, Osiejuk TS, Budka M, Šálek M, Špinka M, Policht R, Syrová M, Blumstein DT. 2019.
566 Measuring individual identity information in animal signals: Overview and performance of
567 available identity metrics. *Methods in Ecology and Evolution* 10: 1558–1570.

568 Lyon BE. 2003. Egg recognition and counting reduce costs of avian conspecific brood
569 parasitism. *Nature* 422: 495–499.

570 Maia R, Gruson H, Endler JA, White TE. 2019. pavo 2: new tools for the spectral and spatial
571 analysis of colour in R. *Methods in Ecology and Evolution* 10: 1097–1107.

572 Marchetti K. 2000. Egg rejection in a passerine bird: size does matter. *Animal Behaviour* 59:
573 877–883.

574 Moksnes A, Røskaft E, Rudolfsen G, Skjelseth S, G. Stokke B, Kleven O, Lisle Gibbs H, Honza
575 M, Taborsky B, Teuschl Y. 2008. Individual female common cuckoos Cuculus canorus lay
576 constant egg types but egg appearance cannot be used to assign eggs to females. *Journal of*
577 *Avian Biology* 39: 238–241.

578 Morales J, Kim SY, Lobato E, Merino S, Tomás G, MARTÍNEZ-de la PUENTE J, Moreno J.
579 2010. On the heritability of blue-green eggshell coloration. *Journal of Evolutionary Biology* 23:
580 1783–1791.

581 Moreno J, Osorno JL. 2003. Avian egg colour and sexual selection: does eggshell pigmentation
582 reflect female condition and genetic quality? *Ecology Letters* 6: 803–806.

583 Nakamura H, Miyazawa Y, Kashiwagi K. 2005. Behavior of radio-tracked Common Cuckoo
584 females during the breeding season in Japan. *Ornithological Science* 4: 31–41.

585 Øien IJ, Moksnes A, Røskaft E. 1995. Evolution of variation in egg color and marking pattern in
586 European passerines: adaptations in a coevolutionary arms race with the cuckoo, Cuculus
587 canorus. *Behavioral Ecology* 6: 166–174.

588 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an
589 inexpensive method for de novo SNP discovery and genotyping in model and non-model
590 species. *PloS one* 7.

591 Petrusková T, Pišvejcová I, Kinštová A, Brinke T, Petrusek A. 2016. Repertoire-based individual
592 acoustic monitoring of a migratory passerine bird with complex song as an efficient tool for
593 tracking territorial dynamics and annual return rates. *Methods in Ecology and Evolution* 7: 274–
594 284.

595 Petrželková A, Pöysä H, Klvaňa P, Albrecht T, Hořák D. 2017. Egg morphology fails to identify
596 nests parasitized by conspecifics in common pochard: a test based on protein fingerprinting and
597 including female relatedness. *Journal of Avian Biology* 48: 229–234.

598 Piálek L, Burress E, Dragová K, Almirón A, Casciotta J, Říčan O. 2019. Phylogenomics of pike
599 cichlids (Cichlidae: Crenicichla) of the C. mandelburgeri species complex: rapid ecological
600 speciation in the Iguazú River and high endemism in the Middle Paraná basin. *Hydrobiologia*
601 832: 355–375.

602 Požgayová M, Piálková R, Honza M, Procházka P. 2018. Sex-specific nestling growth in an
603 obligate brood parasite: Common Cuckoo males grow larger than females. *The Auk:*
604 *Ornithological Advances* 135: 1033–1042.

605 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De
606 Bakker PI, Daly MJ. 2007. PLINK: a tool set for whole-genome association and population-
607 based linkage analyses. *The American journal of human genetics* 81: 559–575.

608 R Development Core Team R. 2018. *R: A language and environment for statistical computing*.
609 R foundation for statistical computing Vienna, Austria.

610 Ramey JA. 2012. clusteval: Evaluation of clustering algorithms. *available at https://CRAN. R-*
611 *project. org/package= clusteval* 666.

612 Rochette NC, Rivera-Colón AG, Catchen JM. 2019. Stacks 2: Analytical methods for paired-end
613 sequencing improve RADseq-based population genomics. *Molecular ecology* 28: 4737–4754.

614 Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image
615 analysis. *Nature methods* 9: 671.

616 Soler M. 2017. Brood Parasitism in Birds: A Coevolutionary Point of View. *Avian Brood*
617 *Parasitism*. Springer, 1–19.

618 Spottiswoode CN, Stevens M. 2010. Visual modeling shows that avian host parents use multiple
619 visual cues in rejecting parasitic eggs. *Proceedings of the National Academy of Sciences* 107:
620 8672–8676.

621 Stone M. 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of the*
622 *Royal Statistical Society: Series B (Methodological)* 36: 111–133.

623 Stowell D, Petrusková T, Šálek M, Linhart P. 2019. Automatic acoustic identification of
624 individuals in multiple species: improving identification across recording conditions. *Journal of*
625 *the Royal Society Interface* 16: 20180940.

626 Sulc M, Hughes AE, Troscianko J, Stetkova G, Prochazka P, Pozgayova Mi, Pialek L, Pialkova
627 R, Brlik V, Honza M. 2020. Automatic identification of bird females using egg phenotype.
628 *bioRxiv*.

629 Šulc M, Štětková G, Procházka P, Požgayová M, Sosnovcová K, Studeckỳ J, Honza M. 2020.
630 Caught on camera: circumstantial evidence for fatal mobbing of an avian brood parasite by a
631 host. *Journal of Vertebrate Biology* 69: 1–6.

632 Šulc M, Troscianko J, Štětková G, Hughes AE, Jelínek V, Capek M, Honza M. 2019. Mimicry
633 cannot explain rejection type in a host–brood parasite system. *Animal Behaviour* 155: 111–118.

634 Troscianko J. 2014. A simple tool for calculating egg shape, volume and surface area from
635 digital images. *Ibis* 156: 874–878.

636 de Villemereuil P, Gimenez O, Doligez B. 2013. Comparing parent–offspring regression with
637 frequentist and Bayesian animal models to estimate heritability in wild populations: a simulation
638 study for Gaussian and binary traits. *Methods in Ecology and Evolution* 4: 260–275.

639 Weinstein BG. 2018. A computer vision for animal ecology. *Journal of Animal Ecology* 87: 533–
640 545.

641 Wyllie I. 1981. *The Cuckoo*. London, UK: Batsford.

642

643 **SUPPORTING INFORMATION**

644 Additional Supporting Information may be found in the online version of this article at the

645 publisher's web-site.

Appendix S1. Code for ImageJ software used for analyzing egg pattern, including pattern energy and skew.

Appendix S2. Data used for pedigree analysis in Colony software.

Appendix S3. Standardized photographs of all cuckoo eggs used in all phenotype analyses.

Appendix S4. All data about cuckoo eggs and analyses performed. This includes phenotype and laying information about all cuckoo eggs, results of human and automatic clustering and genetic distances of individual adult cuckoo females.

Appendix S5. Statistical code for R software.