Title: Automated Workflows using Quantitative Colour Pattern Analysis (QCPA): A Guide to Batch Processing and Downstream Data Analysis

Cedric P. van den Berg[1,2,3] , Nicholas D. Condon[4], Cara Conradsen[1], Thomas E. White[2], Karen L. Cheney[1]

[1] School of the Environment, The University of Queensland, St. Lucia campus, St Lucia QLD 4072, Australia

[2] School of Life and Environmental Sciences, The University of Sydney, Camperdown NSW 2050, Australia

[3] School of Biological Sciences, University of Bristol, 24 Tyndall Avenue, BS8 1TQ, UK

[4] Institute for Molecular Bioscience, The University of Queensland, 306 Carmody Rd, St Lucia QLD 4067, Australia

Corresponding author: Cedric van den Berg, c.vandenberg@bristol.ac.uk

Running title: Automated workflows using Quantitative Colour Pattern Analysis (QCPA)

Keywords: ImageJ, java, R software, automation, animal colouration, quantitative colour pattern analysis

Manuscript length: 3003 words

Figures & tables: 2 figures

Electronic supplement: Manual v1.0.pdf / Worked examples v1.0.pdf

Abstract

Animal and plant colouration presents a striking dimension of phenotypic variation, the study of which has driven general advances in ecology, evolution, and animal behaviour. Quantitative Colour Pattern Analysis (QCPA) is a dynamic framework for analysing colour patterns through the eyes of non-human observers. However, its extensive array of user-defined image processing and analysis tools means image analysis is often time-consuming. This hinders the full use of analytical power provided by QCPA and its application to large datasets. Here, we offer a robust and comprehensive batch script, allowing users to automate many QCPA workflows. We further provide a set of useful R scripts for downstream data extraction and analysis. These scripts will empower users to exploit the full analytical power of QCPA and facilitate the development of customised semi-automated workflows. Such quantitatively scaled workflows are crucial for exploring colour pattern space and developing ever-richer models of visual perception in animals other than humans. These advances will, in turn, facilitate testing hypotheses on the function and evolution of vision and signals at quantitative and qualitative scales, which are otherwise computationally unfeasible.

1. Introduction

Understanding the perception of visual information by non-human observers is crucial to studying the ecology and evolution of animal and plant colouration. The last two decades have seen the creation and widespread adaptation of tools and methods that allow researchers to simulate or approximate aspects of animal vision, such as colour contrast perception (e.g. Endler and Mielke, 2005; Gawryszewski, 2018; Kemp et al., 2015; Renoult et al., 2017; Vorobyev and Osorio, 1998) and spatial vision (Godfrey et al. 1987; Caves and Johnsen 2018). These advances coincide with the steady development of colour pattern analyses (e.g. Chan et al., 2018; Endler, 1991; Endler, 2012; Stoddard et al., 2014; Van Belleghem et al., 2018; van den Berg et al., 2020) as well as their integration into increasingly comprehensive collections of tools and functions such as Pavo (Maia et al. 2019) or the Multispectral Image Calibration and Analysis toolbox (MICA) (Troscianko and Stevens 2015).

Quantitative Colour Pattern Analysis (QCPA) (van den Berg et al. 2020b) is a recent and powerful addition to this landscape. It provides a dynamic analytical framework for analysing visual scenes through the eyes of ecologically relevant observers and is integrated into the MICA toolbox (Troscianko and Stevens 2015). It allows users to choose preferred tools and analyses uniquely suitable for analysing spatiochromatic information. Briefly, it achieves this by taking full-spectrum images as its input (thereby generating '.mspec' files) before applying a suite of models which can consider the spectral, spatial, and temporal sensation of non-human viewers, to ultimately give image-based and numerical outputs that summarise the structure of visual scenes and stimuli (van den Berg et al. 2020b).

Preparing and processing calibrated images in the framework, however, remains tedious. This is due to the time needed to translate calibrated digital images into cone catch

images, user-guided input for identifying regions of interest (ROIs), and the subsequent application of image processing and analyses. Saving and labelling multiple output files, as well as extracting data for subsequent statistical analysis, also remains a tiresome process. As a result, applying the QCPA at quantitative scales, such as analysing hundreds or thousands of images at multiple viewing distances, is nearly impossible in its original form.

An image currently needs hours of manual, repetitive work to obtain the full numerical output from the various types of analyses in the QCPA framework. This is especially true when considering multiple viewing distances and regions of interest (ROIs). Consequently, most studies using the framework do not consider more than a few dozen individual observations with a limited subset of available image statistics (e.g., Nokelainen et al., 2021; Rodríguez-Morales et al., 2021). This is problematic as quantifying ecologically relevant effects, such as colour pattern variability in natural populations or the analysis of behavioural data, requires sufficiently large sample sizes to provide adequate statistical power. This holds particularly true for a method producing hundreds of image statistics from a single observation, such as the QCPA, where the number of available image statistics easily outnumbers observations (van den Berg et al. 2022). Considering all possibly relevant image statistics and deducing a relevant set of parameters is often the desired approach instead of pre-emptively narrowing down the number of considered image statistics (van den Berg et al. 2022). However, such deductive approaches require adequate statistical solutions and remain an intriguing challenge in using complex colour pattern spaces (Stoddard and Osorio 2019; van den Berg et al. 2020b, 2022).

The QCPA framework is, therefore, in urgent need of automation to facilitate its complete application to large datasets. However, unlike other image analysis tools available

to visual ecologists, such as PAVO (Maia et al. 2019), MICA is an ImageJ (Schneider et al. 2012) plugin written in Java (Arnold et al. 2005). It is intentionally tailored towards accessibility via graphical user interfaces (GUIs) rather than command line prompts. Unlike R software (R Core Team 2021) or Python (van Rossum 1995), frequently taught in undergraduate and HDR Biology degrees worldwide, Java is not a programming language that biologists are readily familiar with. Therefore, writing scripts and macros for a complex network of GUI-guided tools in an unfamiliar language is a hurdle for most users of the QCPA framework seeking to batch-process large volumes of image data. This creates the potential for unintentional gatekeeping by researchers with access to programmers familiar with writing scripts in JAVA and an adequate understanding of the framework's mechanisms. The result is a hidden custom of informally traded scripts, which undermines fast and equal access to methods in the field.

Here, we present a dynamic batch-processing extension to QCPA, allowing users to apply almost the entire QCPA framework or selected parts to large datasets via a flexible, GUI-guided input. We further provide a set of R scripts that efficiently extract numerical data from the consecutively stored output files for downstream analysis. We describe the functionality of both—the QCPA batch script and complementary R scripts—and provide detailed worked examples. While this script does not provide a complete solution for every possible application of the QCPA, we hope it will offer a viable solution to most. This batch script should also help researchers customise their automated pipelines, stimulating the open exchange of programming solutions via platforms such as open-access publications, open-access platforms such as GitHub or the dedicated user forum for the MICA toolbox (www.empiricalimaging.com).

2. QCPA batch script

2.1. Intended use

The QCPA batch script is intended to quantify spatiochromatic information of an animal against its visual background. As such, the script can also be applied to images without an animal or object of interest or an animal or object itself. The batch script is intended for users with moderate experience in the manual application of QCPA.

The script allows the user to choose the following desired analytical outputs individually: 1) Colour Adjacency Analysis (CAA), Visual Contrast Analysis (VCA), Boundary Strength Analysis (BSA) and particle analysis; 2) Local Edge Intensity Analysis (LEIA); 3. Colour Maps; and 4) GabRat. For a detailed discussion of these analyses, please see the original publications and their modifications for QCPA as listed in van den Berg et al. (2020).

Notably, the script requires cone mapping functions derived from calibrated cameras (several standard profiles included in the MICA toolbox) and currently does not permit the use of chart-based models. While covering many analyses available in QCPA and the MICA toolbox, it does not provide a comprehensive library of automation. However, the script is designed to be easily modified by the user. Such modifications are explicitly invited from the community, and we encourage their open sharing on community platforms in the interest of mutual benefit among researchers in the field.

2.2. Data preparation

Image processing in QCPA is specific to a variety of user choices specifying the properties of the observer's visual system, the light environment in which a picture was taken as well as the intended target light environment, observer viewing distance, and a variety of parameters

127 and processing choices that vary depending on the desired analyses. The batch script uses
128 three approaches to minimise the need to repeat the number of times such input is required:

129 1. Pre-defined folder structure
130 2. Auxiliary files
131 3. Batch processing graphical user interface (Batch GUI)

132 *2.2.1 Folder structure*

133 The batch script requires input data to be organised in a specific way. This refers to
134 the naming of files and the folder structure. For example, the script expects individual
135 observations (i.e., animals or objects) to be grouped within an overarching folder (i.e., species
136 or site) and file naming must be uniform. This allows the script to reliably detect individual
137 observations while allowing the user to structure their data according to sites or taxa (see the
138 provided datasets for specific examples).

139 *2.2.2 Auxiliary files*

140 The batch script uses the MICA toolbox's pre-existing approach to associate several
141 files with each image to be analysed. In addition to the already existing need for a folder with
142 the specified ROIs and the corresponding .mspec of each image, the batch script requires two
143 additional text files specifying the rotation of each image and the cone mapping function (see
144 the manual for details). These allow users to standardise the orientation of all images before
145 analysis and permit multiple cone mapping functions within the same dataset (e.g., multiple
146 lighting environments).

147 *2.2.3. Batch GUI*

148 The batch script allows the user to specify which of the available analyses they want
149 to conduct and what settings they want to use. The Batch GUI consists of a central interface
150 and a suite of dedicated, tool-specific GUIs activated by user choices. These tool-specific GUIs
151 provide helpful information and allow running all analyses with independent settings.
152 Furthermore, the batch script remembers past user input and will pre-fill the previous choices
153 by the user, helping to test settings and maintain track of chosen settings in case of
154 incomplete runs. See the manual and worked examples for detailed instructions on how to
155 use them. Users familiar with QCPA and MICA toolbox will find many GUIs similar or identical
156 to existing ones.

157 2.3. Workflow recommendations

158 The Batch GUI, manual and worked examples contain a suite of recommendations in
159 addition to those in van den Berg et al. (2020) and the empiricalimaging.com website that
160 users might find informative. We recommend that the user validate the numerical accuracy
161 of the batch script output on a small dataset before analysing large datasets. This also allows
162 the user to confirm that the folder structure and auxiliary files have been arranged correctly.
163 As processing large datasets can take many hours, days, or even weeks, we recommend that
164 the user consider running sub-sections of larger datasets on different instances, such as
165 multiple computers, servers, or cloud-based computing facilities. This will ultimately save
166 time and prevent data loss while helping to detect faulty files in large datasets faster.

167 Like all approaches to visual modelling, image analysis with QCPA requires many
168 different input choices by the user. Keeping a record of these choices is crucial for two
169 reasons. First, it allows the user to keep track of settings, facilitating record-keeping and
170 collaboration. Second, it provides for the publication of repeatable research. The latter is

171 often an issue in visual modelling studies, as modelling choices are often poorly documented
172 (White et al. 2015). To this end, the QCPA batch script keeps various detailed log files in the
173 data output. These can easily be added to the supplementary information of any publications
174 or uploaded/shared as part of the data.

175

176 **3. R script library**

177 We provide a set of tailored R-scripts enabling the extraction and compilation of batch QCPA
178 output data into .csv tables that can be used for downstream analysis. The efficient handling
179 of QCPA batch output is a crucial step in using QCPA at larger scales, as various outputs are
180 stored in files across a dedicated data set structure. The scripts and functions we provide here
181 are formulated and commented on in a way that is aimed to facilitate modification by the
182 user. These scripts and functions are intended to provide guidance for people with limited
183 experience in managing large and complex datasets. However, these scripts and functions
184 also aim to provide a starting point for more versed users to tailor their own data analysis
185 pipelines. A detailed manual for using the R scripts can be found in the electronic Supplement
186 or here: https://github.com/CaraConradsen/QCPA-r-script.

187 The R environment provided includes the following functionality:

188 1. Compiling CAA, VCA, BSA, and particle analysis data.

189 2. Compiling data from LEIA

190 3. Compiling data from GabRat

191 4. Compiling data from Particle Analysis and secondary data output

192 5. Merging, data cleaning and reshaping various subsets of data

6. Exploratory data visualisation and analysis

7. Exporting compiled files

4. Worked examples

We provide detailed worked examples with test data and corresponding toolbox files, such as cone mapping models. These aim to highlight the functionality of the batch script and allow users to familiarise themselves with examples related to their needs. Instructions for the worked examples can be found in the electronic Supplement or here: https://github.com/cedricvandenberg/QCPA-batch-script.

4.1 Trichromatic example without UV (sea slugs seen by a triggerfish, *Rhinecanthus aculeatus*)

Here, we showcase the use of the batch script by applying all modules in the script to investigate how a triggerfish (*R. aculeatus*) perceives two species of cryptic sea slugs (*Aphelodoris varia, Aplysia sp.*) from the east coast of Australia (Nelson Bay, NSW) photographed underwater (Fig. 1).

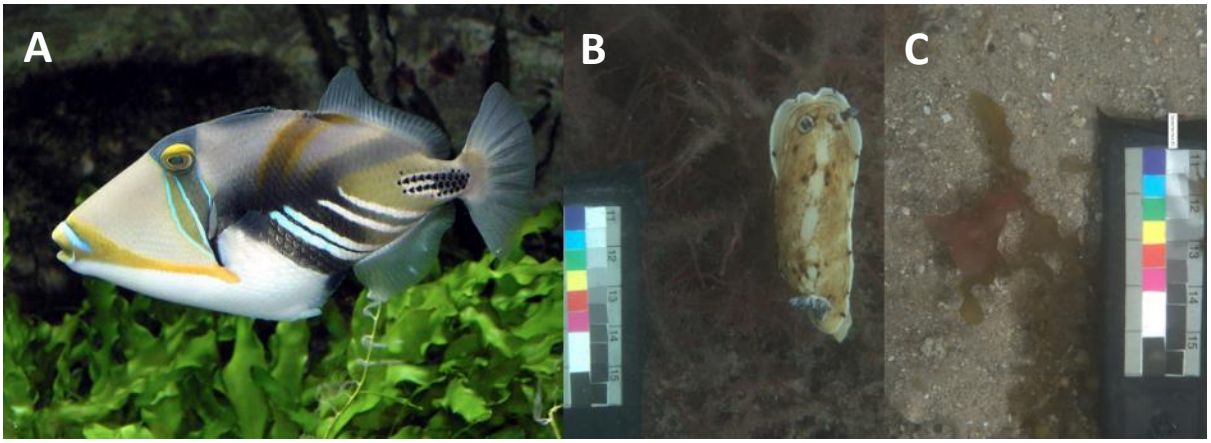

Figure 1 A) *Rhinecanthus aculeatus* (A. Pingstone, public domain). Calibrated images from the test dataset of *Aphelodoris varia* (B) and *Aplysia sp.* (C) by CvdB.

208 4.2 Tetrachromatic example with UV (spiders seen by a bluetit, *Parus major*)

209 This example uses all batch script modules to analyse cryptic spiders (*Tamopsis brisbanensis*)

210 found on eucalypt trees in Sydney, NSW (Australia) (Fig. 2).

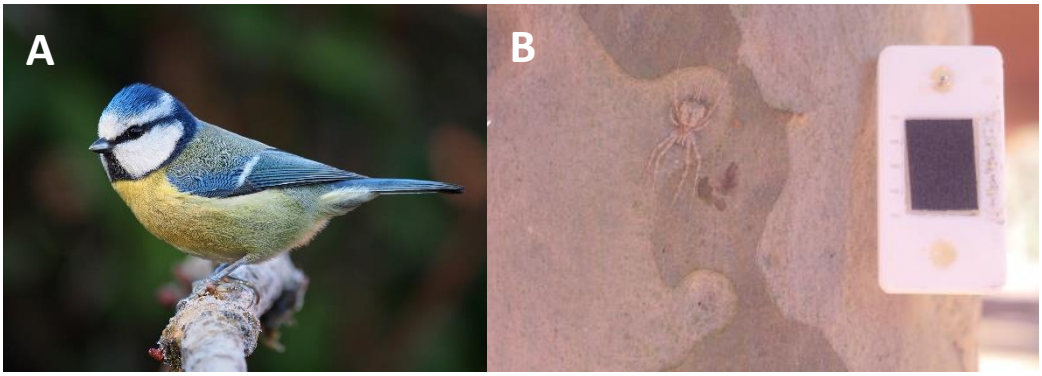

Figure 2 A) *Parus major* (© Francis C. Franklin / CC-BY-SA-3.0); B) Calibrated image of *Tamopsis brisbanensis,* taken by A. Aceves (with permission).

211

212 **5. Discussion**

213 We provide a user-friendly, open-source batch processing extension to QCPA, allowing for the

214 semi-automated analysis of large datasets within the QCPA framework. This greatly facilitates

215 the full-scale use of the analytical power of the framework as intended by its original design

216 (van den Berg et al. 2020b). The open-source code of the script is presented and structured

217 with the deliberate purpose of enabling users to create customised versions of the batch

218 script, which might be more suitable to their intended purposes. The batch script comes with

219 a detailed manual guiding users in preparing and analysing the image data. We further

220 provide three worked examples and corresponding test data, covering many contexts where

221 users will likely use the script. Lastly, the user is provided with a set of R-scripts aiding in the

222 downstream analysis of the data generated by the batch script. These scripts are presented

223 and structured in a way that facilitates users to customise them to their needs.

The script enables the semi-automated analysis of large image datasets, computing hundreds of colour pattern descriptors at multiple viewing distances for dozens, hundreds, or even thousands of observations. This has previously been impossible due to the time-consuming manual processing of images in QCPA. As QCPA and other colour pattern analyses using visual modelling approaches (e.g. Maia et al., 2019; Troscianko and Stevens, 2015) merely relate to approximations of early-stage visual processing (e.g. Endler and Mielke, 2005; Marr, 2010; van den Berg et al., 2020a; van den Berg et al., 2020b; Vorobyev and Osorio, 1998), considering the context and task-dependent relationships with and between a broad range of descriptors is critical (see van den Berg et al., 2022 for discussion). Thus, the presented batch script provides a crucial step towards such analytical approaches.

As QCPA provides a conceptual approach towards combining visual modelling and colour pattern analyses, automating workflows and advancing customisability represent essential stepping stones towards the continued development and refinement of the framework and its contribution towards alternative methods. By providing an inviting, well-annotated, intuitive, open-access and open-source environment, we hope to facilitate the application of and, importantly, eventual modifications and additions to the framework and its combined use with other existing methodologies. Cumulatively, this round-table philosophy will not only advance the development and testing of methods in the field but, crucially, will facilitate the increased critical understanding of and familiarity with existing methodology among members of the field. On a more tangible note, this script can easily be adjusted to enable the analysis of different data structures or include processing steps and analyses available in QCPA or the MICA toolbox currently not considered in the script. The R scripts and functions provided here provide useful guidance for users with variable levels of experience in data wrangling. However, more extensive forms of such customised data

extraction functions will provide an important step towards the dynamic use of large-scale QCPA output across multiple colour pattern analysis frameworks in the R environment, such as PAVO.

6. Acknowledgements

We thank Alfonso Aceves for providing the test data for the UV examples.

7. Declarations

7.1 Funding

This work was supported by the Australian Research Council (FT190199313 to KLC), Hermon Slade Foundation grant to TW (HSF20082), and a UQ Research Stimulus (Allocation Two) Fellowship awarded to CvdB. NC was supported as a CZI Imaging Scientist by grant number 2020-225648 from the Chan Zuckerberg Initiative DAF, an advised fund of the Silicon Valley Community Foundation.

7.2 Conflicts of interest

The authors declare no conflict of interest.

7.3. Ethics approval

Not applicable

7.4 Consent to participate

Not applicable

7.5 Consent for publication

All authors consent to the publication of the manuscript and associated documents.

7.6 Availability of data and material

All data and supplementary materials are available at the respective repositories listed below.

7.7 Code availability

All code is available at the respective repositories listed below.

8. Author's contributions

CvdB – Overall concept (lead), testing, coding, original draft, manuscript writing & review, Funding, NC –coding batch script (lead), Manuscript review, CC – R scripts (lead), Manuscript review, TEW – R scripts, Manuscript review, Funding, KLC – Original concept, Funding, Manuscript review.

9. Supplementary data

Batch script ImageJ plugin, Batch script manual, worked examples guide:

https://github.com/cedricvandenberg/QCPA-batch-script

R scripts and functions:

https://github.com/CaraConradsen/QCPA-r-script

Test data for the worked examples:

https://doi.org/10.48610/3cdcc1f

10. References

Arnold K, Gosling J, Holmes D (2005) The Java programming language. Addison Wesley Professional

Caves EM, Johnsen S (2018) AcuityView: An r package for portraying the effects of visual acuity on scenes observed by an animal. Methods Ecol Evol 9:793–797. https://doi.org/10.1111/2041-210X.12911

Chan IZW, Stevens M, Todd PA (2018) PAT- GEOM: A Software Package for the Analysis of Animal Patterns. Methods Ecol Evol 2041–210X.13131. https://doi.org/10.1111/2041-210X.13131

Endler JA (2012) A framework for analysing colour pattern geometry: Adjacent colours. Biol J Linn Soc 107:233–253. https://doi.org/10.1111/j.1095-8312.2012.01937.x

Endler JA (1991) Interactions between predators and prey. In: Krebs JR, Davies NB (eds) Behavioural Ecology. Blackwell Scientific, Oxford, pp 169–196

Endler JA, Mielke PW (2005) Comparing entire colour patterns as birds see them. Biol J Linn Soc 86:405–431. https://doi.org/10.1111/j.1095-8312.2005.00540.x

Gawryszewski FM (2018) Color vision models: Some simulations, a general n -dimensional model, and the colourvision R package. Ecol Evol 8:8159–8170. https://doi.org/10.1002/ece3.4288

Godfrey D, Lythgoe JN, Rumball D a (1987) Zebra stripes and tiger stripes: the spatial frequency distribution of the pattern compared to that of the background is significant in display and crypsis. Biol J Linn Soc 32:427–433. https://doi.org/10.1111/j.1095-8312.1987.tb00442.x

Kemp DJ, Herberstein ME, Fleishman LJ, et al (2015) An integrative framework for the appraisal of coloration in nature. Am Nat 185:705–724. https://doi.org/10.1086/681021

Maia R, Gruson H, Endler JA, White TE (2019) pavo 2: New tools for the spectral and spatial analysis of colour in r. Methods Ecol Evol 0–2. https://doi.org/10.1111/2041-210X.13174

Marr D (2010) Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. The MIT Press

Nokelainen O, Scott-Samuel NE, Nie Y, et al (2021) The giant panda is cryptic. Sci Rep 11:1–10. https://doi.org/10.1038/s41598-021-00742-4

R Core Team (2021) R: A Language and Environment for Statistical Computing

Renoult JP, Kelber A, Schaefer HM (2017) Colour spaces in ecology and evolutionary biology. Biol Rev 92:292–315. https://doi.org/10.1111/brv.12230

Rodríguez-Morales D, Tapia-McClung H, Robledo-Ospina LE, Rao D (2021) Colour and motion affect a dune wasp's ability to detect its cryptic spider predators. Sci Rep 11:1–9. https://doi.org/10.1038/s41598-021-94926-7

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675. https://doi.org/10.1038/nmeth.2089

Stoddard MC, Kilner RM, Town C (2014) Pattern recognition algorithm reveals how birds evolve individual egg pattern signatures. Nat Commun 5:1–10. https://doi.org/10.1038/ncomms5117

Stoddard MC, Osorio D (2019) Animal Coloration Patterns: Linking Spatial Vision to Quantitative Analysis. Am Nat 193:000–000. https://doi.org/10.1086/701300

Troscianko J, Stevens M (2015) Image calibration and analysis toolbox – a free software suite for objectively measuring reflectance, colour and pattern. Methods Ecol Evol 6:1320–1331. https://doi.org/10.1111/2041-210X.12439

Van Belleghem SM, Papa R, Ortiz-Zuazaga H, et al (2018) Patternize : An R package for quantifying color pattern variation. Methods Ecol Evol 9:390–398. https://doi.org/10.1111/2041-210X.12853

van den Berg CP, Endler JA, Papinczak DEJ, Cheney KL (2022) Using colour pattern edge contrast statistics to predict detection speed and success in triggerfish (*Rhinecanthus aculeatus*). J Exp Biol. https://doi.org/10.1242/jeb.244677

van den Berg CP, Hollenkamp M, Mitchell LJ, et al (2020a) More than noise: Context-dependant luminance contrast discrimination in a coral reef fish (*Rhinecanthus aculeatus*). J Exp Biol 223:jeb.232090. https://doi.org/10.1242/jeb.232090

van den Berg CP, Troscianko J, Endler JA, et al (2020b) Quantitative Colour Pattern Analysis (QCPA): A comprehensive framework for the analysis of colour patterns in nature. Methods Ecol Evol 11:316–332. https://doi.org/10.1111/2041-210X.13328

van Rossum G (1995) Python tutorial, Technical Report CS-R9526. Cent voor Wiskd en Inform

Vorobyev M, Osorio D (1998) Receptor noise as a determinant of colour thresholds. Philos Trans R Soc B Biol Sci 265:351–358. https://doi.org/10.1098/rspb.1998.0302

White TE, Dalrymple RL, Noble DWA, et al (2015) Reproducible research in the study of

biological coloration. Anim Behav 106:51–57.
https://doi.org/10.1016/j.anbehav.2015.05.007