bioRxiv preprint doi: https://doi.org/10.1101/121962; this version posted March 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

`Patternize`: An `R` package for quantifying color pattern variation

Steven M. Van Belleghem[1,2,3], Riccardo Papa[2], Humberto Ortiz-Zuazaga[2], Frederik Hendrickx[4,5], Chris D. Jiggins[1], W. Owen McMillan[6] and Brian A. Counterman[3]

[1.] Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, United Kingdom

[2.] Department of Biology, Center for Applied Tropical Ecology and Conservation, University of Puerto Rico, Rio Piedras, Puerto Rico

[3.] Department of Biological Sciences, Mississippi State University, 295 Lee Boulevard, Mississippi State, MS 39762, USA

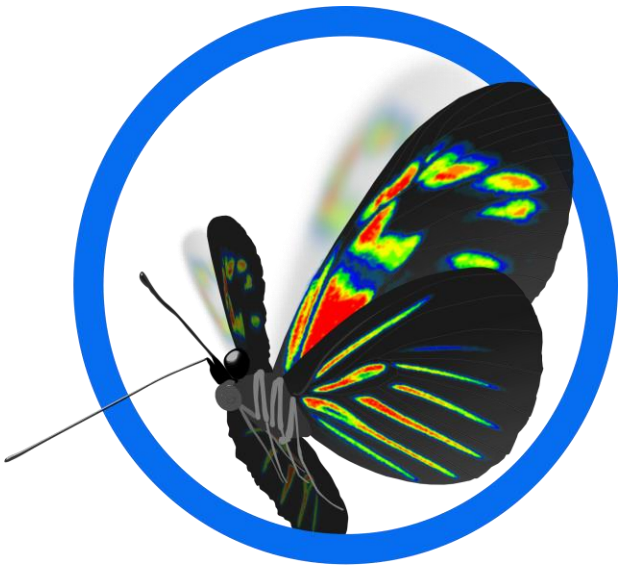
[4.] Terrestrial Ecology Unit, Biology Department, Ghent University, Gent, Belgium

[5.] Royal Belgian Institute of Natural Sciences, Brussel, Belgium

[6.] Smithsonian Tropical Research Institute, Apartado 0843-03092, Panamá, Panama

Keywords: color patterns, heat maps, landmarks, image registration, image segmentation

Corresponding author: vanbelleghemsteven@hotmail.com



bioRxiv preprint doi: https://doi.org/10.1101/121962; this version posted March 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Summary

1. The use of image data to quantify, study and compare variation in the colors and patterns of organisms requires the alignment of images to establish homology, followed by color-based segmentation of images. Here we describe an R package for image alignment and segmentation that has applications to quantify color patterns in a wide range of organisms.

2. `patternize` is an `R` package that quantifies variation in color patterns obtained from image data. `patternize` first defines homology between pattern positions across specimens either through fixed landmarks or automated image registration. Pattern identification is performed by categorizing the distribution of colors using either an RGB threshold or unsupervised image segmentation.

3. We demonstrate that `patternize` can be used for quantification of the color patterns in a variety of organisms by analyzing image data for butterflies, guppies and spiders. Image data can be compared between sets of specimens, visualized as heatmaps and analyzed using principal component analysis (PCA).

4. `patternize` has potential applications for fine scale quantification of color pattern phenotypes in population comparisons, genetic association studies and investigating the genetic basis of color pattern expression across a wide range of organisms.

bioRxiv preprint doi: https://doi.org/10.1101/121962; this version posted March 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Natural populations often harbor great phenotypic diversity. Variations in color and patterns are of the more vivid examples of morphological variability in nature. Taxa as diverse as spiders (De Busschere *et al.* 2012; Cotoras *et al.* 2016), insects (Katakura *et al.* 1994; Williams 2007), fish (Endler 1983; Houde 1987), amphibians and reptiles (Calsbeek *et al.* 2008; Stapley *et al.* 2011; Ng *et al.* 2012; Allen *et al.* 2013; Rabbani *et al.* 2015), mammals (Hoekstra *et al.* 2006; Nekaris & Jaffe 2007; Allen *et al.* 2015) and plants (Clegg & Durbin 2000; Mascó *et al.* 2004) display natural variation in pigment or structural colorations. The distribution of colors in specific patterns play an important role in mate preference (Endler 1983; Kronforst *et al.* 2006), thermal regulation (Forsman *et al.* 2002), aposematism (Rojas *et al.* 2015) and crypsis (Nosil & Crespi 2006) and represent evolutionary adaptations that in many cases have promoted diversification within lineages. In butterflies, for example, the evolution of scale cells and the spatial coordinate system that controls wing pigmentation has played an important role in their diversification into over 16,000 species of butterflies and 160,000 species of moths (Nijhout 1991).

Measuring phenotypic variation in organismal color patterns can provide insights into their underlying developmental and genetic architecture (Klingenberg 2010). However, precisely quantifying color pattern variation is challenging. Consistent comparisons of color patterns from images requires the (1) homologous alignment and (2) color-based segmentation of the images. Homologous alignment can be performed by transforming one image onto another. This transformation can be obtained from fixed sets of landmarks or advanced image registration techniques, which can be stored and utilized to align color patterns extracted from the images. Image segmentation concerns the categorization of pixels by color. Previously, examples of color pattern quantification have been developed for *Heliconius* butterflies (Le Poul *et al.* 2014) and primates (Allen *et al.* 2015). However, these applications are not easily accessible for use in other organisms. Similarly, advanced solutions are available for biomedical image analysis (Modat *et al.* 2010a; Schindelin *et al.* 2012, 2015), but are not tailored towards quantifying color pattern variation.

Here, we develop `patternize`, an approach to quantification of color pattern variation from 2D images using the R statistical computing environment (Core Team 2013). The package provides utilities to extract, transform and superimpose color patterns. While transformations are obtained

bioRxiv preprint doi: https://doi.org/10.1101/121962; this version posted March 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

from fixed landmarks or automated image registration, color-based segmentation of the patterns is performed by using threshold RGB (Red, Blue and Green) values or unsupervised classification of pixels into a set of clusters. By extracting and aligning color patterns from image data of large numbers of samples, `patternize` provides quantitative measures of variation in color patterns that can be used for population comparisons, genetic association studies and investigating dominance and epigenetic interactions of color pattern expression in a wide range of organisms. We demonstrate the utility of the package with *Heliconius* butterflies and more challenging guppy fish and Galápagos wolf spiders.

Alignment

Superimposing color patterns to quantify variation in their expression requires the homologous alignment of the anatomical structures they occur in. Image transformations for this alignment can be obtained from landmark based transformations and image registration techniques. *Landmark based transformations* use discrete anatomical points that are homologous among individuals in the analysis. Non-rigid, but uniform transformations from one set of 'source' landmarks to a set of 'target' landmarks such as *affine* transformations include translation, rotation, scaling and skewing (Hazewinkel 2001). Additionally, non-uniform changes in shape between the source and target landmarks can be accounted for by storing the transformation as if it were 'the bending of a thin sheet of metal', the so-called *thin plate spline* (TPS) transformation (Duchon 1976). Both the affine and TPS transformation can be calculated from fixed sets of landmarks (Figure 1A). We implemented these landmark transformations using utilities provided by the R package `Morpho` (Schlager 2016). Landmarks can be transformed using an arbitrarily chosen reference sample or an average landmark shape obtained from a set of samples. The average landmark shape is obtained by means of Procrustes superimposition of the samples (Goodall 1991).

Alternative to landmark based methods, fast and accurate *image registration* techniques are available for calculating a transformation from a source to target image based on either intensity patterns or features such as points, lines or contours present in the images (Goshtasby 2005) (Figure 1A). We use a computation efficient intensity-based image registration technique implemented in the NiftyReg image registration library (TIG 2016) and made available in R through the `RNiftyReg` package (Clayden *et al.* 2017). This methodology calculates the global

bioRxiv preprint doi: https://doi.org/10.1101/121962; this version posted March 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

transformation of an image by finding correspondences between sub-volumes of the two images (Modat *et al.* 2010a,b). Correspondence is assessed using intensity-based similarity measures and used to calculate the transformation parameters through a least trimmed square (LTS) regression method (Modat *et al.* 2010a,b). The number of corresponding sub-volumes to be included or considered as outliers in the calculation of the transformation can be varied by the user. The global transform calculated by NiftyReg can be rigid (i.e. including translation, rotation and scaling) or affine (i.e. translation, rotation, scaling and skewing). NiftyReg also provides the possibility to perform local nonlinear transformations (Modat *et al.* 2010b). However, these local nonlinear transformations are not used in our implementation because the use of intensity measures would result in warping the color patterns and losing correct assessment of the color pattern variation.

Comparison of the landmark and image registration approach applied to the wing color patterns displayed by *Heliconius* butterflies shows that both methods perform well (Figure 1B). The TPS transformation used in the landmark approach resulted in a better fit to the internal structures of the wing (i.e. wing veins). The slight offset between the color pattern and vein position in the image registration approach likely resulted from a bias in the linear transformation towards aligning the outline of the wing and not including non-uniform changes in shape within the wing. While the landmark based approach is computationally faster, automated image registration removes the need for labor intensive setting of landmarks on each image. Moreover, image registration reduces any variation introduced by differences in how users manually place image landmarks. Care should be taken, however, with automated image registration, as it can be highly sensitive to artefacts in the background. For cases in which the background differs starkly from the studied object, functionality is included that allows to remove the background using RGB cutoff values. The package also allows the user to review the image registration progress to assess the quality of the automatic registration.

bioRxiv preprint doi: https://doi.org/10.1101/121962; this version posted March 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

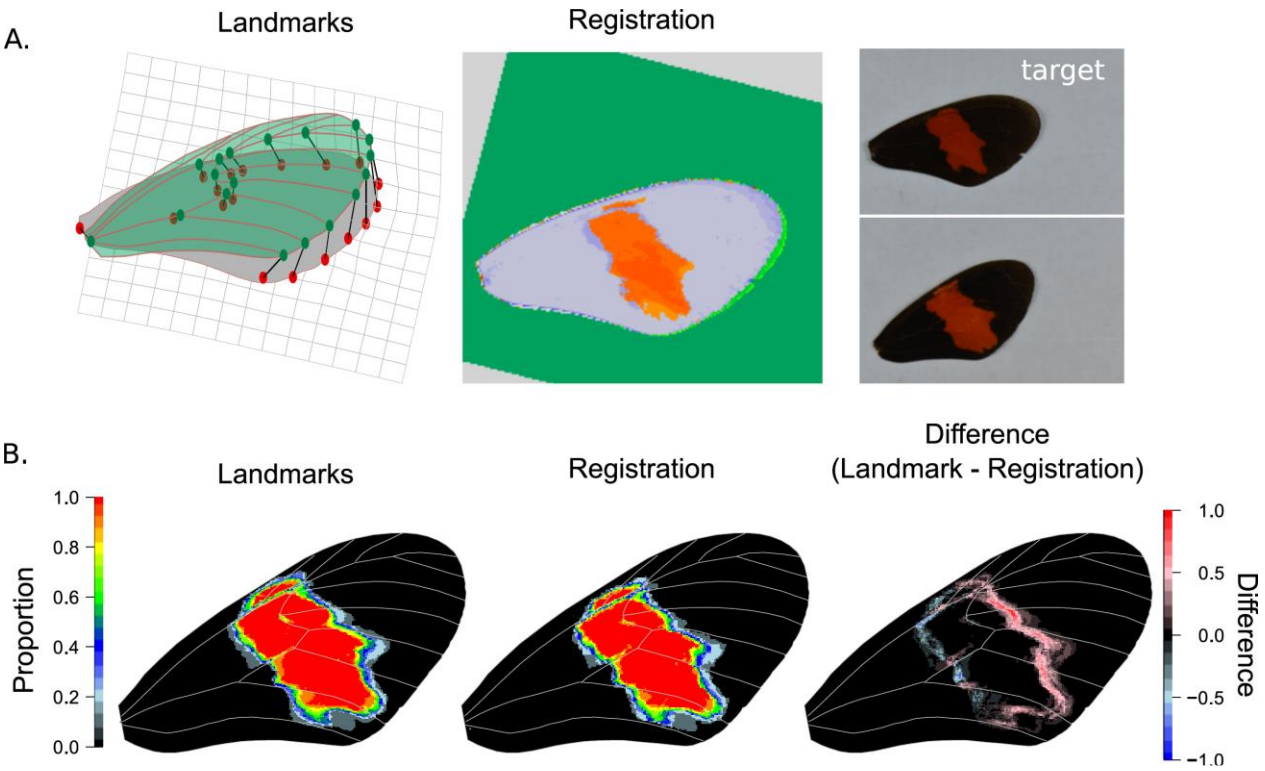


Figure 1. Image transformation and quantification of color pattern variation. (A.) Illustration of transformation strategies of a source (green) image to a target (gray) image. The thin plate spline (TPS) transformation from the source to target landmarks is illustrated by the deformed grid and can be used to transform the image or extracted color pattern. Image registration attempts to find common patterns in images and align the source (green) image to the pixel coordinate system of the target (gray) image. Note the extracted color pattern in red. **(B.)** Example comparison between landmark approach for color pattern alignment for ten butterfly wings of male *Heliconius erato hydara*. For the landmark approach, we used TPS transformation. For the image registration approach, we used affine transformation and 75% of sub-volumes included as inliers.

Color pattern extraction

Studying variation in color patterns requires the correct identification of the color boundaries. Color boundaries can be extracted from images or the trait of interest using RGB thresholds. By selecting pixels within a specified color range (specified as RGB value and offset) we provide a basic image segmentation approach that works well for extracting distinct color patterns. Additionally, for distinct color patterns, the specified RGB value can be iteratively recalculated as the average for the extracted color pixels. This latter approach permits patterns to be easily combined when extracted from sets of images that may have been taken under different light conditions resulting in differences in intensity and contrast. Here, we demonstrated the utility of this approach for color patterns expressed by *Heliconius erato* butterflies and hybrid phenotypes. Our application allows

bioRxiv preprint doi: https://doi.org/10.1101/121962; this version posted March 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to differentiate the two groups of butterflies and indicates overexpression of the color pattern in hybrids that have red instead of yellow scales in their forewing (Figure 2).

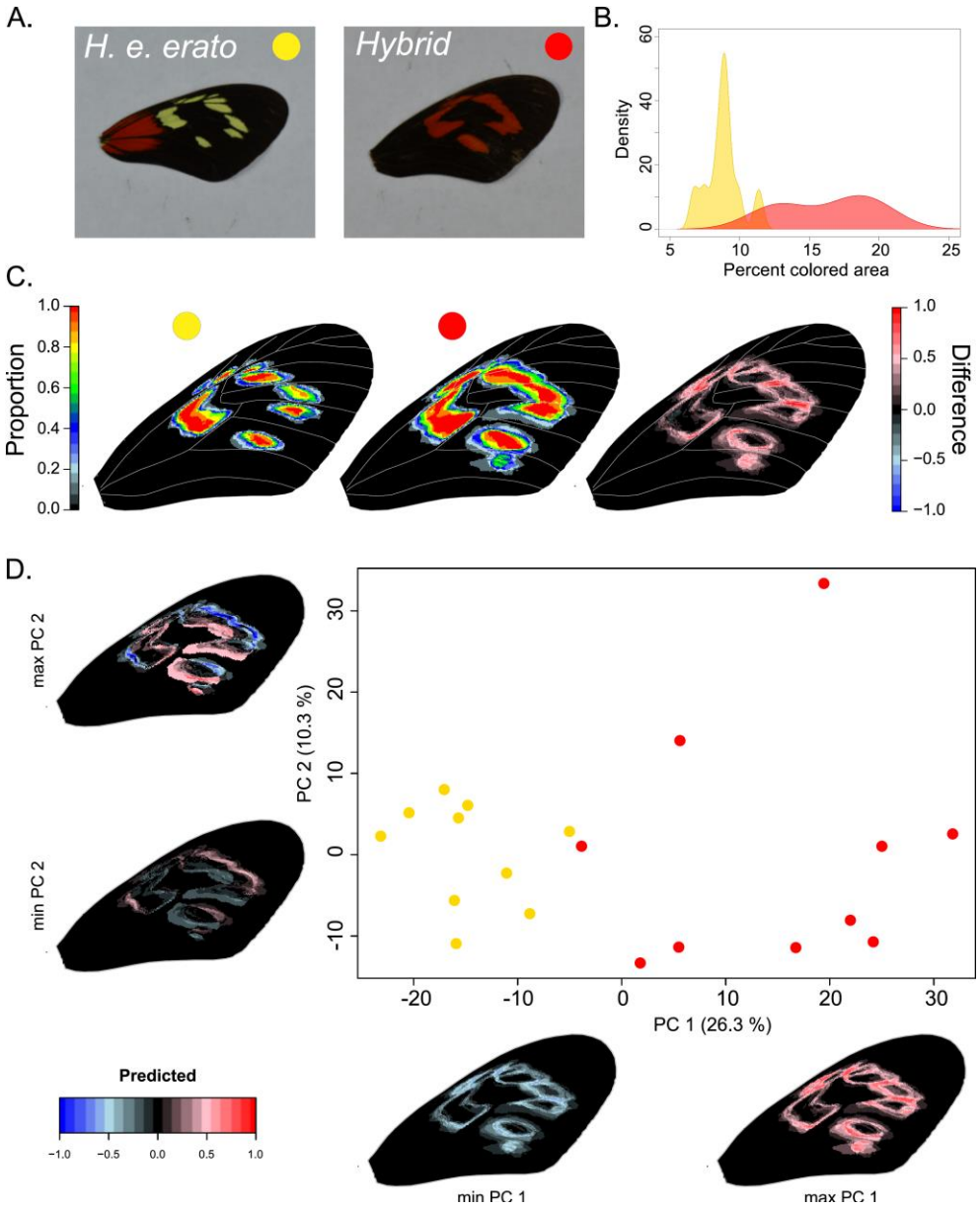


Figure 2. Example of image registration and threshold color extraction in *Heliconius erato erato* (n = 10) and hybrid (n = 10) butterflies (French Guiana). (A.) Example of original images. The hybrid represents a naturally occurring backcross in a hybrid zone with *H. e. hydara* (see Figure 1) that results in red color expression in the forewing band. **(B.)** The percentage of area in the wing expressing yellow in *H. e. erato* and red in the hybrid shows a consistent expansion of the forewing band when red is expressed. **(C.)** Similarly, visualizing the variation in color pattern expression in a heatmap shows a consistently larger pattern in the hybrid phenotypes (*H. e. erato*: left, hybrid: middle, hybrid minus *H. e. erato*: right). **(D.)** Principal component analysis (PCA) confirms that the main axis of variation (PC1) is related to size of the pattern (yellow or red in *H. e. erato* and hybrids, respectively) and separates the *H. e. erato* and hybrid samples. The second principal component (PC2) axis highlights more complex shape differences in the forewing band among the samples as demonstrated by the shape changes of the color patterns along the principal component axis.

bioRxiv preprint doi: https://doi.org/10.1101/121962; this version posted March 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We also implemented an unsupervised approach for color-based image segmentation by using *k*-means clustering (Hartigan & Wong 1979). This algorithm partitions an image into *k* clusters by iteratively assigning each pixel in the image to the cluster that minimizes the distance between the pixel and the cluster centers. Cluster centers are recalculated each iteration by averaging all pixels in the cluster until convergence.

We implemented *k*-means clustering using the R package `stats` (Core Team 2013). Clusters are first obtained from a reference image and then used as initial cluster centers for the further analysis. This allows the program to match clusters that represent the same color pattern in different images. For *k*-means clustering, the number of clusters must be defined manually. For organisms with less distinct pattern boundaries, this is best done by testing different numbers of clusters and choosing a number that best assigns pixels to color patterns.

To characterize variation in color patterns among samples, we implemented linear principal component analysis (PCA). For a specific color or *k*-means color cluster, PCA can be performed on the binary representation of the aligned color pattern rasters obtained from each sample (Figure 2-4). In this matrix, pixel coordinates that have the color of interest in a sample get a value of one, whereas pixel coordinates without the color get the value zero assigned. The variance-covariance matrix obtained from the binary matrix for a color is suitable for PCA to visualize the main variations in color pattern boundaries among or between groups of samples. Subsequently, predicted color pattern changes can be visualized along the principal component (PC) axis (Johnson & Wichern 2007). Positive values present a higher predicted expression of the pattern, whereas negative values present the absence of the pattern. Note that parts of the color patterns that are expressed in all considered samples get a predicted value of zero, as these pixels do not contribute variance for the PCA analysis.

Guppies and spiders

To assess the general utility of our application across taxa, we applied the automated registration and *k*-means clustering approach to groups with more complex body shape and color pattern variation; guppy fish and Galápagos wolf spiders. Males of the guppy (*Poecilia reticulate*) vary greatly in their ornamental patterns that have evolved in response to both natural and sexual

bioRxiv preprint doi: https://doi.org/10.1101/121962; this version posted March 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

selection. Several mutants have been described among male guppies that affect color pattern expression. Manually quantifying the differences in color pattern expression among these mutations has provided valuable insights into the developmental basis and interactions of the involved genes (Kottler *et al.* 2013). Here, we summarized and compared the black and orange color patterns expressed in wild type (WT) versus *golden* mutants of *P. reticulate* males using images obtained from Kottler *et al.* 2013 (images were used from backcrosses obtained from *golden blue* mutant females with heterozygous males from crossing *golden blue* with inbred wild-derived Cumána populations) (Figure 3). All images were aligned to a target image using image registration and colors were *k*-means clustered into seven groups. Before *k*-means color clustering, the background was masked using the outline of the guppy in the target image. Our analysis of the black and orange color cluster strongly matched the description presented in Kottler *et al.* 2013, demonstrating the absence of a posterior orange spot in *golden* mutants backcrossed into a Cumána population genetic background and more diffuse and shifted black ornaments in the *golden* mutants.

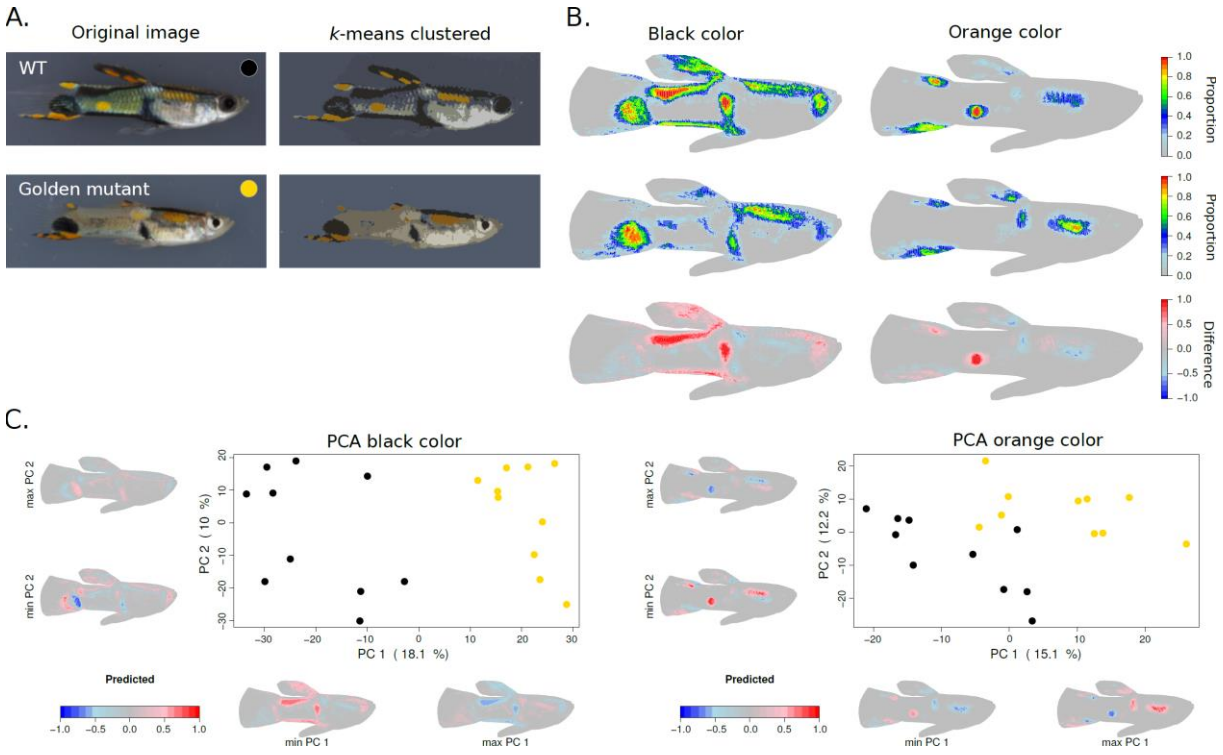


Figure 3. Example of image registration and *k*-means clustering of colors in guppies (*Poecilia reticulata*). (A.) Original image of a wild type (WT) and *golden* mutant guppy and their *k*-means clustered representation (clusters = 7). Clusters are first obtained from a reference image and then used as initial cluster centers for the further analysis. This allows the program to match clusters that represent the same color pattern in different images. **(B.)** Heat maps and difference between WT (n=10) and golden mutant (n=10) for black and orange color clusters. **(D.)** PCA analysis of the pixel matrices obtained for the black (left) and orange (right) color clusters. Images were obtained with permission from (Kottler *et al.* 2013).

bioRxiv preprint doi: https://doi.org/10.1101/121962; this version posted March 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Wolf spiders of the genus *Hogna* inhabit high elevation and coastal habitats on the Galápagos islands Santa Cruz and San Cristobal (De Busschere *et al.* 2010). Despite the phylogenetically close relationship of the high elevation and coastal populations within both islands, morphometric analysis, including measurements of color intensity, have highlighted striking parallel phenotypic divergence between the high elevation and coastal species between the islands (De Busschere *et al.* 2012). Coastal species appear to be paler with a more conspicuous median band on the carapace compared to high elevation species. Here, we demonstrate the robustness of automated image registration by aligning the highly variable images of the wolf spiders (Figure 4). By focusing on correspondence between the images, the automated image registration technique, as implemented in NiftyReg (TIG 2016), manages to align the spider's carapace, which is morphologically the most consistent part in the images. By assigning colors in the spiders to only two clusters, we show a similar pattern as described in De Busschere *et al.* 2012 in which the coastal species show a consistently broader and more conspicuous median band on the carapace and pale lateral bands compared to the high elevation species.

bioRxiv preprint doi: https://doi.org/10.1101/121962; this version posted March 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

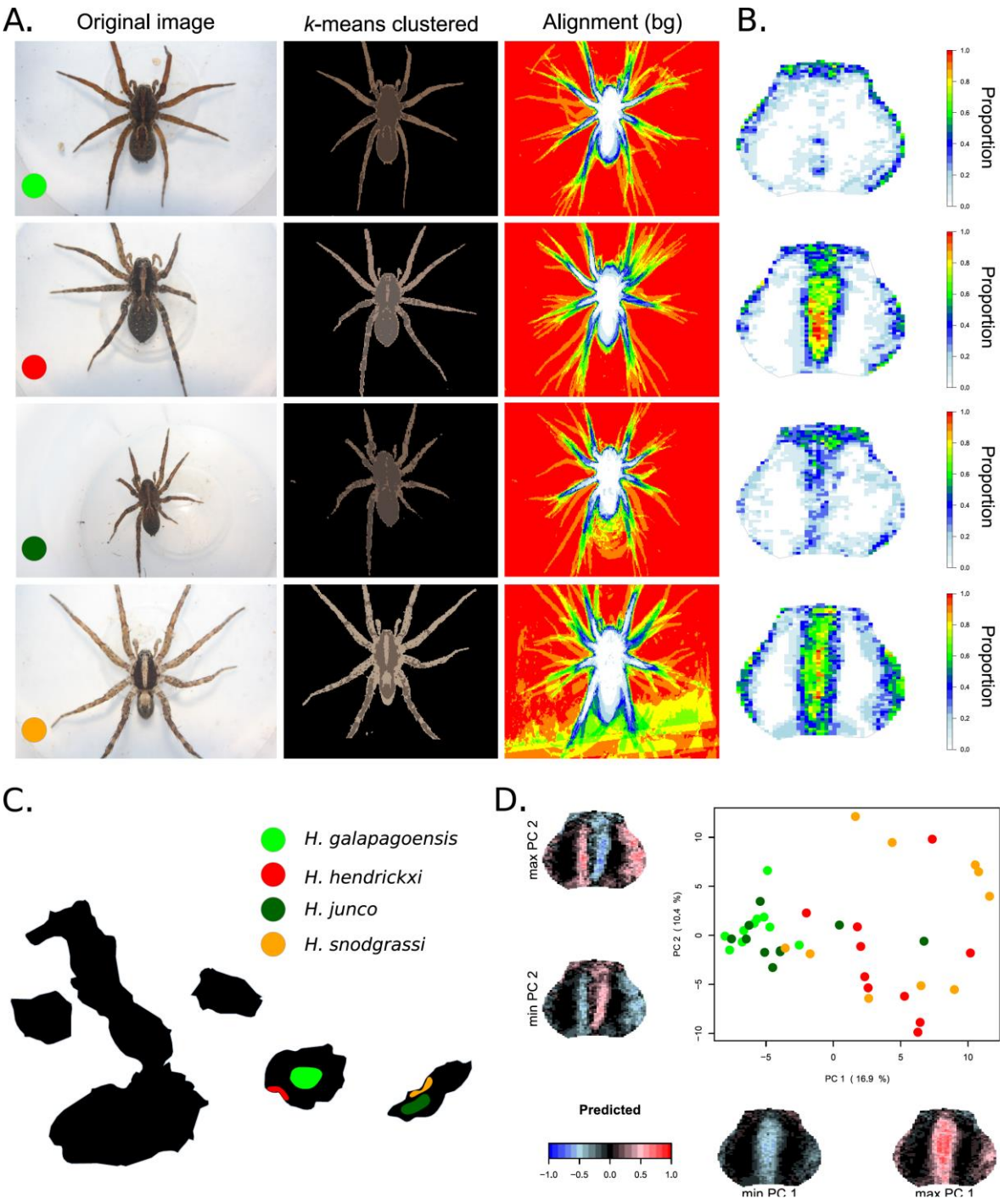


Figure 4. Example of image registration and *k*-means clustering of the color pattern of Galápagos wolf spiders (*Hogna*). (A.) From left to right: example of original image (10 images were used for each species), *k*-means clustered image (k = 3) with removed background and alignment of the color cluster that corresponds to the background (bg) color. **(B.)** Heatmap corresponding to the lightest color cluster focused on the carapace. **(C.)** Map of the Galápagos islands with colors indicating the distribution of four *Hogna* species, two high elevation species (light and dark green) and two coastal species (red and orange). **(D.)** PCA analysis of the pixel matrices obtained for the lightest color cluster demonstrates that the coastal (*H. hendrickxi* and *H. snodgrassi*) and high-elevation (*H. galapagoensis* and *H. junco*) species cluster phenotypically together and share, respectively, the presence and absence of a pale median band on their carapace. Images were obtained with permission from (De Busschere *et al.* 2012).

bioRxiv preprint doi: https://doi.org/10.1101/121962; this version posted March 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Concluding remarks

`patternize` provides an unbiased, fast and user-friendly approach for color pattern analysis that is applicable to a wide variety of organisms. `patternize` allows downsampling of the images, which decreases computation time for both calculating the image transformations and the color pattern extraction. Automated image registration reduces the need for labor intensive landmark setting. However, because automated registration uses intensity patterns in the images, care should be taken by standardizing the experimental setup. Setting correct RGB or cluster parameters may impact results and should be optimized for each analysis. Appropriate RGB and offset values can be obtained, for instance, by extracting RGB histograms for image areas of interest (Schindelin *et al.* 2012). Using few or many *k*-means clusters may, respectively, result in grouping colors of interest or assigning multiple clusters to a single pattern of interest. The output of the main `patternize` functions are `raster` objects (Hijmans 2016) that provide for a wide range of downstream analyses. We provide functions to intersect (masking) the extracted patterns with defined outlines (polygons), to calculate the relative area in which the pattern is expressed and to carry out principal component analysis (PCA). Overall, we hope that this readily accessible R package will provide useful tools for the community of researchers working on color and pattern variation in animals.

Acknowledgments

We kindly thank Jon Clayden for help with implementing RNiftyReg, Stefan Schlager for help with Morpho, Heather Smith for help with digitizing the *Heliconius* images and Verena Kottler and Charlotte De Busschere for allowing us to use the guppy and *Hogna* images. SMVB and BAC were funded by NSF grant DEB-1257839. HOZ was supported in part by NIH grant 5P20GM103475-13. All authors declared that they have no conflict of interest.

bioRxiv preprint doi: https://doi.org/10.1101/121962; this version posted March 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Data accessibility

The package is available as library("patternize") on CRAN (cran.r-project.org/web/packages/patternize). The code, ongoing developments and data and code used for the examples can be accessed through GitHub (github.com/StevenVB12/patternize; github.com/StevenVB12/patternize-examples). Bug reports and feature requests can be sent using the GitHub issue tracker.

Author contributions

SMVB and BAC conceived the development of the package. SMVB wrote the code. SMVB, RP and BAC wrote the manuscript. HOZ helped improving the code. SMVB, BAC, FH and RP conceived data acquisition. HOZ, FH, CDJ and WOM contributed helpful comments for building the package and writing the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

References

Allen, W.L., Baddeley, R., Scott-samuel, N.E. & Cuthill, I.C. (2013). The evolution and function of pattern diversity in snakes. *Behavioral Ecology*, **24**, 1237–1250.

Allen, W.L., Higham, J.P. & Allen, W.L. (2015). Assessing the potential information content of multicomponent visual signals: a machine learning approach. *Proceedings of the Royal Society B*, **282**, 20142284.

Calsbeek, R., Bonneaud, C. & Smith, T.B. (2008). Differential fitness effects of immunocompetence and neighbourhood density in alternative female lizard morphs. 103–109.

Clayden, J., Modat, M., Presles, B., Anthopoulos, T. & Daga, P. (2017). RNiftyReg: Image Registration Using the 'NiftyReg' Library. R package version 2.5.0. https://CRAN.R-project.org/package=RNiftyReg.

Clegg, M.T. & Durbin, M.L. (2000). Flower color variation: A model for the experimental study of evolution. *Proceedings of the National Academy of Sciences*, **97**, 7016–7023.

Core Team, R. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Cotoras, D.D., Brewer, M.S., Croucher, P.J.P., Geoff, S., Lindberg, D.R. & Gillespie, R.G. (2016). Convergent evolution in the colour polymorphism of *Selkirkiella* spiders (Theridiidae) from the South American temperate rainforest. *Biological Journal of the Linnean Societyo*.

De Busschere, C., Baert, L., Van Belleghem, S.M., Dekoninck, W. & Hendrickx, F. (2012). Parallel phenotypic evolution in a wolf spider radiation on Galápagos. *Biological Journal of the Linnean Society*, **106**, 123–136.

De Busschere, C., Hendrickx, F., Van Belleghem, S.M., Backeljau, T., Lens, L. & Baert, L. (2010). Parallel habitat specialization within the wolf spider genus *Hogna* from the Galápagos. *Molecular ecology*, **19**, 4029–4045.

Duchon, J. (1976). Splines minimizing rotation invariant semi-norms in Sobolev spaces. *Volume 571 of the series Lecture Notes in Mathematicss* (eds W. Schempp & K. Zeller), pp. 85–100. Springer.

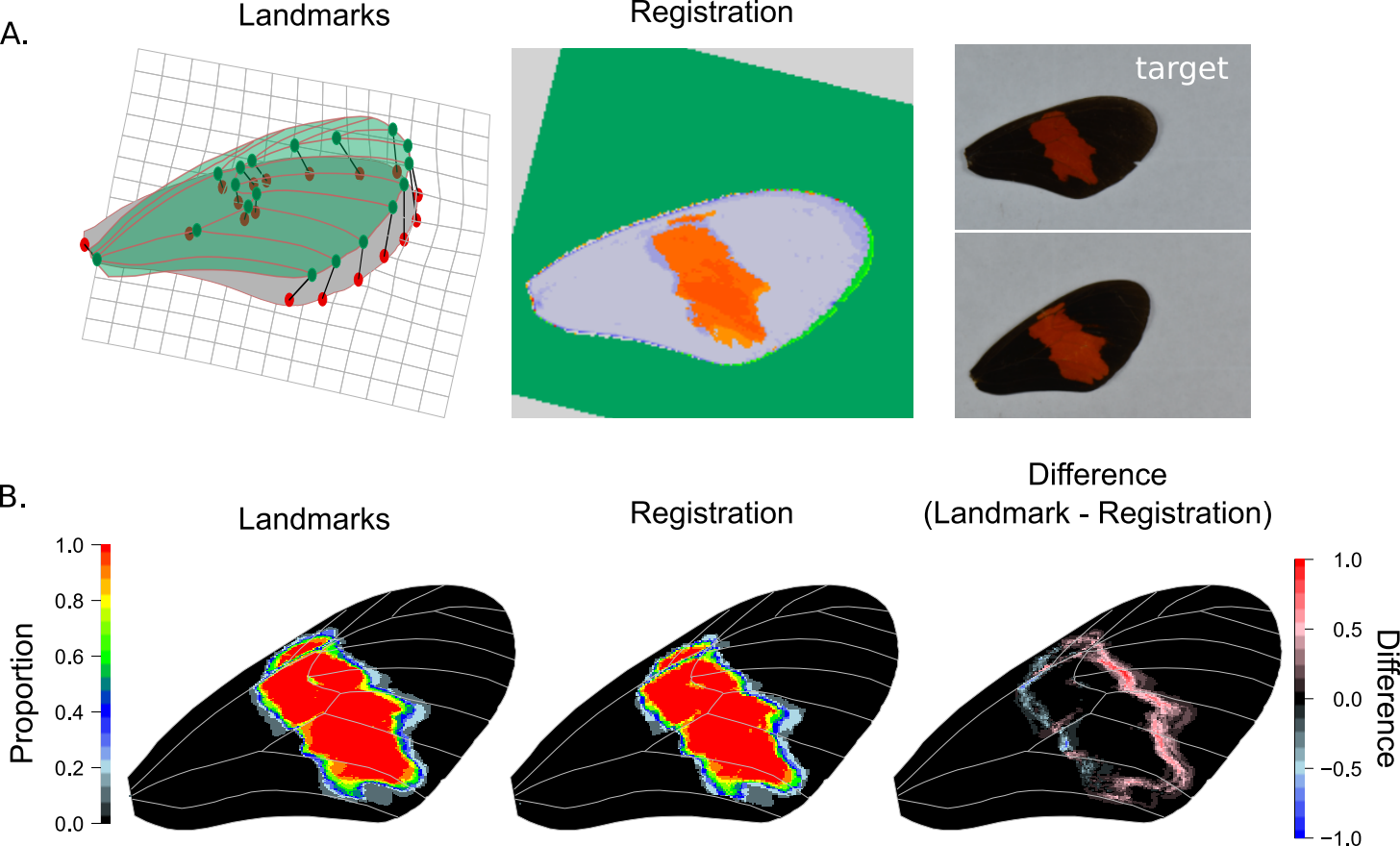
Endler, J.A. (1983). Natural and sexual selection on color patterns in poeciliid fishes. *Environmental Biology of Fishes*, **9**, 173–190.

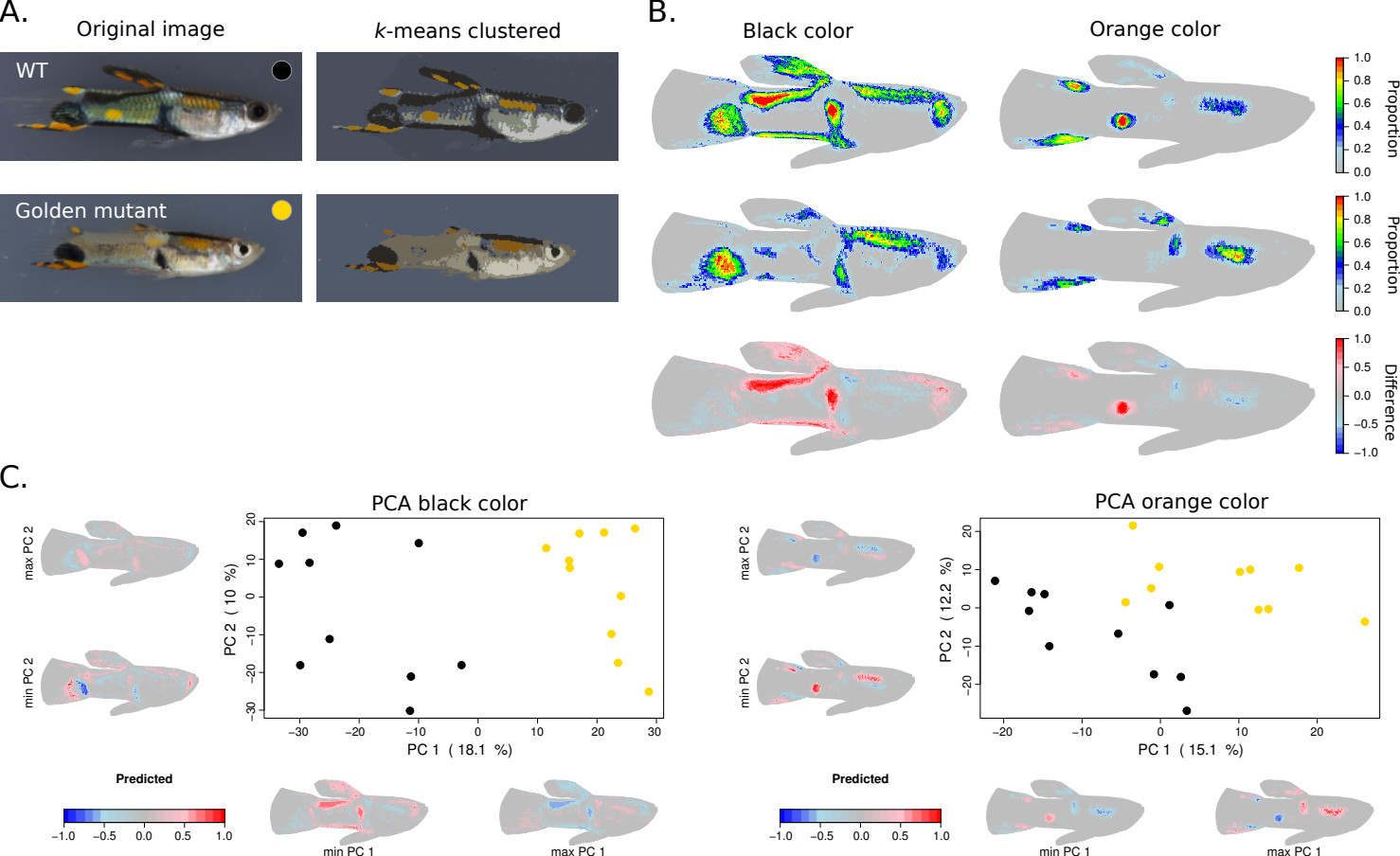
Forsman, A., Ringblom, K., Civantos, E. & Ahnesjö, J. (2002). Coevolution of color pattern and thermoregulatory behavior in polymorphic pygmy grasshoppers *Tetrix undulata*. *Evolution*, **56**, 349–360.

Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society. Series B*, **53**, 285–339.

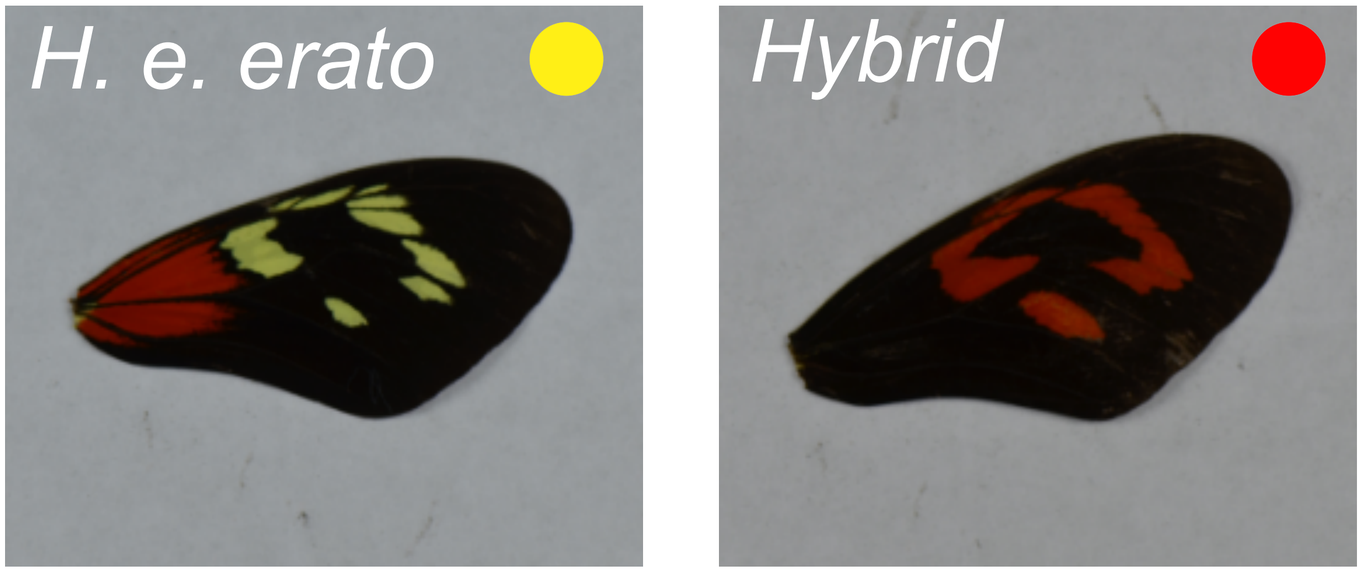
bioRxiv preprint doi: https://doi.org/10.1101/121962; this version posted March 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

277 Goshtasby, A. (2005). *2-D and 3-D image registration: for medical, remote sensing, and industrial applications*.
278 Wiley, Hoboken, NJ.
279 Hartigan, J.A. & Wong, M.A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal*
280 *Statistical Society. Series B*, **28**, 100–108.
281 Hazewinkel, M. (Ed.). (2001). Affine transformation. *Encyclopedia of Mathematics*. Springer.
282 Hijmans, R.J. (2016). raster: Geographic data analysis and modeling. R package version 2.5-8. http://cran.r-
283 project.org/package=raster.
284 Hoekstra, H.E., Hirschmann, R.J., Bundey, R. a, Insel, P. a & Crossland, J.P. (2006). A single amino acid mutation
285 contributes to adaptive beach mouse color pattern. *Science*, **313**, 101–104.
286 Houde, A.E. (1987). Mate choice based upon naturally occurring color-pattern variation in a guppy population.
287 *Evolution*, **41**, 1–10.
288 Johnson, R.A. & Wichern, D.W. (2007). *Applied multivariate statistical analysis*, 6th Edition. Pearson.
289 Katakura, H., Saitoh, S., Nakamura, K. & Abbas, I. (1994). Multivariate analyses of elytral spot patterns in the
290 phytophagous ladybird beetle *Epilachna vigintioctopunctata* (Coleoptera, Coccinellidae) in the province of
291 Sumatra Barat, Indonesia. *Zoological science*, **11**, 889–894.
292 Klingenberg, C.P. (2010). Evolution and development of shape: integrating quantitative approaches. *Nature*, **11**,
293 623–635.
294 Kottler, V.A., Fadeev, A., Weigel, D. & Dreyer, C. (2013). Pigment pattern formation in the guppy, *Poecilia*
295 *reticulata*, involves the Kita and Csf1ra receptor tyrosine kinases. *Genetics*, **194**, 631–646.
296 Kronforst, M.R., Young, L.G., Kapan, D.D., McNeely, C., O'Neill, R.J. & Gilbert, L.E. (2006). Linkage of butterfly
297 mate preference and wing color preference cue at the genomic location of *wingless*. *Proceedings of the*
298 *National Academy of Sciences of the United States of America*, **103**, 6575–6580.
299 Le Poul, Y., Whibley, A., Chouteau, M., Prunier, F., Llaurens, V. & Joron, M. (2014). Evolution of dominance
300 mechanisms at a butterfly mimicry supergene. *Nature Communications*, **5**, 1–8.
301 Mascó, M., Noy-Meir, I. & Sérsic, A.N. (2004). Geographic variation in flower color patterns within *Calceolaria*
302 *uniflora* Lam . in Southern Patagonia. *Plant Systematics and Evolution*, **244**, 77–91.
303 Modat, M., Mcclelland, J. & Ourselin, S. (2010a). Lung registration using the NiftyReg package. *Medical Image*
304 *Analysis for the Clinic: A Grand Challenge, Workshop Proc. from MICCAI 2010*, 33–42.
305 Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C. & Ourselin, S. (2010b).
306 Fast free-form deformation using graphics processing units. *Computer Methods and Programs in Biomedicine*,
307 **98**, 278–284.
308 Nekaris, K.A.I. & Jaffe, S. (2007). Unexpected diversity of slow lorises (*Nycticebus spp.*) within the Javan pet trade:
309 implica- tions for slow loris taxonomy. *Contributions to Zoology*, **76**, 187–196.
310 Ng, J., Landeen, E.L., Logsdon, R.M. & Glor, R.E. (2012). Correlation between *Anolis* lizard dewlap phenotype and
311 environmental variation indicates adaptive divergence of a signal important to sexual selection and species
312 recognition. *Evolution*, **67**, 573–582.
313 Nijhout, H.F. (1991). *The development and evolution of butterfly wing patterns*. Smithsonian Institution Press.
314 Nosil, P. & Crespi, B.J. (2006). Experimental evidence that predation promotes divergence in adaptive radiation.
315 *Proceedings of the National Academy of Sciences*, **103**, 9090–9095.
316 Rabbani, M., Zacharczenko, B., Green, D.M., Abbani, M.O.R. & Acharczenko, B.R.Z. (2015). Color pattern
317 variation in a cryptic amphibian, *Anaxyrus fowleri*. *Journal of Herpetology*, **49**, 649–654.
318 Rojas, B., Valkonen, J. & Nokelainen, O. (2015). Aposematism. Quick guide. *Current biology*, **25**, R350–R351.
319 Schindelin, J., Arganda-carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C.,
320 Saalfeld, S., Schmid, B., Tinevez, J., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P. & Cardona, A.
321 (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, **9**, 676–682.
322 Schindelin, J., Rueden, C.T., Hiner, M.C. & Eliceiri, K.W. (2015). The ImageJ ecosystem: An open platform for
323 biomedical image analysis. *Molecular reproduction & Development*, **82**, 518–529.
324 Schlager, S. (2016). Morpho: Calculations and visualisations related to geometric morphometrics. R package version
325 2.3.1.1. http://cran.r-project.org/package=Morpho.
326 Stapley, J., Wordley, C. & Slate, J. (2011). No evidence of genetic differentiation between anoles with different
327 dewlap color patterns. *Journal of Heredity*, **102**, 118–124.
328 TIG. (2016). Nifty Reg.
329 Williams, P. (2007). The distribution of bumblebee colour patterns worldwide: possible significance for
330 thermoregulation, crypsis, and warning mimicry. *Biological Journal of the Linnean Society*, **92**, 97–118.
331

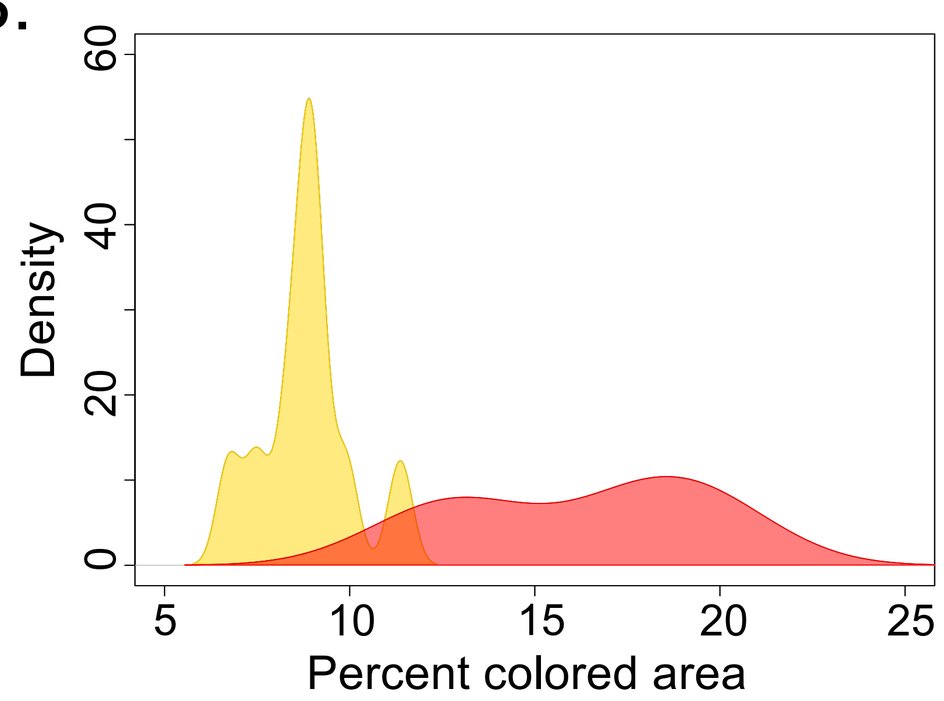

A.
Landmarks
Registration
target
B.
Landmarks
Registration
Difference
(Landmark - Registration)
Proportion
1.0
0.8
0.6
0.4
0.2
0.0
1.0
0.5
0.0
−0.5
−1.0
Difference


A.
Original image
k-means clustered
WT
Golden mutant
B.
Black color
Orange color
Proportion
Proportion
Difference
C.
PCA black color
PCA orange color
max PC 2
min PC 2
PC 2 (10 %)
PC 1 (18.1 %)
PC 2 (12.2 %)
PC 1 (15.1 %)
Predicted
min PC 1
max PC 1

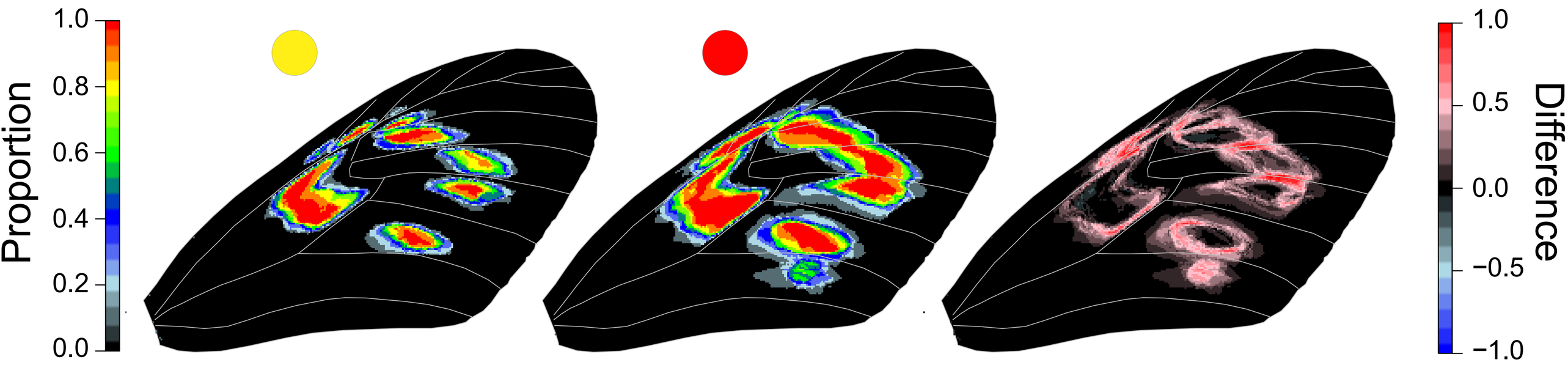
A.



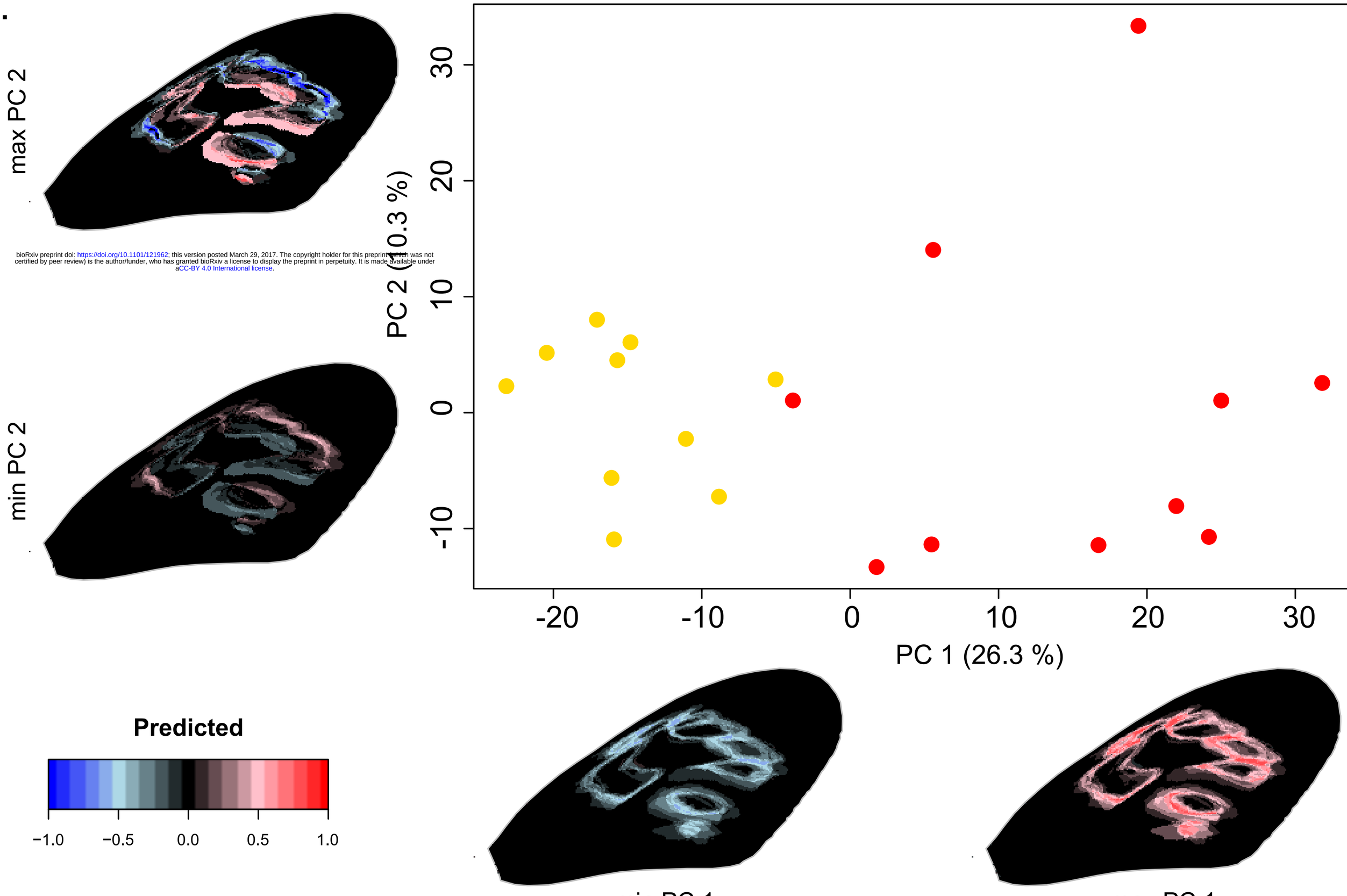
B.



C.



D.



bioRxiv preprint doi: https://doi.org/10.1101/121962; this version posted March 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/121962; this version posted March 29, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

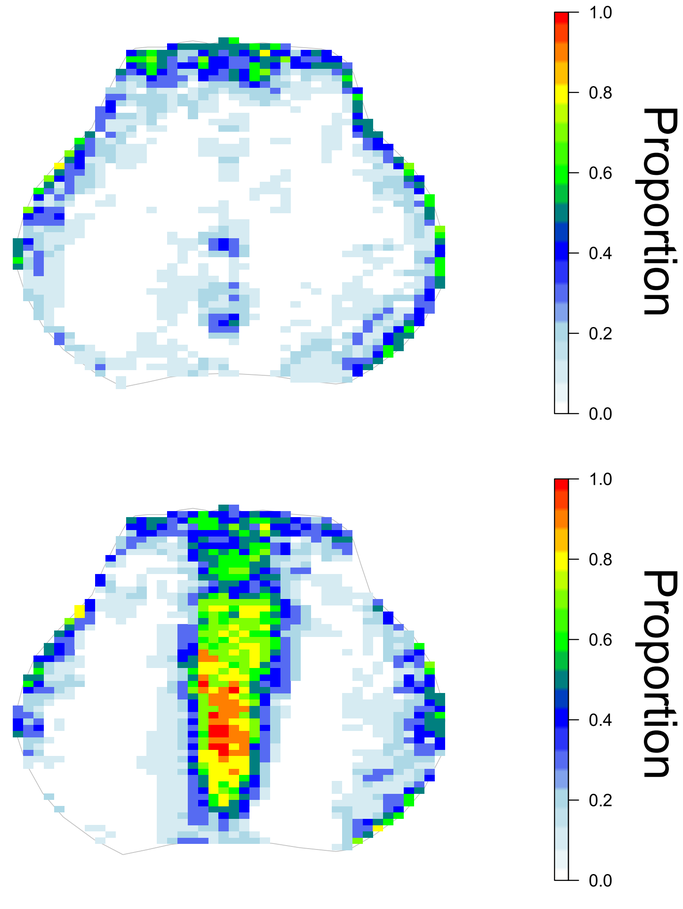
A.

Original image

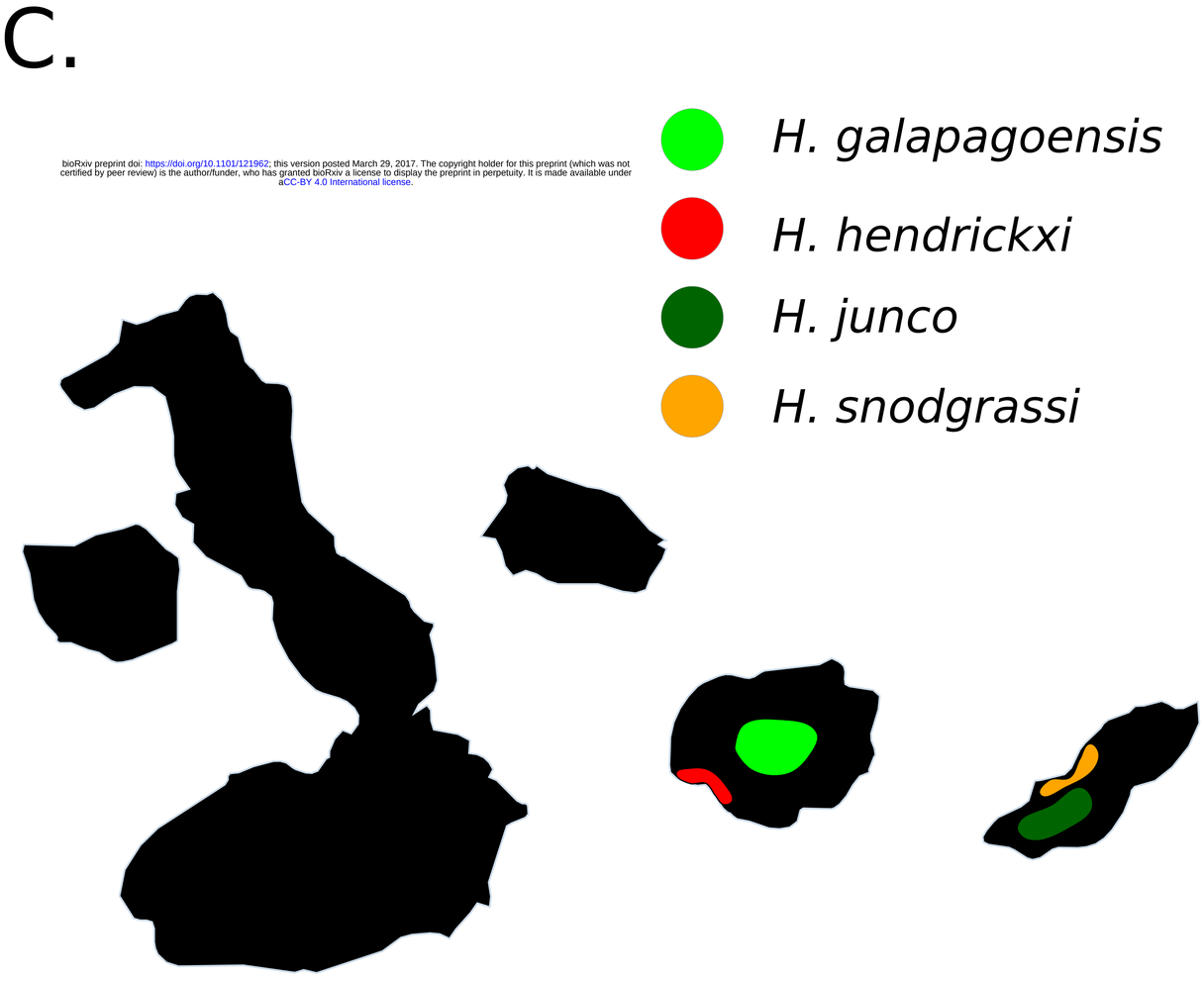
k-means clustered

Alignment (bg)

B.



C.



D.