Image analysis of the ventral colour pattern discriminates between Spectacled salamanders, *Salamandrina perspicillata* and *S. terdigitata* (Amphibia, Salamandridae)

Claudio Angelini¹,²,*, Corrado Costa³, Stefano Raimondi¹, Paolo Menesatti³, Carlo Utzeri¹

Abstract. In the present study, we applied statistical methods to quantitative image analysis of the persistent and individual ventral colour pattern of *Salamandrina* salamanders, in order to discriminate between individuals of the two species belonging to this genus. Pictures of 238 individuals from three populations of *S. perspicillata* and pictures of 95 *S. terdigitata* from two populations were analysed. Partial least squares discriminant analysis (PLSDA) classified 98.78% of individuals into the correct species. PLSDA reaches lower percentages of correct classification when applied to discriminate individuals from different populations of the same species (74.14% for *S. perspicillata*, 78.26% for *S. terdigitata*). An ANOVA analysis of colour abundances in different body sectors reveals significant differences between species. The results show that colour pattern has a specific basis, the most discriminant areas being the head and the pectoral girdle. We discuss these results in the light of the proposed evolutionary scenarios of the species, and suggest that ventral colour patterns were driven by founder effect.

Keywords: animal coloration, Partial Least Square Discriminant Analysis, species recognition, topographical image analysis.

Introduction

Animal coloration has been a topic of interest and research in biology for over a century. Colour pattern is one of the most frequently used characters for traditional animal identification at different taxonomic levels, from intraspecific morphs upwards. Usually, the use of colour pattern implies a loss of information about the image of the animal as it is. This is due to the description of coloration patterns on the basis of dichotomic or discrete variables (e.g., Ortolani, 1999; Schaefer, Vences and Veith, 2002), or to analyses focused on separate specific regions of interest of the body (e.g., Macedonnia, Brandt and Clark, 2002; Vásquez and Pfenning, 2007), or to analyses that do not take into account the relative position of the pattern elements (e.g., Endler, 1978; Rudh, Rogell and Höglund, 2007; Wollenberg et al., 2008). Nevertheless, it is difficult to apply such methods when dealing with animals that have complex colour patterns, as the majority of amphibians have. It is known that amphibian colour patterns play an important role in intra-specific communication (Summers et al., 1999) as well in protection against predators (Summers and Clough, 2001). Here, we apply a quantitative image analysis (Glasbey and Horgan, 1995) that permits retention of the information on both colours and their position in the image analysed, to study the ventral colour pattern of the urodele *Salamandrina*.

The genus *Salamandrina* (Fitzinger, 1826) is endemic to the Italian peninsula. Formerly considered monotypic, it has been recently split into two species on the basis of genetic analyses (Mattoccia, Romano and Sbordoni, 2005; Nascetti, Zangari and Canestrelli, 2005): *S. perspicillata* (Savi, 1821) found in northern and central Italy, and *S. terdigitata* (Lacépède, 1788) in the south. So far, no morphological differences have been described to distinguish the two
species, nor using colour pattern. Both species of *Salamandrina* are entirely black on the dorsal side, with a yellowish V-shaped patch on the head. The ventral coloration consists of irregular black, red and white patches: the head is usually black with one or more white patches along the mandibular edge, the trunk is mostly black with one or more white patches, and the areas around the cloaca, the tail and the limbs are red. The ventral color pattern is unique to each individual and persistent throughout its lifetime (Vanni, Nistri and Zagaglioni, 1997). By using a quantitative image analysis, in a former study (Costa et al., 2009) we found that in *S. perspicillata* the colour patterns are population-specific. Since it is not possible to distinguish the two species by coloration, and the qualitative image analysis is not applicable to such complex colour patterns as in *Salamandrina*, in the present study we applied two methods based on quantitative image analysis to evaluate if the individual ventral colour pattern can be also used for species identification. The first method we applied was modified after Costa et al. (2009), which was successfully used to discriminate among five populations of *S. perspicillata*. This method consists of a topographical colour pattern analysis. The second method we applied is based on the ANOVA analysis of the colour abundance inside body sectors, which we used as a descriptive (human sight) complement to the first analysis. Our main purposes were to evaluate if the colour pattern, properly analysed, can represent a character to distinguish the two otherwise identical species of *Salamandrina*, and to refine an analysis method that is applicable to all kinds of colour patterns, even the more complex.

**Material and methods**

We analysed 238 ventral pictures from females of three *Salamandrina perspicillata* populations (San Rocco, SRO; Acqua della Chiesa, AdC; Ciccopano, CI), and 95 pictures of females from two *S. terdigitata* populations (torrente Cerasuolo, CE; and torrente Rosa, RO) (fig. 1). We used also 24 pictures of females from another population of *S. perspicillata* (San Francesco, SF) to test the multivariate model (see below) on an independent dataset. *Salamandrina* does not exhibit any known distinct secondary sexual characters (including colour), but it is possible to sex the females when observed in the water body used for depositing the eggs, as only females enter the water (Lanza, 1983; Zuffi, 1999). Consequently, in order to avoid any bias to analysis due to unknown colour differences between the sexes, we dealt only with females that were sexed for sure.

**Pre-processing**

We used the pre-processing procedure reported and described in details by Costa et al. (2009), that will be briefly summarized below.

Analyses were based on colour digital photographs (1200 × 1600 pixels; 300 pixels per inch) of the ventral side of salamanders. Animals were placed horizontally, as much as possible following the main body axis, and the camera was parallel to the animals. Salamanders were photographed in the field and then released. Limbs and tail were removed from the photographs.

In *Salamandrina* the colour patches show well defined borders and the three colours, black, red and white, are well distinguishable. The colours of pictures were transformed into pure black, red and white. To segment the red, green, and blue (RGB) images into three-colour images, the K Nearest Neighbours (KNN) supervised multivariate clustering method was applied (Belur, 1991). Each pixel was assigned to the colour most common among its KNN (k = 5). The classification procedure and the reconstruction of the three-colour images were performed in Matlab (Shakhnarovich, Darrell and Indyk, 2006).

The shape and colour pattern of each individual were morphologically adjusted to a standard view by means of the consensus configuration of the whole sample (Costa et al., 2009). The final RGB images were 886 × 1544 pixels at 72 p.p.i., for a total amount of 7539 pixels examined.

**Topographical analysis**

In order to check if individuals could be correctly assigned to their own species on the basis of colour pattern (i.e., position and colour of each pixel), a Partial Least Square Discriminant Analysis (PLSDA) was applied. PLSDA consists of a classical partial least squares (PLS) regression analysis where the response variable is categorical (Y-block; replaced by a set of dummy variables describing the categories), thus expressing the class membership of the statistical units (Sjöström, Wold and Söderström, 1986; Sabatier, Vivein and Amenta, 2003; Aguzzi et al., 2009). The model includes a calibration phase and a cross-validation phase for which residual errors (root mean square error) are calculated (RMSECV). The prediction ability of PLSDA also depends on the number (v) of latent variables (LV) used in the model. The optimal v value was determined by predicting the results for independent sets of samples (test set) for different values of v, and determining the value of v for which the highest percentage of correct classification was found in the test set. PLSDA calculates a “prediction probability” and a classification threshold for each group modelled. The
PLSDA analysis provides the percentage of correct classification and the loadings of each species on each LV. This analysis also expresses the statistical parameters indicating the modelling efficiency indicated by sensitivity and specificity parameters. The sensitivity is the percentage of the individuals of a category accepted by the class model. The specificity is the percentage of the individuals of the categories different from the modelled one, rejected by the class model.

This analysis was performed using Matlab (rel. 7.1, PLSToolbox Eigenvector rel. 4.0) on the single pixel colour variables (X-block) on the superimposed configuration of each individual. The dataset was divided into two subsets: the first, containing 75% of individuals of each species, was used for the class modelling and validation; the second was used for the independent test. To optimally select the 25% test set, the Kennard and Stone (1969) algorithm was applied. This algorithm belongs to the family of space-filling algorithms and is based on Euclidean distances between data. These algorithms select objects without the a priori knowledge of a regression model. The hypothesis is that as the true model is rather complex, it requires a uniform distribution of objects in the information space (see also Costa et al., 2008, and Menesatti et al., 2008, for further details on this methodology). The PLSDA model was also applied to 24 pictures of a S. perspicillata population (SF) used as an independent dataset.

Two additional PLSDA models were built to evaluate the interpopulation differences in the colour pattern within the same species. For this intra-specific PLSDA analysis, we used the same procedure as above.

**Colour abundance**

We subdivided the standardized images of individual colour pattern among the areas shown in fig. 2. Then, we calculated the proportion of black, red and white pixels for each area. In order to test if differences exist between the species concerning the colour composition of each area, we applied an ANOVA for mixture experiments (the data met the assumptions of the ANOVA). This is necessary when analysing components that must sum up to a constant, as in our case. The analysis of mixture components is the same as a multiple regression that does not include an intercept term. This analysis also provides the size of the parameters, i.e., the load of each colour in explaining the observed pattern of variance. We applied a two-way repeated measures ANOVA in order to test if the main-effect colour of each body area...
Figure 2. Body areas for the analysis of colour abundance of the ventral pattern of Salamandrina spp. (resulting from the ANOVA for mixture experiments) was symmetrically distributed between the right and left sides, and if species differed in this aspect.

Results

Topographical analysis

Table 1 shows the characteristics and the main results of the selected PLSDA model (5 LV). The cumulative percentage of variance for the Y-block is 97.9%. The model has a high specificity (100%) and sensitivity (99.7%), and a low mean classification error (0.0014). Figure 3 shows the scores of the 82 test individuals on the first two LVs (85.89% of the cumulative variance on the Y-block), which provided a good discrimination between the two species (98.78% of correct classification). The partial overlapping of the two groups is due to the lack of the other 3 LVs that complete the model. The PLSDA model is highly efficient in terms of percentage of correct classification as only one individual of S. perspicillata out of 82 was misclassified in the independent test. Figure 4 shows the loadings of each pixel on each LV, that is the contribution to the classification of each pixel (the intensity of white is proportionate to the pixel’s contribution). LV1 explains 60.26% of the total variance of the X-block and 57.23% on the Y-block. The loadings of the pixels (fig. 3) indicate that the most important areas in discriminating between species are the region along the mandibular edge and the y-shaped area between the neck and the pectoral girdle. The same model correctly classified 21 out of 24 (87.5%) individuals from SF population into S. perspicillata.

Two additional models were built in order to discriminate females of S. perspicillata into the three populations, and females of S. terdigitata into the two populations (table 1). For S. perspicillata, the percentage of correct classification of the 58 test individuals was 74.14% (model built on 180 calibration individuals from three populations). The percentage of correct classification of the 33 test individuals belonging to two populations of S. terdigitata was 78.26% (model built on 72 calibration individuals).

Colour abundance

Table 2 reports the results of ANOVA for mixture experiments for each area. In all cases the linear model had the best fit. The colour composition of each area is different between species, except area I (see also fig. 5). The two-way ANOVAs revealed that the main colour is asymmetrically distributed between left and right sides (area I more abundant at the right; II at the left; III at the right; V at the right; whole body at the right). We also found interaction between the factors “body side” and “species” for area V (S. perspicillata showed...
Table 1. Characteristics and main results of the PLSDA models used to identify individuals belonging to each of the two species of *Salamandrina* (2 species column), to each of the three populations of *S. perspicillata* (pop. *Sp* column) and to each of the two populations of *S. terdigitata* (pop. *St* column).

<table>
<thead>
<tr>
<th></th>
<th>2 species</th>
<th>Pop. <em>Sp</em></th>
<th>Pop. <em>St</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of individuals for 75% modelling and validation</td>
<td>251</td>
<td>180</td>
<td>71</td>
</tr>
<tr>
<td>No. of individuals for 25% independent test</td>
<td>82</td>
<td>58</td>
<td>24</td>
</tr>
<tr>
<td>No. of classes</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Random probability (%)</td>
<td>50</td>
<td>33.33</td>
<td>50</td>
</tr>
<tr>
<td>No. of LV</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Cumulated variance X-block (%)</td>
<td>62.57</td>
<td>62.95</td>
<td>65.20</td>
</tr>
<tr>
<td>Cumulated variance Y-block (%)</td>
<td>97.90</td>
<td>86.70</td>
<td>91.55</td>
</tr>
<tr>
<td>Mean specificity (%)</td>
<td>100</td>
<td>98.06</td>
<td>100</td>
</tr>
<tr>
<td>Mean sensitivity (%)</td>
<td>99.70</td>
<td>98.90</td>
<td>98.55</td>
</tr>
<tr>
<td>Mean classification error (calibration)</td>
<td>0.0014</td>
<td>0.0153</td>
<td>0.0074</td>
</tr>
<tr>
<td>RMSEC</td>
<td>0.1026</td>
<td>0.2090</td>
<td>0.2055</td>
</tr>
<tr>
<td>Correctly classified model (%)</td>
<td>100</td>
<td>99.67</td>
<td>98.61</td>
</tr>
<tr>
<td>No. misclassified (model 75%)</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Correctly classified independent test (%)</td>
<td>98.78</td>
<td>74.14</td>
<td>78.26</td>
</tr>
<tr>
<td>No. misclassified (test 25%)</td>
<td>1 (Sp in St)</td>
<td>15</td>
<td>5</td>
</tr>
</tbody>
</table>

Figure 3. Partial least squares discriminant analysis: plot of individual PLSDA scores of *Salamandrina* spp. individuals on LV1 and LV2.
Figure 4. Partial least squares discriminant analysis: PLSDA loadings of the pixels for each latent variable: white intensity is related to a higher contribution to the classification; (a) shows the pixels that have a loading greater than 215 on the LV1.

Table 2. Colour abundance analysis of the ventral pattern of *Salamandrina* spp. based on ANOVA for mixture experiments. Areas are designated as in fig. 2. Two-way ANOVA was performed for the main colour (r-l = comparison between right and left sides; r-l* species = interaction between factors “r-l” and “species”) (** indicates $P < 0.05$, - - - indicates $P > 0.05$, after Bonferroni’s correction).

<table>
<thead>
<tr>
<th>Area</th>
<th>$F_{2,330}$</th>
<th>$P$</th>
<th>Main colour</th>
<th>Two-way ANOVA$_{1,331}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>3.24</td>
<td>---</td>
<td>black</td>
<td>r-l: **<em>; r-l</em> species: - -</td>
</tr>
<tr>
<td>II</td>
<td>20.57</td>
<td>***</td>
<td>black</td>
<td>r-l: **<em>; r-l</em> species: - -</td>
</tr>
<tr>
<td>III</td>
<td>21.91</td>
<td>***</td>
<td>white</td>
<td>r-l: **<em>; r-l</em> species: - -</td>
</tr>
<tr>
<td>IV</td>
<td>15.06</td>
<td>***</td>
<td>white</td>
<td>r-l: - - -; r-l* species: - -</td>
</tr>
<tr>
<td>V</td>
<td>17.09</td>
<td>***</td>
<td>red</td>
<td>r-l: **<em>; r-l</em> species: ***</td>
</tr>
<tr>
<td>Whole</td>
<td>16.74</td>
<td>***</td>
<td>black</td>
<td>r-l: **<em>; r-l</em> species: ***</td>
</tr>
</tbody>
</table>
Figure 5. Triangular plots representing colour proportion in each body area and in the entire body for *S. perspicillata* (grey circles) and *S. terdigitata* (black triangles) (*b* = black, *r* = red, *w* = white).

more red at the right side, *S. terdigitata* at the left side) and for the whole body. An ANOVA for colour abundance was also applied to the pixels that have a loading greater than 215 on the LV1 (the maximum loading value was 255) from PLSDA (fig. 4a) and, as expected, their colour abundances are significantly different between species (*F*₂,₃₃₀ = 7.14, *P* < 0.001).

**Discussion**

In the present study the differences in ventral coloration patterns between the two species of the genus *Salamandrina* were studied in order to find an efficient model of classification. The PLSDA-based model showed a high efficiency of classification in the internal test (one individual misclassified out of 82), as well as for the additional SF population (three individuals misclassified). Costa et al. (2009) found that the very high inter-individual variability of the colour pattern of *S. perspicillata* had a population basis, with almost 89% correct classification of individuals among five populations. In the present paper, the PLSDA shows a similar level of efficiency when applied at the population level, but individuals of different species are discriminated better (nearly 99%) than individuals of different populations within the same species, thus indicating that colour pattern has an even stronger specific basis.

Previously, no morphological differences were known between the two *Salamandrina* species. We found interspecific differences in the distribution of colours on the ventral area (table 2 and fig. 5); and PLSDA indicated that the areas where the species are particularly different are the anterior part of the head and the pectoral girdle, the same areas that Costa et al. (2009) found highly discriminating among populations. Utzeri, Antonelli and Angelini (2005) observed salamanders displaying the anterior part of their ventral surface, by raising on the fore limbs or standing up on their hind limbs for several minutes. On these bases, Costa et al. (2009) suggested that the coloration of the anterior region could be related to intraspecific communication, thus being subject
to micro-evolutionary forces, whilst the rest of the trunk could be involved in aposematic signalling for which just the presence of the colour patches and not their relative position may be important. Our findings that the coloration of the anterior region is also the most discriminating area between the species seem to confirm that this area plays a different role from the rest of the trunk. *Salamandrina salamanders* are very philopatric (Angelini, 2006) and even close populations keep isolated, showing little if any gene flow (Hauswaldt et al., 2008). This led Costa et al. (2009) to suggest that genetic drift is probably the main force fixing the colour pattern of the anterior ventral region of the body at the population level. Interestingly, the proposed speciation scenario that Mattoccia, Romano and Sbordoni (2005) and Nascetti, Zangari and Canestrelli (2005) have in common (two different dispersal events from Sardinia towards the Italian peninsula) should have implied two different (groups of) founder populations carrying their own characteristic coloration patterns. If so, it is very likely that genetic drift acted also in differentiating the colour pattern of the anterior region between the two species. Further, if the anterior colour pattern is actually involved in intra-specific communication, this would have implied a sort of “obliged” reinforcement, similar to the process known to lead to sympatric speciation (Servedio and Noor, 2003; Gray and McKinnon, 2006), thus fixing the extant colour pattern of each species.

We think our results are also relevant within the framework of (colour) polymorphism. The colour pattern of *Salamandrina*, similarly to other species (e.g., of the genus *Bombina*, among amphibians) shows such a high inter-individual variability that discrete morphs are not identifiable within it, that is, it is not strictly referable to as polymorphism. Nevertheless, our results, as well as those of Costa et al. (2009), indicate that even if morphs are not identifiable by sight within *Salamandrina* species, the individuals tend to share common coloration if belonging to the same population, i.e., we can recognize population-morphs. On the other hand, very likely even simple colour patterns that are easy to arrange in morphs (e.g., *Salamandra atra atra* vs. *S. a. aurorae*) might show high intra-morph variability when analysed with appropriate tools, e.g. by spectral analyses (Endler, 1990) or by our topographical analysis. In short, the improvement of analysis instruments makes colour polymorphism a topic worth reexamining. In turn, the improvement of statistical analyses, and related software, makes it feasible to extract and assess far more information from coloration than was previously possible. Furthermore, the very high percentages of correct classification we obtained indicate that statistical modelling is a promising tool also for the systematic biologist. This is preliminary confirmed by the very good performance (87.5%) of the PLSDA model when applied to the individuals of the external test population SF.

The PLSDA analysis is widely applied in a broad range of research fields (agricultural products, marine biology, microarray data, spectroscopy), and recently also in image analysis (Xing, Saeys and De Baerdemaeker, 2007; Costa et al., 2008, 2009; Menesatti et al., 2008; Aguzzi et al., 2009), in which it allows a topographical approach to colour differences. Nevertheless, it does not permit discrimination of the species based on human sight. The opposite is true for the colour abundance analysis: it is based on human sight, but it is purely descriptive, since it takes into account only the amount of different colours but not their exact position. However, the ANOVA examining colour abundance showed that species differ in this aspect, also. We think that supporting the multivariate classification in image analysis with descriptive means (see also Costa et al., 2009) is a promising approach to distinguish individuals based on human sight. Furthermore, the warping procedure we used, where geometric morphometry is applied only for shape standardization, has a great potential of applicability to other research fields dealing with image processing such as
hypespectral imaging, colorimetric and texture mapping and also for the automated individual identification.

Acknowledgements. Authors are grateful to S. Hauswaldt (Technical University of Braunschweig), whose first revision of the manuscript really improved our paper, and to D. Antonelli and B. Cari for the field help.

References


Received: June 12, 2009. Accepted: December 11, 2009.