

**Futile searches, obvious findings: a 2D digital exploration of morphological indistinctiveness in the (sub)genus *Viduoliva* Petuch & Sargent, 1986 (Olividae: Olivinae)**

Cesare Brizio

World Biodiversity Association, Museo Civico di Storia Naturale di Verona,  
Lungadige Porta Vittoria, 9, 37129 Verona, Italy. <https://orcid.org/0000-0002-7042-9684>  
[briziocesare@gmail.com](mailto:briziocesare@gmail.com)

**ABSTRACT** “The Olivinae problem” (here defined - for the namesake family - as the particular difficulty of specific attribution of a shell based on its morphology) originates from the concurrence of interspecifically low, intraspecifically high variability and allometric growth in overall smooth, ellipsoidal shells lacking spines, varices, costae or other species-specific pronounced external features. To explore the potential of 2D digital morphometrics techniques in tackling the issue, around 200 dorsal view silhouette photographs of as many specimens of 13 species of *Viduoliva* were shot. For each silhouette, an *ad hoc* program generated three shape indexes and a shape vector based on 100 equally spaced points on the shell contour. Multivariate analysis of the shape indexes via 3D scatter plots resulted in deeply intermeshing clusters where only pairs of the most dissimilar species could be separated reasonably well. The shape vectors, besides being used to generate an “average *Viduoliva* shape”, were input in another *ad hoc* program that grouped them by similarity thanks to a few techniques of agglomerative clustering based on difference metrics. The indecisive results confirmed the morphological indistinctiveness of most *Viduoliva* species. The Python software developed for this experiment is freely available for download and possible improvement or repurposing.

**KEY WORDS** Digital morphometrics, Olivinae, *Viduoliva*, multivariate analysis, shape vectors

## INTRODUCTION

Nature can't be fooled. In the creative rewording and reframing by the father of Cladistics, Willi Hennig, the successive speciation events that, in time, build the phylogenetic trees, at individual level boil down to tokogeny, the biological relationship between parent and offspring (see Hennig, 1966). Species exist inasmuch as successful interbreeding is sustained in time. Hence, intricate as it may seem, there is just one definitive and unambiguous proof that two anisogamous organisms belong in different species: their failure to interbreed successfully with production of fertile offspring – the phenomenon on which, among other things, the

sympatry criterion for species separation is based. Any other way to deal with interspecific boundaries may be defined as:

- indirect, whenever the actual impossibility of fertilization is demonstrated by a thorough description of the physical barriers that impede copulation, as customarily done, *e.g.*, in many branches of entomology (traumatic insemination should be separately considered);
- inductive, whenever the impossibility or ineffectiveness of copulation is assumed, as in the case of purely genomics methods (Lesk, 2017), that require the strictest caution: even when whole-genome sequencing is available, using estimated gene flow as a proxy for reproductive isolation can be questioned outside of the

rigorous methodological frame by Singhal *et al.* (2025) and bibliography therein. For candidate new species, expeditious genomics (such as any DNA barcoding technique) does not provide certainties, but statistical evidence of the plausibility of some predefined hypothesis regarding speciation. Phillips *et al.* (2022) demonstrated that the reliability of such evidence is particularly sensible to statistical issues, that can be mitigated by a good competence in statistical genomics (see also Collins & Cruickshank, 2012; Cheng Z *et al.*, 2023; Cong Q *et al.*, 2017; Meyer & Paulay, 2005; Nymoen *et al.*, 2024), and by integrative frameworks as advocated, *e.g.*, by Caruso *et al.* (2024).

The typical conchological studies are purely morphological and, consequently, inductive investigations, especially when they do not involve malacology proper (most conchologist don't like to be schooled by malacologists in that respect...). The choice of not considering the animal nor its lifecycle may leave the researcher at grasps with an indistinctive shell morphology. Such indistinctiveness – the effect of high intraspecific variability and low interspecific variability, exacerbated by allometric growth – is the hallmark (hereinafter, “the Olivinae problem”) of many genera and species in the subfamily Olivinae, whose systematics in the last few centuries, not by chance, has always been contentious – to say the least.

Instituted at page 157 of Petuch & Sargent (1986), *Viduoliva* (as most of the newly-created subgenera by Petuch & Sargent, 1986) would have deserved the rank of full genus (Ed Petuch, pers. comm. 15 November 2021). Formally, *Oliva (Viduoliva)* Petuch & Sargent, 1986 is currently recognized by the World Register of Marine Species as an alternative representation of *Oliva* Bruguière, 1789. It was chosen as the subject of this experiment for the availability of

a sufficient number of specimens in the author's collection (around 200, all dexiotropic), and because it epitomizes the “Olivinae problem”.

For a few decades, information technology has come to the rescue of the human experts in the field of morphology, originating special areas of research, including digital morphology and digital morphometrics, particularly well-developed in the field of medical sciences but also for investigations in archeology, botany, zoology and paleontology. Among the seminal works, those by Norman MacLeod, who developed the concepts of landmark, eigenshape and eigenanalysis in the context of biological image analysis (MacLeod 1999, 2005, 2007, 2009a, 2009b, 2015). It should be noted that we are not talking of automated image recognition nor of deep-learning image classifiers: even though some algorithms and technicalities are successfully employed in both fields, the basic challenge in numerical morphology is the digital representation of the shape at a size-agnostic level of abstraction higher than the image itself. Such level is well beyond purely linear measurements such as length and width and – by leveraging and expanding abstractions such as shape indexes, landmarks, contours – aims at conveying all the peculiar, intrinsic characters that differentiate one shape from another, and allow the researcher to treat each entity as a point, or as a cloud of points, in a high-dimensional shape space.

Human expert classification rests on several concurrent subtle morphological clues, most of which develop in the three-dimensional domain and are not well-suited for discrete numerical representation. In the two-dimensional domain, many shape analysis techniques emerged in recent years. Regarding outline shapes, Salili-James *et al.* (2022) in their comprehensive

work compare geometric morphometrics, semi-landmarks eigenshapes analysis, and three diffeomorphic (Hirsch, 1997) methods – large deformation diffeomorphic metric mapping (LDDMM), square-root velocity functions (SRVF) and geometric currents – also taking advantage of machine learning techniques.

Similar to the lap time of a sportscar depends on the mechanical grip offered by the track surface, the efficacy of computational methods for shape recognition correlates directly with the degree of difference in a set of objects. In Salili-James *et al.* (2022) the software classifiers exceeded the discriminative power of the human experts participating in the study, but each set of outlines (leaves, vases, seashells) considered in their study comprised only highly different shapes. In that respect, Olivinae are a very slippery track. If – as it can reasonably be expected – the “Olivinae problem” reverberates in the digital domain, there’s no certainty that even the most sophisticated digital morphometrics techniques can generate unequivocal groupings separated by clear-cut boundaries.

Even more so for this experiment, based on a simplistic 2D approach, outline (“silhouette”) shape analysis, limited to one combination of orientation and axial rotation per specimen. Simplicity seconds both the limited experience of the author, and the modest variance of the shapes of *Viduoliva*: techniques such as SRVF are often too complex for an exploratory project and cannot remediate the lack of actual morphological discontinuities. The methods proposed here were chosen also because the extraction of meaningful data (indexes, shape vectors, difference metrics) from the raster images could be integrally managed by the software, with no human intervention. Two approaches were explored:

- Purely computational shape indexes were based on a few of the most typical features

of Olivinae shells: the only expected outcomes of a multivariate analysis of very similar shape indexes are somehow distinct clouds/clusters of similar values, with limited overlap and – hopefully – a denser core, susceptible of Principal Components Analysis or other statistical techniques. As we will see, such expectation proved optimistic.

- A similarity analysis based on shape vectors, inspired by the work of Thanh and Quan (2025), who provided clear indications about a programming language (Python) implementation of their methods. By taking into account the entirety of the silhouette outline, it can be expected to provide a more sophisticated concept of “digital similarity”, and for sure can be used to compute the outline of an “average *Viduoliva*” (see the Results section).

However reductive, our approach will provide an interesting opportunity to explore some of the digital morphometrics methods cited above and, at the same time, it will give an objective, numerical dimension to the “Olivinae problem” that puzzles the human experts.

A final and probably superfluous warning in the era of the so-called “artificial intelligence”: being entirely predetermined by the input data and by the underlying algorithm, computer outputs are not “demonstrations”, but intrinsically unassertive, merely passive representations. The computer cannot corroborate any conclusion, but can just arrange and display the available data in the format deemed more effective for the desired purpose. No “intelligence” there!

## MATERIALS AND METHODS

One hundred and ninety-nine *Viduoliva* shells, attributed by human experts to thirteen species (Table 1) were chosen for this experiment. The allometric growth of Olivinae would require to

restrict any comparative morphological study to a consistent set of adult specimens with exactly the same number of volutions, factoring out gerontic as well as juvenile shells. That is

not our case: even though blatantly immature specimens have been excluded, photographed specimens may differ in the number of degrees of teleoconch volution.

Species number	Species name	Number of specimens
1	<i>Viduoliva elegans</i> (Lamarck, 1811)	19
2	<i>Viduoliva indomalaysica</i> Petuch & Sargent, 1986	8
3	<i>Viduoliva keeni</i> (Marrat, 1870)	20
4	<i>Viduoliva macleaya</i> (Duclos 1840)	16
5	<i>Viduoliva mindanaoensis</i> Petuch & Sargent, 1986	27
6	<i>Viduoliva neostina</i> (Duclos 1840)	8
7	<i>Viduoliva raderi</i> Petuch & Sargent, 1986	8
8	<i>Viduoliva reticulata</i> (Röding, 1798)	42
9	<i>Viduoliva rubrolabiata</i> (Fischer, 1903)	2
10	<i>Viduoliva sp.</i>	1
11	<i>Viduoliva tricolor</i> (Lamarck, 1811)	12
12	<i>Viduoliva vidua</i> (Röding, 1798)	31
13	<i>Viduoliva westralis</i> Petuch & Sargent, 1986	5
Total		199

**Table 1.** Species and number of specimens considered in this study.

The use of a video camera – as advocated in a previous publication (Brizio, 2019) - greatly simplified both the setup of the imaging equipment and the capture of each image by providing a constant preview and several user-controlled settings. To favor the programmatic generation of unique, reliable and continuous contours, the shells were photographed in dorsal aspect, with the apertural side facing a backlit background (LED tablet) as illustrated in Figures 1 and 2. In high-contrast backlit photography, parasite reflections may cause peripheral light blotches that would be computed as separate contours by the edge-detection algorithms. To avoid such disturbances from the illuminated area immediately surrounding the shell, an elevated acrylic base, resting on an open black cardboard box of suitable size, was placed on

the LED tablet, and surrounded by an external, slightly bigger open cardboard box. A small, frosted translucent PVC plate was positioned under the shell to improve light diffusion and to obliterate the small defects of the acrylic base, that would otherwise appear in the illuminated background. An 8 MP DeltaPix Invenio 8DII video camera mounted on a fixed stand and equipped with a Sony SAL/SAM 18-55 photo camera objective was used to photograph the shell from above, via the image capture function of the DeltaPix InSight software.

The dorsal silhouette images were oriented according to the apex-up iconographical convention, in which the upper part of the image coincides with the posterior part of the shell: hence, we may consider expressions as

“upper third” and “posterior third” as equivalent. Care was taken in ensuring that the coiling axis was parallel to the horizontal acrylic base, by placing a very small amount of reusable adhesive putty under the tip of the parietal plate (Figure 2). Considering that the attitude of the shell under the camera is determined by the lip edge touching the transparent support, differences in the number of volutions and in the degree of lip development result in a different amount of axial rotation, regardless of the care put in ensuring a consistent vertical orientation of the coiling axis in the images. Furthermore, lip development influences the right-to-left offset of the coiling axis (Figure 2, inset 6), that in dextrotropic shells is strongly displaced to the right (dorsal view, apex up) by a highly developed, gerontic lip. While such unavoidable inconsistencies concur with the intrinsic shape differences in determining the shell silhouette, it’s tentatively assumed that their effect is negligible.

After image capture, the silhouettes were consistently rotated via Photoshop Elements software, to ensure the verticality of the spiral coiling axis: in this experiment, the reference points chosen in dorsal view to pinpoint the axis are the shell apex and the point of deepest indentation (posteriormost / uppermost point) of the anterior notch, as observable in the silhouette. Correct alignment was helped by the use of a transparent set square superposed to the computer screen. Each oriented silhouette was then processed by an *ad hoc* Python program developed by the author, that performed the following operations:

- Blurring the image, to reduce noise.
- Thresholding (transforming in two-color) the image to improve contour detection.
- Detecting silhouette contour by suitably parametrized Canny edge detector (see the References entry Wikipedia Canny Edge Detector, 2026).

- Cropping the image at the minimum bounding box surrounding the contour.
- Resizing to 1000 pixel of vertical height (variable width) to ensure shape comparability regardless of the linear size of the shell. We remind the reader that shape is a set of size-independent features.
- Detecting the contour of the resized image by Canny edge detector.
- Individuating the centroid of the contour.
- Sampling of 100 points at equal intervals (= equal number of contour pixels) on the contour.
- Calculating, for each point, the Euclidean distance from the centroid.
- Saving a 100-points vector of Euclidean distances in a 2D matrix (one line per shell, 100 columns per line – see Figure 3) containing the data to generate a curve for each specimen, as illustrated in Figure 9.

Contour sampling did not occur at fixed angular distances, that would have undersampled feature-rich portions, such as the apical area and the shell shoulder. Instead, in the radial projection from the shape centroid, the angular distance between successive points may vary slightly, depending on the exact shape. Decisively, overall shape similarity and high density of sampling ensure that all the contours based on the sampled points are consistent and comparable, and that each point, starting from the shell apex, was sampled at roughly the same angular distance in every shell. This assumption was confirmed by the successful reconstruction of a calculated shell outline from a 100-points vector, based on the inconstant angular distances derived from the outline of a typical *Viduoliva vidua* specimen. Subsequent program steps included the following:

- Saving the width of the 1000-pixel image in a variable.
- Filling in white the whole contour.
- Applying thresholding to binarize the image (black/white).

- Creating two separate images from the upper (posterior) third and the lower (anterior) two-thirds of the 1000-pixel images.
- Counting the number of white pixels in each image.
- Calculating the shape indexes described below.
- Saving as a csv (comma separated values) file the 2D matrix containing the shape indexes.

**First investigation - multivariate scatter analysis of three shape indexes.** For a clearer explanation of the calculation of the shape indexes, the following definitions apply: all the variables cited contain a number of pixels.

- VL – length of the vertical side of the smallest bounding box.
- HL – length of the horizontal side of the smallest bounding box.
- BA – Box Area, the area of the smallest bounding box (VL \* HL).
- CA – Contour Area, the area inside the contour.

The following shape indexes are calculated:

- **AR – ASPECT RATIO:** VL/HL based on the entire 1000-pixel image, it expresses the degree of slenderness (elongation along the coiling axis) of the shell.
- **PTI – POSTERIOR TAPERING INDEX,** calculated as  $1-(CA/BA)$  with reference to the posterior (“upper”) third of the silhouette. A high, pointed spire would result in a higher value than a blunt, flat spire.
- **ATI – ANTERIOR TAPERING INDEX,** calculated as  $1-(CA/BA)$  with reference to the anterior (“lower”) two-thirds of the silhouette. A fusiform, biconical shell would score a higher ATI than a cylindrical shell.

3D scatter plots based on the three indexes were generated in the RStudio environment thanks to the scatterplot3d library (R language).

Partial plots required the use of the R-language dplyr library for data filtering. The Shape Coefficient Range image was created with Microsoft Excel from a data table.

**Second investigation - Grouping by similarity of Euclidean distance vectors.** The term “cluster” has two different undertones: in the multivariate analyses, it refers to the proximity grouping of data points, as usual for scatter diagrams. Concerning the agglomerative clustering of shape vectors, it refers to the algorithmically generated groups of similar contours.

The second technique evaluates similarity at the scale of the entire Euclidean distance vectors: that way, the entire contour is considered at once. While the grouping by similarity of equidimensional 1D vectors is a classic computational problem, with several source code libraries available for the Python language, none is especially conceived for the treatment of contour vectors, so an exploratory approach was adopted. Depending on the algorithm, the target number of clusters can be user-defined, or emerge from a hierarchical clustering function, in which case a classificatory dendrogram can also be generated. After some fruitless attempts, agglomerative clustering was adopted (see the References entry SciKit Learn Agglomerative Clustering, 2026). In the most frequent implementations, the main phases of the process are three:

- Calculation of pairwise distance among curves, based on the chosen metric;
- Linkage of the pairwise distances, based on the chosen method;
- Clustering (“Labeling”) of the curves, based on the chosen criterion.

When assessing the potential of the software in support of species recognition, both very few and too many clusters are informationally

uninteresting, and it may be wiser to ask (or to set parameters) for roughly as many clusters as species. After trying several different software setups, the figures and the tables were generated as follows: Pairwise Distance algorithm, Euclidean and correlation metrics, average and complete methods, maxclust criterion with 13 cluster (SciPy v1.16.2 Manual, 2026). Furthermore, an approach based on the Fréchet distance (Fréchet, 1906; Eiter & Mannila, 1994) (also known as “shortest leash length”) algorithm was tried with precomputed metric and complete method to generate 9 clusters (Wikipedia Fréchet Distance, 2026; Pypi Project Discrete Fréchet Distance, 2026). One more library available on the Python Package Index, Similarity Measures, was also tested (Pypi Project Similarity Measures, 2026). It allowed to calculate several difference metrics within the set of curves, on which similarity grouping could be based: considering the questionable results of the other approaches, it was not delved into. The collective images of

the contour clusters, as well as the composite photos, were created with Photoshop Elements 4.0.

## RESULTS

**Shape indexes.** The well-known morphological overlap among the *Viduoliva* species is graphically revealed by the range of values of the shape indexes in the examined species (Figure 4).

The 3D scatter plots in Figures 5 – 8 show one dot per specimen, with a species-specific color (see the legend in each image).

Figures 5, 6 and 7 respectively show examples of good, partial or bad separation among clusters of pairs, or triplets, of species.

Figure 8 shows the whole set of specimens, showing two differently oriented views on the full extent of intermeshing clusters.

Species	No. of clusters (Metric = correlation, Method = average)	No. of clusters (Metric = Euclidean, Method = complete)	No. of clusters (Metric = correlation, Method = complete)	No. of clusters (Average of three setups)	No. of specimens	No. of specimens per average cluster	Percent specimens per average cluster	Cluster Spread Ranking	Cluster Spread Points	Reliability factor (% entries vs. <i>V. reticulata</i> )	Cluster Spread Points (corrected)	Cluster Spread Ranking (corrected)
<i>elegans</i>	5	6	6	5.667	19	3.353	17.65%	7	50	45%	22.62	4
<i>indomalaysica</i>	3	5	4	4.000	8	2.000	25.00%	4	80	19%	15.24	7
<i>keeni</i>	5	5	6	5.333	20	3.750	18.75%	6	60	48%	28.57	1
<i>macleaya</i>	5	6	9	6.667	16	2.400	15.00%	9	30	38%	11.43	10
<i>mindanaoensis</i>	5	6	8	6.333	27	4.263	15.79%	8	40	64%	25.71	3
<i>neostina</i>	4	3	4	3.667	8	2.182	27.27%	3	90	19%	17.14	6
<i>raderi</i>	5	4	6	5.000	8	1.600	20.00%	5	70	19%	13.33	8
<i>reticulata</i>	7	7	9	7.667	42	5.478	13.04%	10	20	100%	20	5
<i>rubrolabiata</i>	2	2	2	2.000	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
<i>sp.</i>	1	1	1	1.000	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
<i>tricolor</i>	2	3	5	3.333	12	3.600	30.00%	2	100	29%	28.57	1
<i>vidua</i>	8	10	9	9.000	31	3.444	11.11%	11	10	74%	7.38	11
<i>westralis</i>	2	4	2	2.667	5	1.875	37.50%	1	110	12%	13.10	9

**Table 2.** Results of agglomerative clustering of 199 outline curves. The right-hand columns rank the species by morphometric consistency, excluding those with less than five specimens. See the text for explanations.

**Shape vectors.** The Euclidean distance 1D vectors can be used to generate curves like those in Figure 9. From the whole set of curves, additional vectors with the absolute minimum and absolute maximum for each sampling position were calculated, and the same values were averaged for all the set, originating the “average *Viduoliva*” curve. The three calculated vectors were used to generate as many contour shapes (Figure 10). With some caution, the one based on the average values can be considered a realistic representation of the “average *Viduoliva* contour” for what concerns the set of specimens included in this experiment. While the green shape was obtained by averaging the data point of all the vectors, the points of minimum and maximum outlines have been contributed by one vector or by a few limit cases. Red line and blue line serve only visual reference purposes as hypothetical extremes of variability, and the possibility that a *Viduoliva* with such contours can exist is undemonstrated.

Agglomerative clustering of the shape vectors confirmed the scenario outlined by the multivariate analysis of the shape indexes. At the scale of the whole contour, indistinctiveness is boosted rather than mitigated. Table 2 summarizes the results of agglomerative clustering based on pairwise distance and three different setups, the third and most evenly distributed of which was used to generate Figure 11. The results of the single viable clustering setup based on the Fréchet distance are presented in figure 12.

## DISCUSSION

**Shape indexes.** In the 3D scatter plots, no trace of cluster ellipsoidicity, a precondition for a successful Principal Components Analysis, was observed at specific level. Considering the species in pairs, non-coincident or partly

coincident triplets of indexes can be observed in Figure 4. Consequently, in particular cases, good separation (Figure 5) and partial separation (Figure 6) between species’ clusters can be observed whenever one of the two terms of comparison lays at an extreme of the 3D distribution.

Otherwise, random pairs of species can overlap extensively. Interestingly, there are cases in which the shape indexes clusters of two species intermesh deeply, while in actuality there is good morphological separation based on other characters outside shape analysis: Figure 7 shows the data points of *V. keenii* and *V. vidua*, that are easily distinguished by size (average *V. keenii* are roughly half the size of *V. vidua*) and by ornamentation (contrary to *V. vidua*, *V. keenii* ornamentation has a limited variability).

It's scientifically unacceptable to draw any conclusion based on only one data point in a multivariate analysis: anyway, more on the anecdotal than on the scientific side, it may be interesting to consider the destiny of an unidentified specimen from Semporna, Mabul Island, Malaysian Borneo (*Viduoliva sp.*, pure cyan marker in Figure 8) that human experts could not attribute to any extant *Viduoliva* species. Confirming the lack of clear affinities, in the 3D scatter plot *Viduoliva sp.* falls at the periphery of the general distribution, immediately outside the envelope of the clusters of *V. vidua*, *V. reticulata*, *V. elegans*, *V. macleaya*, corroborating the impression that it may belong to an undescribed species.

Unavoidably, when considering the entire dataset (Figure 8), the “Olivinae problem” emerges in its full extension: although the data points of a few species originate somehow distinct clusters, all the species have data points penetrating the clusters of one or – more frequently – of more than one different species.

**Shape vectors.** In the following section and in Table 2, species with less than five specimens are considered as statistically insignificant. The case of *V. rubrolabiata* is emblematic: one of the just two available specimens sports an unusually high and pointed spire, and in all the program runs the two shells ended in separate clusters, wrongly suggesting an extremely high variability of this rather consistent species.

Figures 11 (based on the correlation-complete pairwise distances) and 12 (based on the precomputed-complete Fréchet distances) show that the agglomerative clustering could not improve the discriminative power of the multivariate analysis. On the contrary, shape vectors look less informative than the triplets of shape indexes. Aggregative clustering reveals the implicit contradiction in the requisite: “Separate all the entries in 13 bins by grouping them by similarity.” When the level of dissimilarity and the nature of the subtle morphometric differences do not provide enough grip to the tires of the clustering algorithms, the thin red line between coalescence and separation jumps wildly. Figures 11 and 12 show small clusters of up to 11 species, revealing that, as far as the shell contour is considered and mathematically speaking, interspecific similarity completely trounces intraspecific consistency.

As a side note, the Fréchet distance algorithm proved capable to agglomerate some horizontally displaced, similar curves that that the Pairwise Distance algorithm assigned to separate clusters.

Despite the poor discriminative power of agglomerative clustering, in a last-ditch attempt to extract some meaningful information from the first columns in Table 2 a question was asked: can the agglomerative clustering provide some sort of quantitative evaluation of the

intraspecific morphometrical consistency? To answer that question, we may observe that, if compared with those of the most variable ones, the contours of species with lower intraspecific variability...

- ...will be spread in fewer clusters – three columns in Table 2 show the number of clusters in which each species is distributed. Since this datum varies depending on the setup parameters, an “Average Number of Clusters” was calculated for the three runs of pairwise distance clustering;
- ...will have a higher number of specimens in each cluster they intercept - considering the number of entries for each species, the column “Number of Specimens per Average Cluster” was calculated.

Regardless of its intrinsic variability, the probability that the shape vectors of a species fall into more than one cluster increases with the number of specimens considered. In our set *V. reticulata* is present with 42 entries. If as many entries would have been available for every species, how would they have performed? Two different approaches, optimistic and pessimistic, were adopted to answer that question, with the results appearing in the right-hand side of Table 2.

- We may optimistically assume that – regardless of the number of entries – the specimens considered in this experiment encompass the entire possible intraspecific range of variation of each species. In the optimistic scenario, the average number of clusters per species will not vary if new specimens are added: hence, despite heavily penalizing the species with more shells that encompass a wider range of clusters, the “Cluster Spread Ranking” in Table 2 can be considered as a reliable distinctiveness classification, in which the first four places are respectively taken by *V. westralis*, *V. tricolor*, *V. neostina* and *V. indomalaysica*.
- If on the contrary we assume that each additional entry may extend the specific

morphometrical variability, we need to quantify the added variance, remembering that the experiment is based on exactly 13 clusters, and consequently getting rid of the following unfounded assumptions:

- Every additional specimen shall not fall in any existing cluster;
- The existing variability (*e.g.*, the number of specimens for average cluster) will be proportionally extended: if 10 specimens of species A span five clusters, 10 additional specimens of species A will span another new five clusters, for a total of ten.

Any attempt at proportioning the increase in diversity by calculated values, results in a new ranking where *V. reticulata*, the only species whose data are unaltered, gets the first place, while all the other species spread on an unrealistically higher number of clusters. The only way out of the conundrum may be the introduction of a “reliability factor”, directly proportional to the number of specimens considered for each species, and consequently expressed in forty-seconds with values ranking from 1 (=42/42) for *V. reticulata* to 0.12 (=5/42) for *V. westralis*. By translating the Cluster Spread Ranking in points, and multiplying such points by the reliability factor, a new Cluster Spread Ranking (corrected) emerges, in which *V. tricolor* and *V. keeni* share the first place, followed by *V. mindanaoensis*, *V. elegans* and *V. reticulata*.

Even though a human observer may subjectively agree on the fact that the best-scoring species have a quite distinctive outline, the results of such a naïve attempt are more suggestive than reliable. As a last comment on Figures 11 and 12, only two species own a monospecific and a bispecific cluster – although very small: *V. elegans* and *V. vidua*.

**Generalities.** The results of exploratory studies such as this experiment may be useless, but are

not worthless. Grasping their possible value is uneasy, even more so for a reader versed in olivology, who should bear in mind that the programs were fed just with black on white silhouettes, all of the same size. Results are size-agnostic, color-agnostic, pattern-agnostic and based only on the dorsal aspect of the shell. Far from considering all the morphological features of *Viduoliva* shells, and from using all the digital morphometrics techniques, the experiment was narrow both on the front of the data and of the methods.

Such limitations and the notorious “Olivinae problem” suggested to set the bar low from the outset. As expected, this experiment was incapable of separating correctly the entirety of the specimens, but failed gracefully, proving effective whenever the subset elements contained the minimum information needed for separation. Otherwise, the data are ice under the wheels of the digital morphometrics sportscar. In the 3D scatter plots, not only some pairs or triplets of species can be resolved but, as graphically illustrated in Figure 7, information about size and color would suffice to unentangle some of the intermeshing clusters. In other words, it’s currently undemonstrated that a human expert, given no other hint than the 199 black and white silhouettes printed on cards, without any reference to the actual shell size, could separate them in exactly thirteen stacks, and put the correct label on each stack. More probably, all things being the same, the expert couldn’t improve the results of this experiment. Subjectively, even in the case of the better separated clusters of two species (Figure 5), the author would have found arduous to assign a silhouette to one species or another. In the case of partially overlapping clusters (Figure 6), even separating the silhouettes relative to the most distant points of the two clusters would be challenging, while separating those in the intermeshing portion

would be more a matter of luck than of actual recognition.

## CONCLUSIONS

A few relevant details of the shape of 199 *Viduoliva* shells were digitally described and analyzed by simple, digital morphometrics methods. Results confirm the pronounced morphological overlap of the species in (sub)genus *Viduoliva*. Considering that *Viduoliva* is morphologically typical of the subfamily, and that indistinctiveness is relevant feature in other (sub)genera of Olivinae, we can confirm on a numerical basis that the “Olivinae problem” is not a psychosis.

Looking at the big picture, the lesson learned once again is: the degree of separability is an intrinsic quality of any dataset. Good software engineering and good methods are indispensable, but software does not create separability: software reveals separability. Hence, whenever data are poorly separable, software devices such as algorithms perform poorly, confirming the unassertive, purely presentational nature of computer assisted classifiers. For any software capability put into play, there will always be an indistinctiveness threshold above which the software shall not discriminate in which grouping an entity may belong.

Recapitulating, the key to effectiveness is more a matter of data than a matter of software or brainpower. The steps to take are clear:

- First, we must describe unambiguously, in human language, all the subtle morphological details that concur to specific attribution - in this area, humans perform poorly, as confirmed by the taxonomic feuds, some of which secular;
- Second, we must translate all those unambiguous descriptions in as many sets of digital quantitative or qualitative data,

the columns of our dataset or the list of features or landmarks for shape description;

- Third, we must populate the dataset with the data derived from the actual specimens, the lines of our dataset, or the points of our digital shapes;
- Fourth, we must adopt the most appropriate digital morphometrics techniques to process such information.

Provided that – in contrast to our experiment – all the four steps are taken, a software classifier can be very effective, and probably outpace, if not outsmart, any human expert.

This experiment – unprecedented in Olivinae studies – shows that, even taking advantage of a degree of precision markedly superior to the human power of quantitative evaluation, distinctive characters in *Viduoliva* are absent or so little developed that they are barely perceptible, at least for what concerns the size-agnostic silhouettes of the shell dorsal outline.

There is no way to overcome this imperceptibility by rearranging data, or by choosing different algorithm parameters. We can rely on digital morphometrics, an honest-by-design discipline that can just reveal separability, cannot create it, nor can remediate the lack of actual morphological discontinuities: slave to the intrinsic discriminative power of the data, when properly practiced it delivers results that aren't artificially boosted nor affected by any bias.

## ACKNOWLEDGEMENTS

The author thanks Vu Quyet Thanh of the Vietnam-Russia Tropical Science and Technology Research Center, Ha Noi, Vietnam, for the exhaustive methodological section in Than & Quan (2025) and for his availability for discussion.

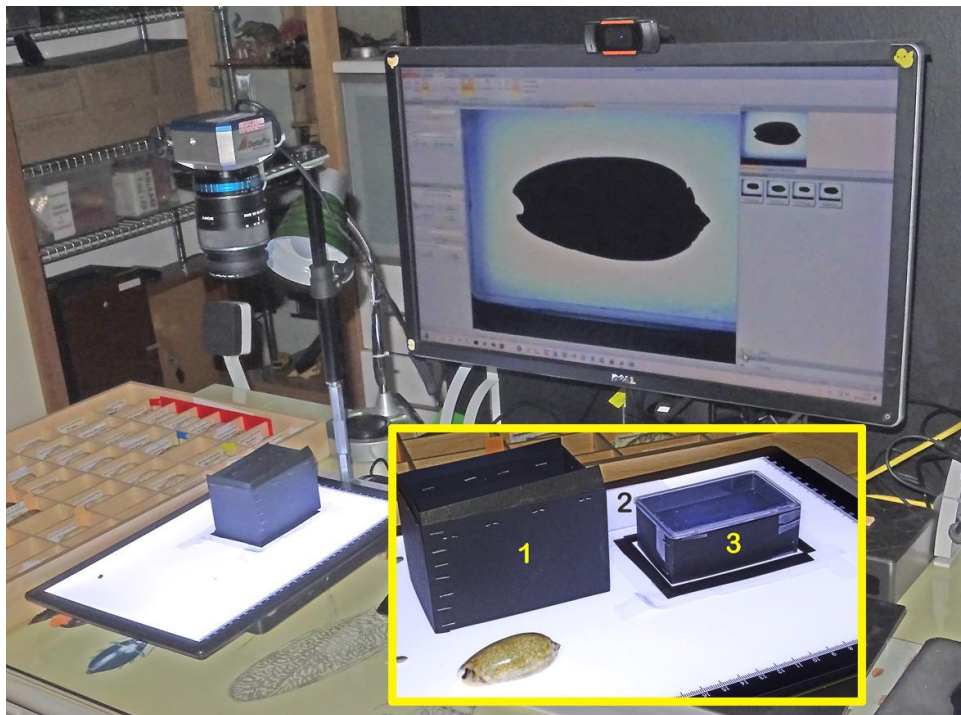
## ADDITIONAL MATERIAL

The Python programs developed for this experiment can be freely downloaded in the “Seashells” and in the “Programming” sections of the author’s website, <https://www.cesarebrizio.it>. Readers are encouraged to improve and leverage for their own projects the source code, that is released under Copyright Commons CC BY-SA 4.0 license.

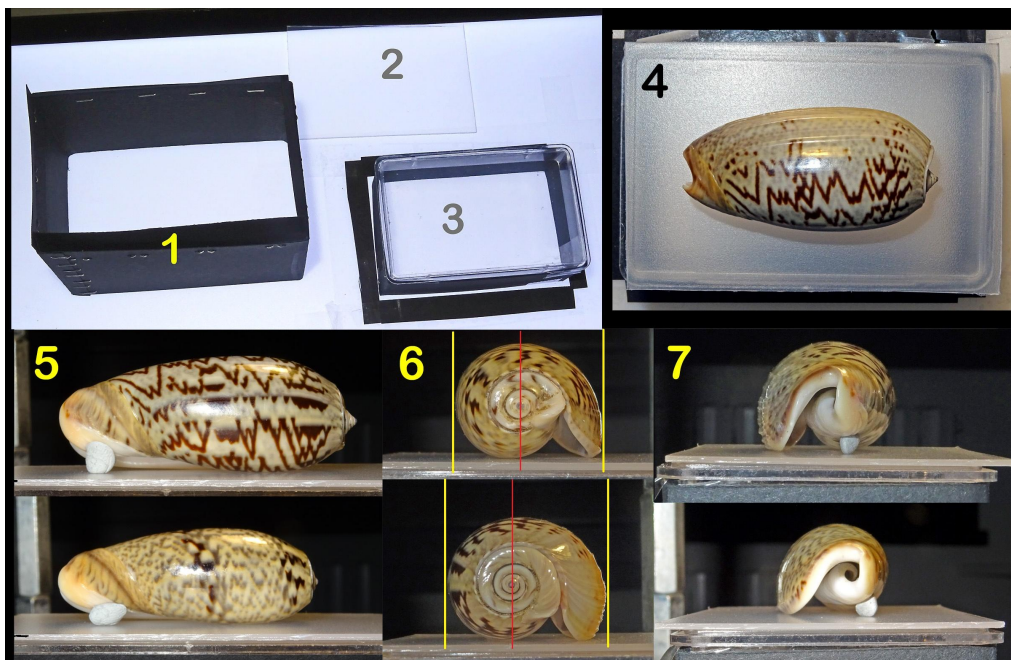
## LITERATURE CITED

- Brizio C. 2019.** Quick, massive, manual digital photography of small (~ 1/10 cm) specimen collection by a professional digital video camera and a photo camera zoom lens. *The Festivus* 51(2):150-163.
- Caruso, V., E. Hartop, C. Chimeno, S. Noori, A. Srivathsan, M. Haas, L. Lee, R. Meier & D. Whitmore 2024.** An integrative framework for dark taxa biodiversity assessment at scale: A case study using *Megaselia* (Diptera, Phoridae). *Insect Conserv Divers.* 17(6):968-987. <https://doi.org/10.1111/icad.12762>
- Cheng, Z., Q. Li, J. Deng, Q. Liu & X. Huang 2023.** The devil is in the details: Problems in DNA barcoding practices indicated by systematic evaluation of insect barcodes. *Front. Ecol. Evol.* 11:1149839:1-11. <https://doi.org/10.3389/fevo.2023.1149839>
- Collins R.A. & R.H. Cruickshank 2012.** The seven deadly sins of DNA barcoding. *Molecular Ecology Resources* 13(6): 969-975. <https://doi.org/10.1111/1755-0998.12046>
- Cong, Q., J. Shen, D. Borek, R.K. Robbins, P.A. Opler, Z. Otwinowski & N.V. Grishin 2017.** When COI barcodes deceive: complete genomes reveal introgression in hairstreaks. *Proc. R. Soc. B* 284:20161735. <http://dx.doi.org/10.1098/rspb.2016.1735>
- Eiter, T. & H. Mannila 1994.** Computing discrete Fréchet distance. Tech. Report CD-TR 94/64, Information Systems Department, Technical University of Vienna.1-7.
- Fréchet M. 1906.** Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo* (1884-1940), 22(1):1-72.
- Hennig W. 1966.** Phylogenetic systematics. University of Illinois Press.
- Hirsch M. 1997.** Differential Topology, Berlin, New York: Springer-Verlag, ISBN 978-0-387-90148-0
- Lesk A.M. 2017.** Introduction to Genomics (Third Edition). Oxford, Oxford University Press.
- MacLeod N. 1999.** Generalizing and extending the eigenshape method of shape visualization and analysis. *Paleobiology* 25:107-138.
- MacLeod N. 2005.** Principal components analysis (eigenanalysis & regression 5). *Palaeontological Association Newsletter* 59:42-54.
- MacLeod N. 2007.** Automated taxon identification in systematics: theory, approaches, and applications. CRC Press, Taylor & Francis Group, London.
- MacLeod N. 2009a.** Who is Procrustes and what has he done with my data? *Palaeontological Association Newsletter* 70:21-36.
- MacLeod N. 2009b.** Form & shape models. *Palaeontological Association Newsletter* 72:14-27.
- MacLeod N. 2015.** The direct analysis of digital images (eigenimage) with a comment on the use of discriminant analysis in morphometrics. *In: Lestrel, P.E. (Ed.), Proceedings of the third*

- international symposium on biological shape analysis. World Scientific, Singapore, 156-182.
- Meyer, C.P. & G. Paulay 2005.** DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biol* 3(12):e422.
- Nymoen, A.R., J.A. Kongsrud, E. Willassen & T. Bakken 2024.** When standard DNA barcodes do not work for species identification: intermixed mitochondrial haplotypes in the *Jaera albifrons* complex (Crustacea: Isopoda) - *Marine Biodiversity* 54(43.):1-17. <https://doi.org/10.1007/s12526-024-01435-7>
- Petuch, E.J. & D.M. Sargent 1986.** Atlas of the Living Olive Shells of the World, The Coastal Education & Research Foundation (CERF), Charlottesville (VA).
- Phillips, J.D., D.J. Gillis & R.H. Hanner 2022.** Lack of Statistical Rigor in DNA Barcoding Likely Invalidates the Presence of a True Species' Barcode Gap. *Front. Ecol. Evol.* 10:859099. <https://doi.org/10.3389/fevo.2022.859099>
- Pypi Project 2026.** Discrete Fréchet Distance. <https://pypi.org/project/Fréchetdist/> (recovered on 5 January 2026).
- Pypi Project 2026.** Similarity Measures. <https://pypi.org/project/similaritymeasure/> (recovered on 5 January 2026).
- Salili-James, A., A. Mackay, E. Rodriguez-Alvarez, D. Rodriguez-Perez, T. Mannack, T.A. Rawlings, A.R. Palmer, J. Todd, T.E. Riutta, C. Macinnis-Ng, Z. Han, M. Davies, Z. Thorpe; S. Marsland, & A.M. Leroi 2022.** Classifying organisms and artefacts by their outline shapes. *J. R. Soc. Interface* 19:20220493. <https://doi.org/10.1098/rsif.2022.0493>
- SciKit Learn 2026.** Agglomerative Clustering <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html> (recovered on 5 January 2026).
- SciPy 2026.** Manual version 1.16.2. <https://docs.scipy.org/doc/scipy/reference/index.html> (recovered on 5 January 2026).
- Singhal, S., A.D. Leaché, M.K. Fujita, C.D. Cadena & F. Zapata 2025.** A Genomic Perspective on species delimitation. *Annu. Rev. Ecol. Evol. Syst.* 2025. 56:467-89. <https://doi.org/10.1146/annurev-ecolsys-102723-055311>
- Thanh, V.Q. & N.V. Quan 2025.** A Novel Method Utilizing Otolith Outline Analysis for Identifying Fish Species. *Turkish Journal of Fisheries and Aquatic Sciences*, 25(12). <https://doi.org/10.4194/TRJFAS26348>
- Wikipedia 2026.** Canny Edge detector. [https://en.wikipedia.org/wiki/Canny\\_edge\\_detector](https://en.wikipedia.org/wiki/Canny_edge_detector) (recovered on 5 January 2026).
- Wikipedia 2026.** Fréchet Distance. [https://en.wikipedia.org/wiki/Fr%C3%A9chet\\_distance](https://en.wikipedia.org/wiki/Fr%C3%A9chet_distance) (recovered on 5 January 2026).



**Figure 1.** Photographic setup for silhouette photography via video camera. Inset legend: 1 - external black cardboard box, 2 - diffuser, 3 - acrylic box cover elevated by a black cardboard frame.



**Figure 2.** Details of the photographic setup for silhouette photography via video camera. 1 - external black cardboard box, 2 - diffuser, 3 - acrylic box cover elevated by a black cardboard frame, 4 - shell positioned on the diffuser, 5 - use of putty to ensure horizontality of the coiling axis, 6 - differently sized and aged specimens in apical view, the different degree of rotation is determined by the lip edge; observe how a highly developed, gerontic lip influences the horizontal offset of the coiling axis, 7 – as 6 in anterior view.

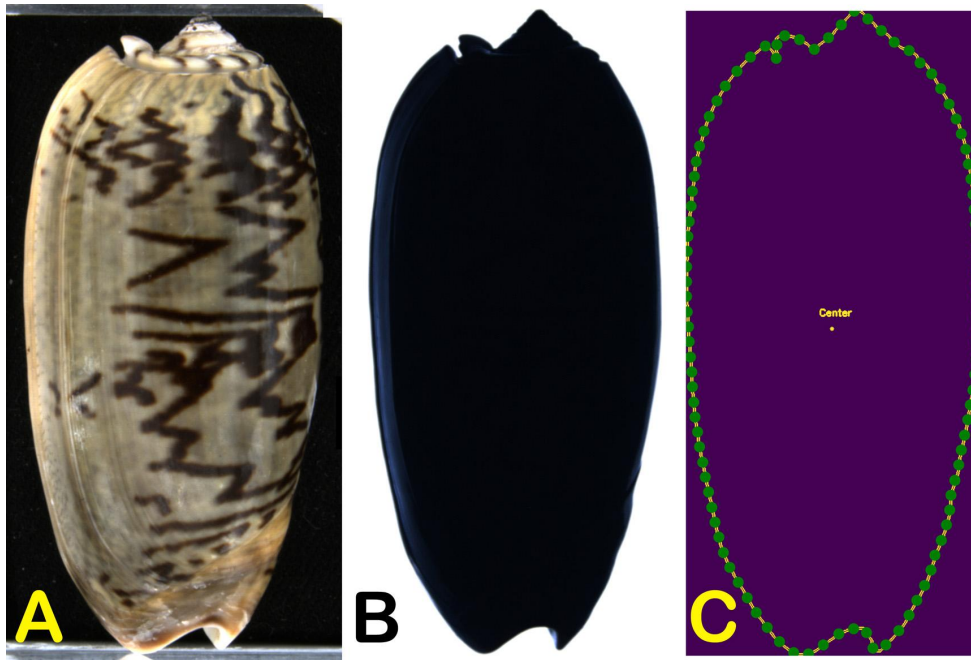


Figure 3. A= a *Viduoliva vidua* as an ordinary photo; B= as a silhouette; C= and as a sampled contour.

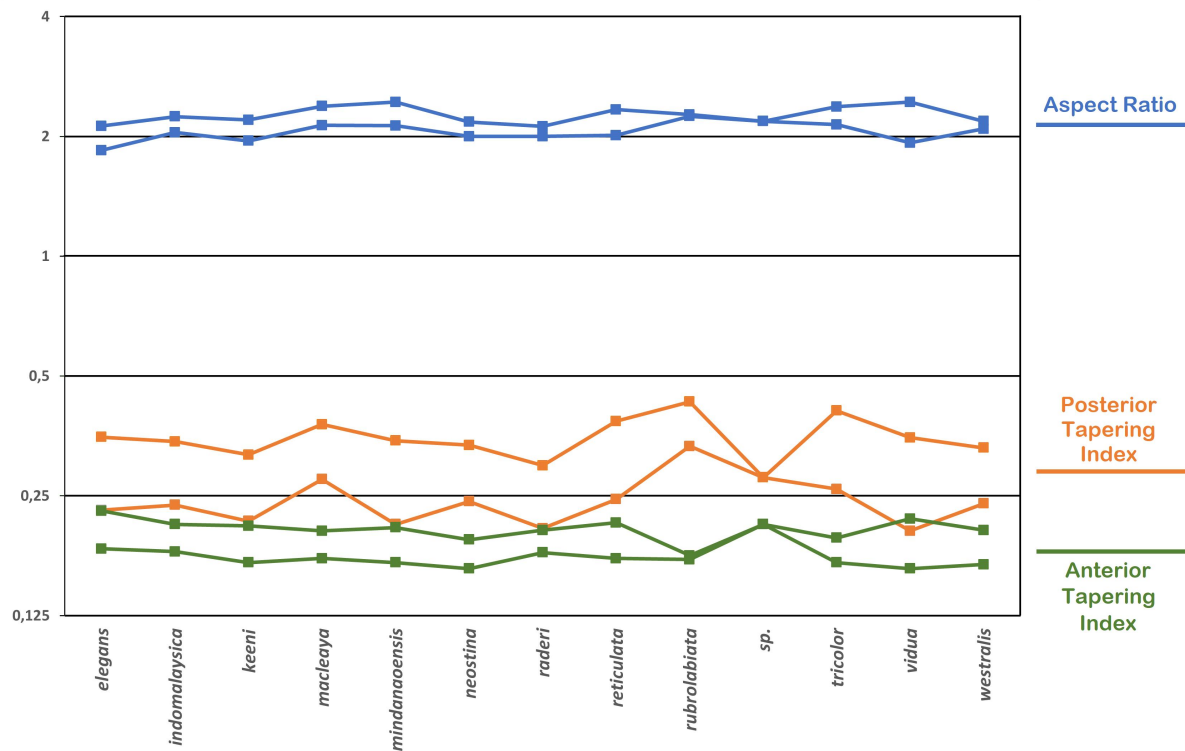


Figure 4. The overlapping range of indexes in most of the species of *Viduoliva*. For each index, the upper line connects the maximum values observed for each species, whose minimum values are connected by the lower line.

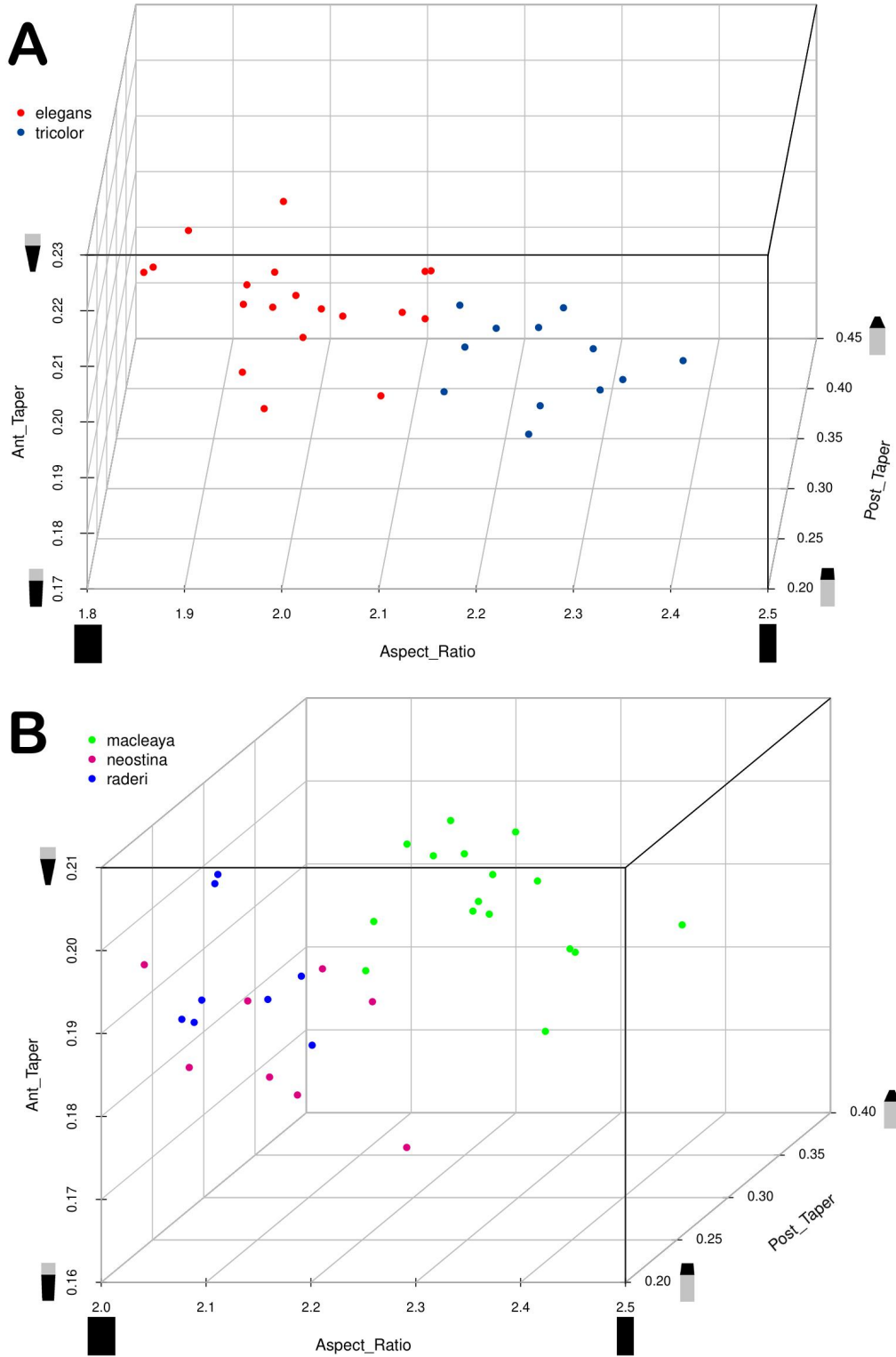


Figure 5. Cases of good separation of clusters. A - *V. elegans* and *V. tricolor*, B - *V. neostina* + *V. neostina raderi* and *V. macleaya*.

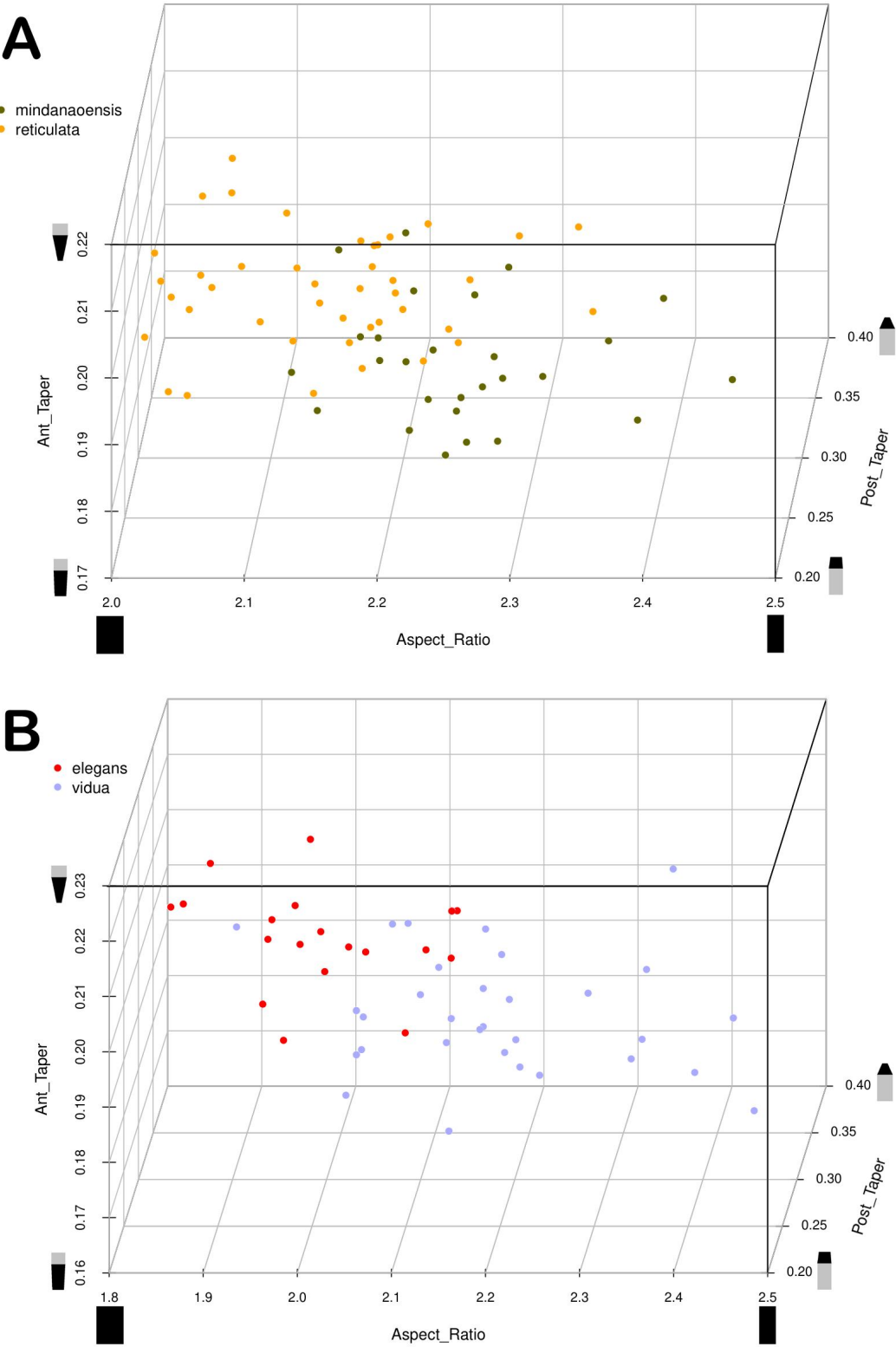


Figure 6. Cases of partial separation of clusters. A - *V. mindanaoensis* and *V. reticulata*, B - *V. elegans* and *V. vidua*.

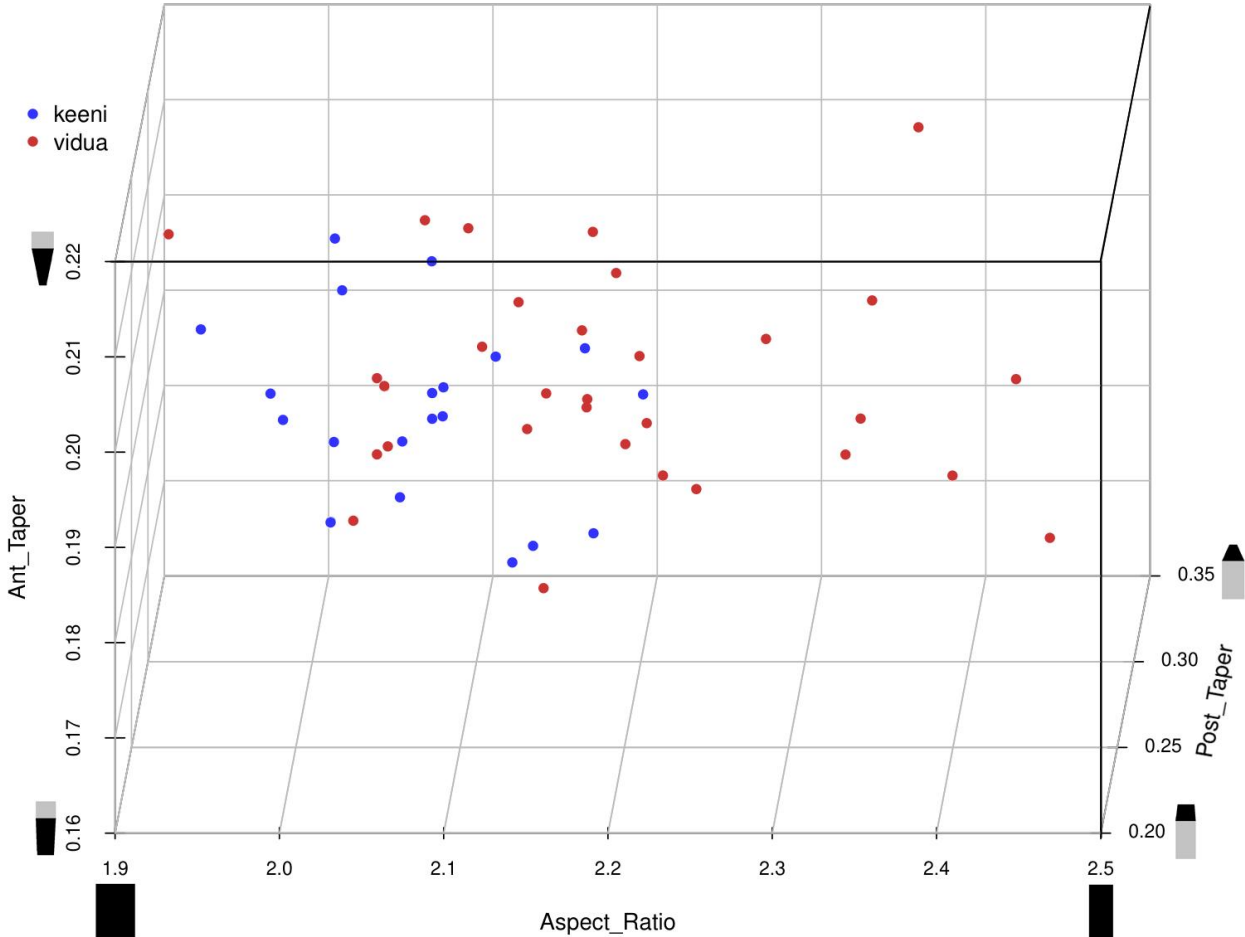
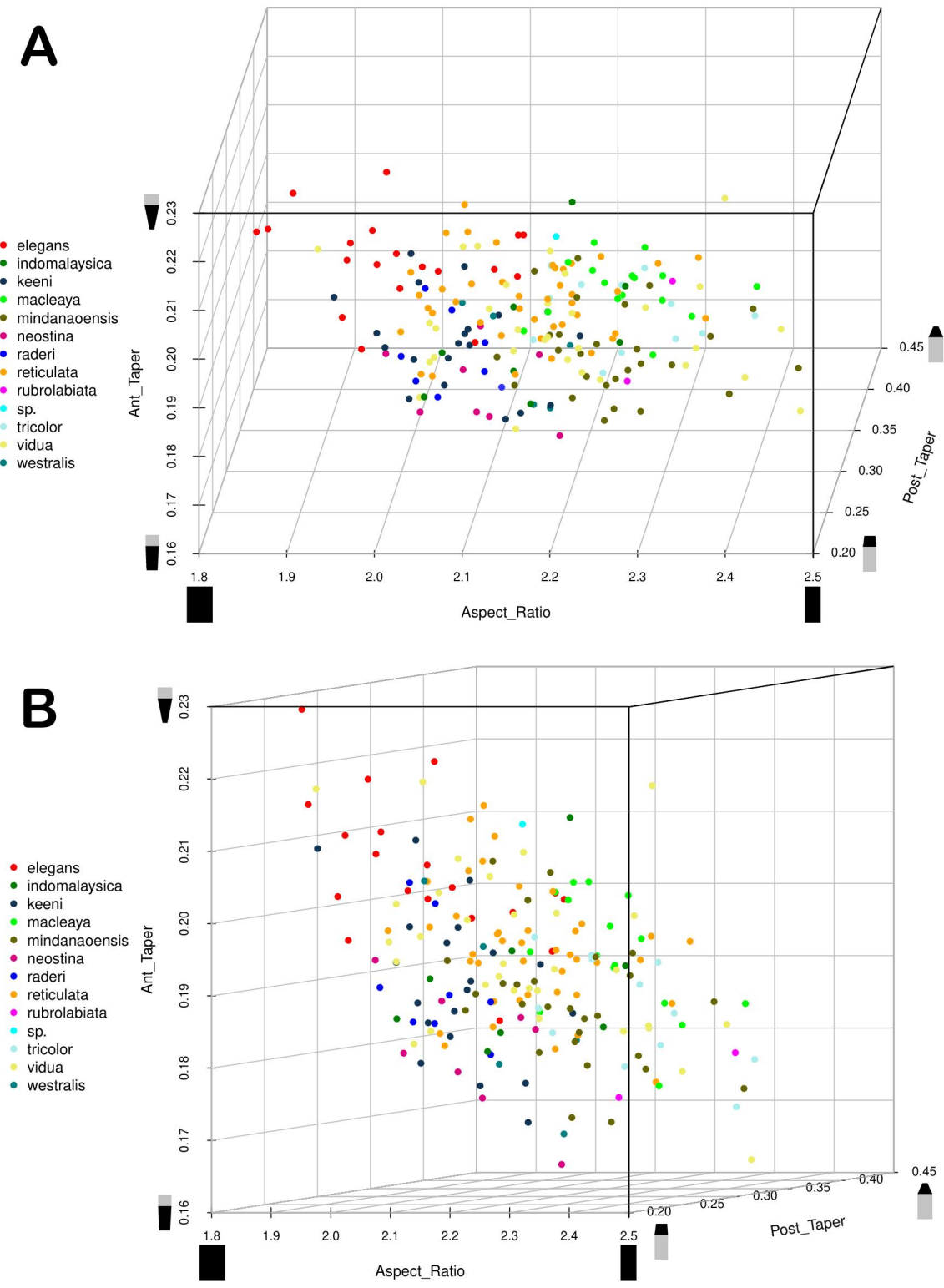
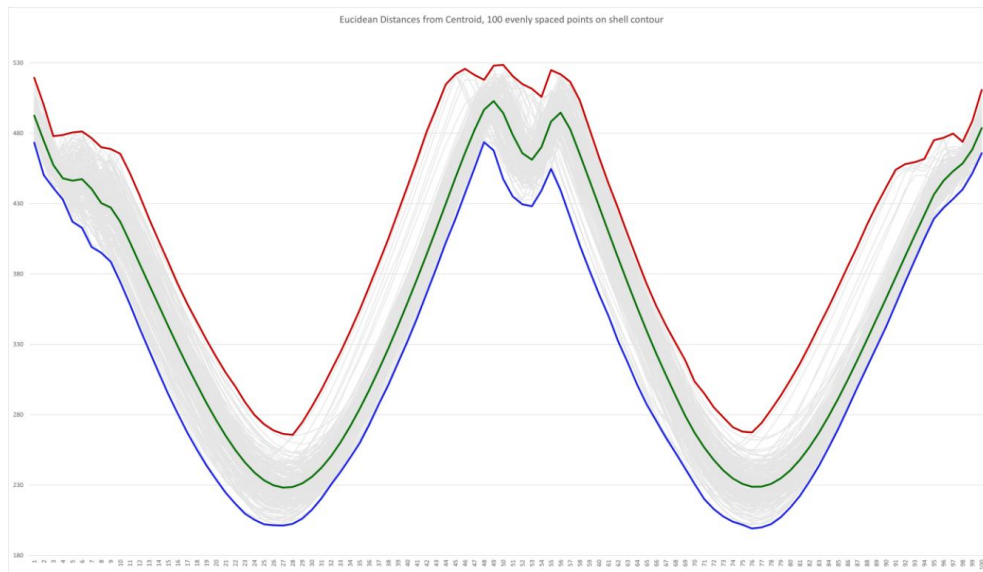


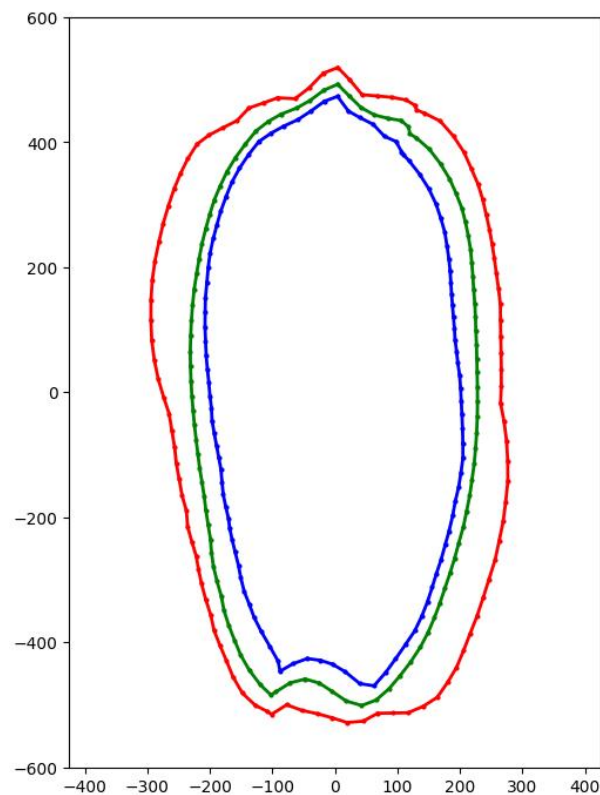
Figure 7. Typical case of overlapping clusters between two random *Viduoliva* species, *V. keeni* and *V. vidua*.



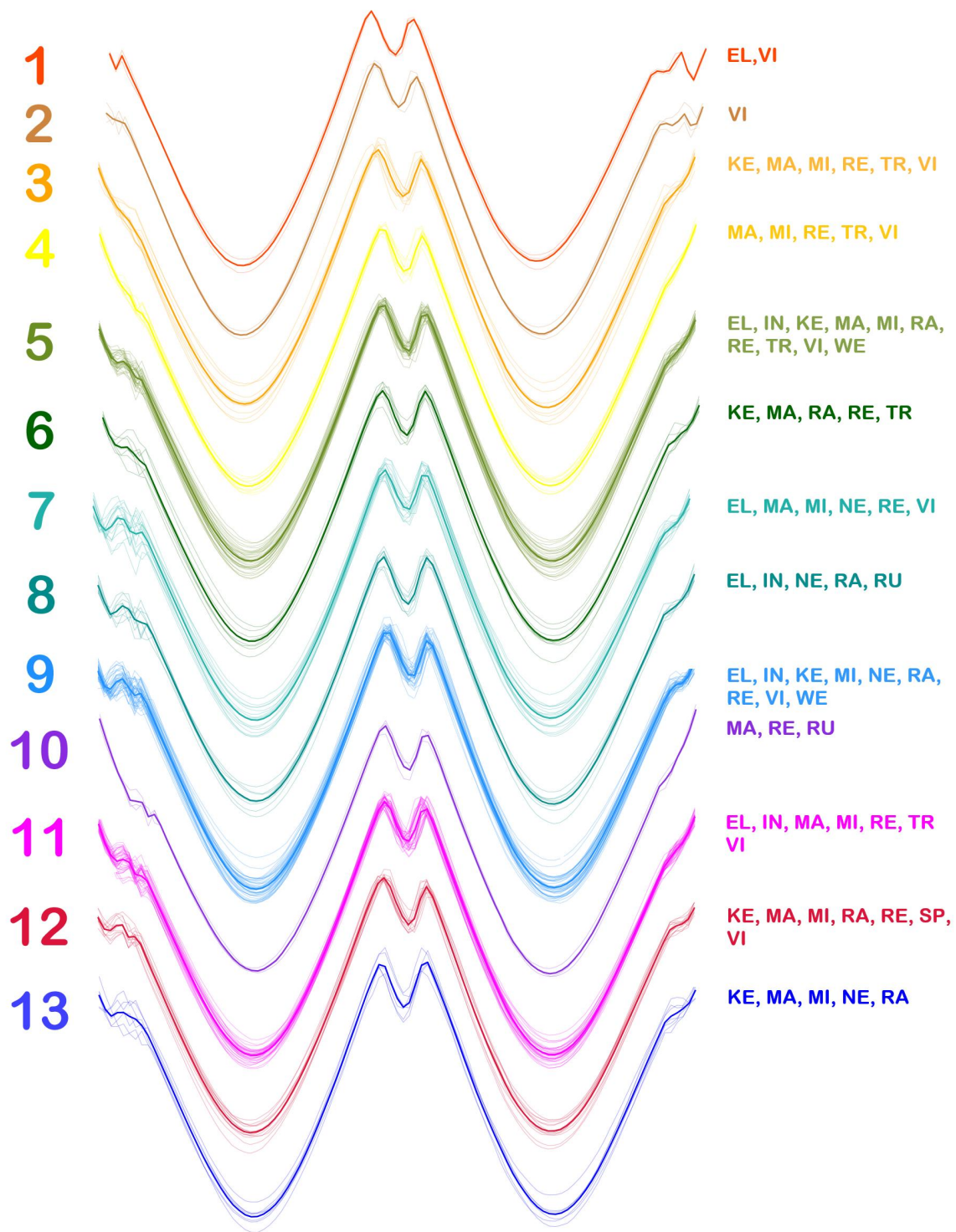
**Figure 8.** Two aspects of the 3D scatter plot of the indexes calculated for all the 199 specimens and 13 species of *Viduolina*.



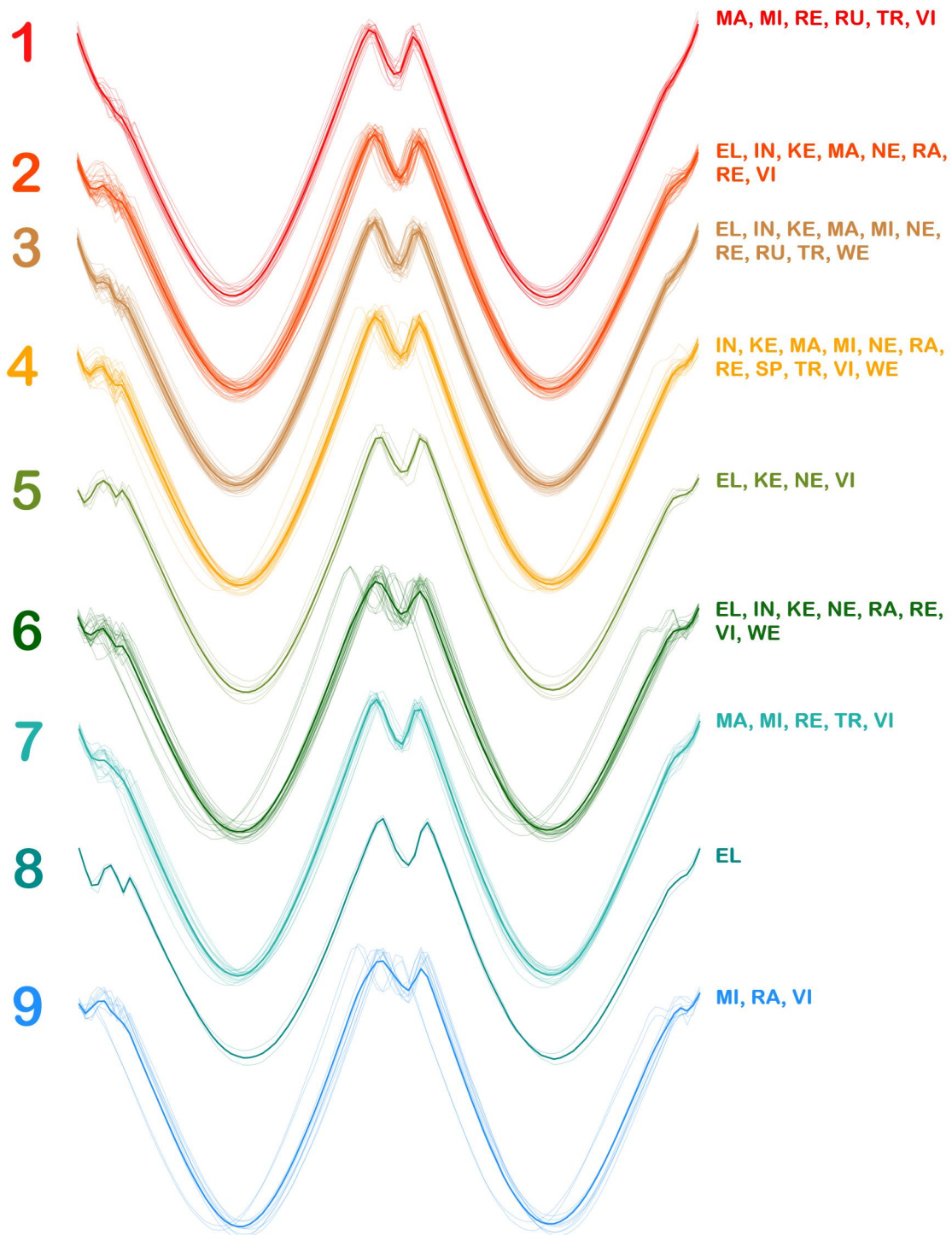
**Figure 9.** Bundle of 199 100-point Euclidean distance vectors from silhouette centroid (grey lines). Blue line - point-by-point minimum values of all the set, green line - point-by-point average values of all the set, red line - point-by-point maximum values of all the set. Vertical column is in pixels.



**Figure 10.** Three calculated *Viduoliva* silhouettes respectively based on maximum values observed (red line), average values observed (green line) and minimum values observed (blue line) in the 100-point Euclidean distance vectors. Only the green outline (“average *Viduoliva*”) can be deemed realistic. Axes graduation is the displacement in pixels from silhouette centroids.



**Figure 11.** The results of agglomerative clustering by pairwise distance (metric=correlation, method=complete) of 199 shape vectors generated from *Viduoliva* silhouettes. See the text for comments. Abbreviations in the left column: EL, *V. elegans*; IN, *V. indomalaysica*; KE, *V. keeni*; MA, *V. macleaya*; MI, *V. mindanaoensis*; NE, *V. neostina*; RA, *V. raderi*; RE, *V. reticulata*; RU, *V. rubrolabiata*; SP, *V. sp.*; TR, *V. tricolor*; VI, *V. vidua*; WE, *V. westralis*.



**Figure 12.** The results of agglomerative clustering by Fréchet distance (metric=precomputed, method=complete) of 199 shape vectors generated from *Viduoliva* silhouettes. See the text for comments. Abbreviations in the left column: EL, *V. elegans*; IN, *V. indomalaysica*; KE, *V. keeni*; MA, *V. macleaya*; MI, *V. mindanaoensis*; NE, *V. neostina*; RA, *V. raderi*; RE, *V. reticulata*; RU, *V. rubrolabiata*; SP, *V. sp.*; TR, *V. tricolor*; VI, *V. vidua*; WE, *V. westralis*.