Combining geometric morphometrics with pattern recognition for the investigation of species complexes

MICHEL BAYLAC\textsuperscript{1*}, CLAIRE VILLEMANT\textsuperscript{1} and GIUSEPPINA SIMBOLOTTI\textsuperscript{2}

\textsuperscript{1}Entomologie, CNRS ESA 8043, Département Systématique et évolution, Muséum National d’Histoire Naturelle, 45 rue Buffon, F-75005 Paris, France
\textsuperscript{2}Dipartimento di Scienze Ambientali, Universita dell’Aquila, Via Vetoio, Localita Coppito 67 100 L’Aquila, Italy

Received 25 July 2002; accepted for publication 11 February 2003

The parasitoid Hymenoptera contain a large number of species complexes that are as yet unresolved by traditional taxonomic studies. Evolutionary studies as well as biological control programmes often require further investigations which cannot rely on simple qualitative morphological characters; in many cases, particularly with dried specimens housed in museums, molecular approaches cannot be used. Recent developments in geometric morphometrics and statistical exploratory approaches open new perspectives for the objective evaluation of morphological characters in this taxonomic context. In this study, geometric morphometrics and pattern recognition approaches were applied to the wing shape and venation of two closely related braconid species considered to differ by subtle qualitative morphological head characters. Exploratory analyses such as kernel density estimates and Gaussian mixture analyses were used to explore the structure of the data in the multivariate morphometric space. Discrimination techniques (linear discriminant functions and neural networks combined with cross-validations) were used to estimate the taxonomic value of qualitative characters. Gaussian mixtures highlighted the existence of two non-overlapping groups. A good congruence was found between one of the two groups and the a priori defined \textit{Bassus tumidulus}. The misclassification rate was higher for \textit{B. tegularis} specimens, which also appeared morphometrically heterogeneous. Discrimination between the two a priori defined species was incomplete with misclassification rates higher than, or equal to, 6%. In most cases, the lack of congruence between species and morphometrically defined subgroups could be related to specimens that exhibited ambiguous qualitative character states. In summary, if two entities are present, they still need to be defined morphologically, while \textit{B. tegularis} heterogeneity calls for further investigation of specimens of known origin and hosts. © 2003 The Linnean Society of London, Biological Journal of the Linnean Society, 2003, \textbf{80}, 89–98.


INTRODUCTION

Hymenopteron parasitoids are a highly speciose group whose species are generally characterized by high levels of host specificity and low levels of morphological differentiation (Unruh & Messing, 1993; Polaszek & Dessart, 1996; Claridge, Dawah & Wilson, 1997). Such a combination of characteristics explains the frequency of unsolved species complexes within the Hymenoptera (Vet \textit{et al}., 1984a,b; Pungerl, 1986; Polaszek & Walker, 1991; Polaszek, Ubeku & Bosque-Perez, 1993; Fernando & Walter, 1997). Many traditional taxonomic investigations group hidden species within a single taxon. Well-documented cases of this, which represent only a small part of the known species complexes, have mostly been evidenced in the course of biological control programmes using morphometrics and genetic investigations (Kenis & Mills, 1998; Kimani-Njogu \textit{et al}., 1997, 2001; Heraty & Polaszek, 2000; Chen, Giles & Greenstone, 2002; Manzari \textit{et al}., 2002). Morphometrics has mostly been used to discriminate between groups defined a priori using morphological or biological information: hosts, geographical origin, phenology, etc. All morphometric investigations have relied on traditional morphomet-
rics. Few partitioned out size (Mosimann & James, 1979; Bookstein, 1989; Klingenberg, 1996), a procedure which enhances the interpretations and may increase the power of the analyses which can otherwise confound two different sources of variability.

The present study deals with a complex of two very closely related species of the braconid genus Bassus: B. tumidulus Nees and B. tegularis Thomson. The genus belongs to the subfamily Agathidinae, species of which are generally polyphagous parasitoids (Nixon, 1986) attacking various species of Lepidoptera. However, host data relating to most of the species are lacking or unreliable. The close morphological similarity between B. tumidulus and B. tegularis induced Nixon (1986) to synonymise the two species. More recently, Simbolotti & van Achterberg (1992) rejected this synonymy on the basis of new discriminant characters: the shape of the scape and the sculpture of the area between the antennal insertions. In B. tegularis, this area appears V-shaped with a median groove, and its surface shows a deep roughness alternating with rather dense punctuations (Fig. 1A). The same area in B. tumidulus appears comparatively less sculptured and less densely punctuated, while the median groove is replaced by a rather smooth depression (Fig. 1C). The scape in B. tegularis (Fig. 1B) is apically more slender than in B. tumidulus (Fig. 1D). However, preparation processes, preservation of specimens as well as conditions of observation (e.g. light, orientation plane) can affect the correct evaluation of the character states. This ambiguity is not unique to the Bassus case, but applies to many insect groups and consequently only experienced people can use such characters reliably. Furthermore, since most qualitative characters are rarely analysed, their distribution properties remain unknown and their discrimination values cannot be assessed.

This has clearly been the case for the Bassus characters. Head and antennal characters involve complex patterns almost impossible to quantify satisfactorily. We therefore turned to a different organ which could be more easily quantified and analysed using morphometric approaches. Our study was based on the wing shape and venation as described by a set of homologous landmarks. Insect wings are mostly two-dimensional structures, easily numerized, and have been shown to provide reliable information at the specific, infra-specific or population levels of various taxa using traditional morphometrics (Guillaumin & Lefebvre, 1974; Ruttner, 1988) or geometric morphometrics (Guillaumin, 1972; Rohlf & Archie, 1984; Baylac & Daufresne, 1996; Haas & Tolley, 1998, Klingenberg et al., 2001; Dujardin, Le Pont & Baylac, 2003). We used genetic morphometric approaches that provide a higher statistical power (Bookstein, 1991; Marcus et al., 1996) and allow for the direct visualization of the geometric transformations of the objects (Rohlf & Marcus, 1993). Geometric morphometrics has proved to be powerful enough to successfully investigate complex taxonomic problems at the species level (Adams & Funk, 1997; Fink & Zelditch, 1997; Fadda & Corti, 2001; Dobigny, Baylac & Denys, 2002), patterns of differentiation between populations of a single species (Baylac & Daufresne, 1996; Corti et al., 1996; Loy, Di Martino & Capolongo, 1996), or interspecific hybridization (Auffray et al., 1996; Monti, Baylac & Lalanne-Cassou, 1998).

In most cases, systematic investigations of species complexes involve a patterning of the morphological variability. This process is almost always done by simple visual inspection, either of the specimens themselves, or of their projections onto multivariate planes when morphometric or other quantitative investigations are used. The main question concerns the validity and the reality of this patterning, since ‘humans are good . . . at discerning subtle patterns that are really there, but equally so at imagining them when they are altogether absent’ (Carl Sagan, Contact 1985, quoted in Cook, 1998). One alternative is to investigate the existence of a ‘natural’ (i.e. based only on morphometric similitude and excluding any a priori taxonomic information) grouping of individuals. Such

Figure 1. Taxonomic characters used to separate Bassus tegularis (A,B) from B. tumidulus (C,D). (A,C) Sculpture of the area between the antennal insertions. (B,D) Shape of scape.
a task is known as classification (Arabie, Hubert & De Soete, 1996; Dunn & Everitt, 2001) or unsupervised learning in the pattern-recognition literature (Fukunaga, 1990; Ripley, 1996). The growing field of pattern recognition has led to the development of powerful exploratory tools which increase the potential for use of morphometric investigations in alpha systematics. Based on the Bassus example, the aim of the present article is to demonstrate that geometric morphometrics combined with the exploratory tools of pattern recognition are well suited to assist objectively in clarifying complex problems such as those frequently encountered in alpha systematics.

MATERIAL AND METHODS

The study dealt with 25 Bassus tegularis and 36 B. tumidulus female specimens of different origin but mainly from France (see Table 1). Species assignment of all specimens was assessed by one of the authors (GS). Hosts for the 55 specimens from the collections of MNHN Paris (France) and RMNH (Naturalis), Leiden (the Netherlands), were unknown. The six remaining individuals (B. tumidulus) were obtained from larvae of the tortricid Sparganothis pilleriana Denis & Schiffermüller in French vineyards (Côte d’Or) in 2001. Sample sizes may appear rather small. They are however, fully representative of usual taxonomic study constraints.

Right forewings were cut at their base. They were temporarily slide mounted and digitized at ×1575 using a video camera connected to a Leitz Periplan microscope. Image definition was 768 × 512 pixels (lines × columns). The coordinates of 15 homologous landmarks (B. tumidulus) were obtained from larvae of the tortricid Sparganothis pilleriana Denis & Schiffermüller in French vineyards (Côte d’Or) in 2001. Sample sizes may appear rather small. They are however, fully representative of usual taxonomic study constraints.

Right forewings were cut at their base. They were temporarily slide mounted and digitized at ×1575 using a video camera connected to a Leitz Periplan microscope. Image definition was 768 × 512 pixels (lines × columns). The coordinates of 15 homologous landmarks (B. tumidulus) were obtained from larvae of the tortricid Sparganothis pilleriana Denis & Schiffermüller in French vineyards (Côte d’Or) in 2001. Sample sizes may appear rather small. They are however, fully representative of usual taxonomic study constraints.

Right forewings were cut at their base. They were temporarily slide mounted and digitized at ×1575 using a video camera connected to a Leitz Periplan microscope. Image definition was 768 × 512 pixels (lines × columns). The coordinates of 15 homologous landmarks (B. tumidulus) were obtained from larvae of the tortricid Sparganothis pilleriana Denis & Schiffermüller in French vineyards (Côte d’Or) in 2001. Sample sizes may appear rather small. They are however, fully representative of usual taxonomic study constraints.

Right forewings were cut at their base. They were temporarily slide mounted and digitized at ×1575 using a video camera connected to a Leitz Periplan microscope. Image definition was 768 × 512 pixels (lines × columns). The coordinates of 15 homologous landmarks (B. tumidulus) were obtained from larvae of the tortricid Sparganothis pilleriana Denis & Schiffermüller in French vineyards (Côte d’Or) in 2001. Sample sizes may appear rather small. They are however, fully representative of usual taxonomic study constraints.

Right forewings were cut at their base. They were temporarily slide mounted and digitized at ×1575 using a video camera connected to a Leitz Periplan microscope. Image definition was 768 × 512 pixels (lines × columns). The coordinates of 15 homologous landmarks (B. tumidulus) were obtained from larvae of the tortricid Sparganothis pilleriana Denis & Schiffermüller in French vineyards (Côte d’Or) in 2001. Sample sizes may appear rather small. They are however, fully representative of usual taxonomic study constraints.

Right forewings were cut at their base. They were temporarily slide mounted and digitized at ×1575 using a video camera connected to a Leitz Periplan microscope. Image definition was 768 × 512 pixels (lines × columns). The coordinates of 15 homologous landmarks (B. tumidulus) were obtained from larvae of the tortricid Sparganothis pilleriana Denis & Schiffermüller in French vineyards (Côte d’Or) in 2001. Sample sizes may appear rather small. They are however, fully representative of usual taxonomic study constraints.

Right forewings were cut at their base. They were temporarily slide mounted and digitized at ×1575 using a video camera connected to a Leitz Periplan microscope. Image definition was 768 × 512 pixels (lines × columns). The coordinates of 15 homologous landmarks (B. tumidulus) were obtained from larvae of the tortricid Sparganothis pilleriana Denis & Schiffermüller in French vineyards (Côte d’Or) in 2001. Sample sizes may appear rather small. They are however, fully representative of usual taxonomic study constraints.

Right forewings were cut at their base. They were temporarily slide mounted and digitized at ×1575 using a video camera connected to a Leitz Periplan microscope. Image definition was 768 × 512 pixels (lines × columns). The coordinates of 15 homologous landmarks (B. tumidulus) were obtained from larvae of the tortricid Sparganothis pilleriana Denis & Schiffermüller in French vineyards (Côte d’Or) in 2001. Sample sizes may appear rather small. They are however, fully representative of usual taxonomic study constraints.

Table 1.

<table>
<thead>
<tr>
<th>Country of origin</th>
<th>Bassus tegularis</th>
<th>Bassus tumidulus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italy</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>France</td>
<td>17</td>
<td>24</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Macedonia</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Unknown origin</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Origin of French specimens</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile de France</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Côte d’Or</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Var</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Others</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 1. Geographic origin of the Bassus specimens

The study dealt with 25 Bassus tegularis and 36 B. tumidulus female specimens of different origin but mainly from France (see Table 1). Species assignment of all specimens was assessed by one of the authors (GS). Hosts for the 55 specimens from the collections of MNHN Paris (France) and RMNH (Naturalis), Leiden (the Netherlands), were unknown. The six remaining individuals (B. tumidulus) were obtained from larvae of the tortricid Sparganothis pilleriana Denis & Schiffermüller in French vineyards (Côte d’Or) in 2001. Sample sizes may appear rather small. They are however, fully representative of usual taxonomic study constraints.

Right forewings were cut at their base. They were temporarily slide mounted and digitized at ×1575 using a video camera connected to a Leitz Periplan microscope. Image definition was 768 × 512 pixels (lines × columns). The coordinates of 15 homologous landmarks (B. tumidulus) were obtained from larvae of the tortricid Sparganothis pilleriana Denis & Schiffermüller in French vineyards (Côte d’Or) in 2001. Sample sizes may appear rather small. They are however, fully representative of usual taxonomic study constraints.

Table 1. Geographic origin of the Bassus specimens

<table>
<thead>
<tr>
<th>Country of origin</th>
<th>Bassus tegularis</th>
<th>Bassus tumidulus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italy</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>France</td>
<td>17</td>
<td>24</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Macedonia</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Unknown origin</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Origin of French specimens</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile de France</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Côte d’Or</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Var</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Others</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 1. Geographic origin of the Bassus specimens

The study dealt with 25 Bassus tegularis and 36 B. tumidulus female specimens of different origin but mainly from France (see Table 1). Species assignment of all specimens was assessed by one of the authors (GS). Hosts for the 55 specimens from the collections of MNHN Paris (France) and RMNH (Naturalis), Leiden (the Netherlands), were unknown. The six remaining individuals (B. tumidulus) were obtained from larvae of the tortricid Sparganothis pilleriana Denis & Schiffermüller in French vineyards (Côte d’Or) in 2001. Sample sizes may appear rather small. They are however, fully representative of usual taxonomic study constraints.

Right forewings were cut at their base. They were temporarily slide mounted and digitized at ×1575 using a video camera connected to a Leitz Periplan microscope. Image definition was 768 × 512 pixels (lines × columns). The coordinates of 15 homologous landmarks (B. tumidulus) were obtained from larvae of the tortricid Sparganothis pilleriana Denis & Schiffermüller in French vineyards (Côte d’Or) in 2001. Sample sizes may appear rather small. They are however, fully representative of usual taxonomic study constraints.

Right forewings were cut at their base. They were temporarily slide mounted and digitized at ×1575 using a video camera connected to a Leitz Periplan microscope. Image definition was 768 × 512 pixels (lines × columns). The coordinates of 15 homologous landmarks (B. tumidulus) were obtained from larvae of the tortricid Sparganothis pilleriana Denis & Schiffermüller in French vineyards (Côte d’Or) in 2001. Sample sizes may appear rather small. They are however, fully representative of usual taxonomic study constraints.
Ininary principal component analysis (PCA) of the covariance matrix of the Procrustes residuals provided a dimension reduction (Fukunaga, 1990) of the morphometric space. As shown for example in Dobigny et al. (2002) and Friess & Baylac (2003), statistical analyses were done in the reduced space and used subsets of the first principal components instead of Procrustes residuals. Log-transformed centroid size was included or excluded successively. For discriminant and neural networks analyses, the exact number of retained components was the minimum number which maximized the cross-validated classification percentages.

Exploratory investigations were used to gain insight into the structure and patterning of the data. These used approaches based on the analysis of density of points in the reduced morphometric space: kernel density estimates (KDE) and Gaussian mixture analyses (Ripley, 1996; Dunn & Everitt, 2001). Essentially, KDEs are smoothing devices. They may be applied to 1-D or 2-D data in order to delineate patterns of density and may be seen as a form of visual non-parametric spatial clustering (Wand & Jones, 1995). Local densities are estimated around each point of the distribution using a window of defined size. Enclosed points are weighted inversely to their distance from the window centre using various symmetric functions, among which Gaussian are the most widely used. Local density estimates are totaled over the whole distribution range, and density patterns are visualized using pseudo-colourization, surfacing or contouring representations. The size of the window controls the degree of smoothing, while the choice of a kernel function is given by practical rules (Wand & Jones, 1995).

Histograms are also an example of density estimators, but they have well known drawbacks, such as the need for larger sample sizes, and the dependence on origin and bin-width. KDEs depend upon window size but in a less critical manner. KDEs were applied to the successive planes of the reduced space in order to investigate the multidimensional morphometric space.

Gaussian mixtures can apply either to uni- or multivariate data. They are well adapted to the unsupervised delineation of clusters of points for which sizes and shapes may vary from group to group. The single restriction is that within-group distributions are multivariate normally distributed (McLachlan & Basford, 1988; Banfield & Raftery, 1993). The selection of the most probable number of groups, as well as of their distributional properties, are carried out using maximum likelihood estimations (Fraley & Raftery, 1998). However, since the maximum mixture likelihood increases with the number of parameters in the model, a Bayesian information criteria (BIC) which penalizes the complexity of the model is used instead. BIC is not a statistical test, and the comparison between alternative models is operated using standard conventions (Fraley & Raftery, 1998): less than 2, the difference is considered to be weak, it is considered to be positive between 2 and 6, strong between 6 and 10, and very strong when greater to 10.

The fact that only Gaussian groups are selected may appear to be a restrictive condition. However, Gaussian assumptions are more likely to apply with morphometric data. Furthermore, size elimination may contribute to the normalization of the data, while projections of non-normal data in a reduced space, as per-

Figure 2. Location of the 15 landmarks on a right forewing of a Bassus tumidulus female. White lines correspond to the links used to visualize the shape transformations.
formed by PCA, tend towards normality (Diaconis & Freedman, 1984). Nonetheless, potential pitfalls of Gaussian mixtures cannot be denied, but current alternative approaches like hierarchical clustering or k-means also suffer from known restrictions and drawbacks (Celeux et al., 1989), such as the search for spherical, non-oblique, well-separated and homogeneous groups, and occasionally from chaining effects (Bock, 1996; Dunn & Everitt, 2001).

Classification rates between the two a priori defined species used cross-validated discriminant functions and neural networks. Neural networks are becoming standard tools due to their high discriminative power and their less stringent statistical assumptions (Ripley, 1996). Finally, the congruence between 'natural' groups and species was analysed by confusion matrices, while misclassified specimens were checked both for their morphological character states and for their wing patterns.

Statistical and pattern recognition analyses used the R statistical language (Ihaka & Gentleman, 1996; Venables & Ripley, 1997) version 1.41 for Windows (http://cran.r-project.org/). Misclassification rates were calculated using leave-one-out cross-validation procedures, which provide higher but more realistic and unbiased estimates than direct estimations. For neural networks, the number of units in the hidden layer, the decay (lambda) parameter, as well as the function (linear, threshold or logistic) were tuned interactively in order to maximize the leave-one-out cross-validated classification rates.

RESULTS

CENTROID SIZE

The ANOVA yielded highly significant results ($F = 30.56, \text{d.f.} = 1.59, P < 10^{-4}$). Specimens attributed to *B. tumidulus* have larger wings (mean centroid-size = 6.22) than those attributed to *B. tegularis* (mean = 6.07). Although the variances remained comparable (0.012 and 0.011, respectively), *B. tegularis* appeared less homogeneous due to some outliers.

SHAPE

Exploratory analyses

43.47% of the shape variance was explained by the two first PCA axes extracted from the covariance matrix, which explained most of the between-group differences (Fig. 3). A total of 11 axes was required to recover 90% of the shape variability. The first axis mainly described a common dominant within-group variability. The shape deformations along this axis mainly involved the pterostigma and the corresponding anterior wing margin. This variability corresponded to an opposition between large pterostigmas, whose base appeared shifted towards the anterior

Figure 3. Plot of the two a priori species limits onto the first plane of the principal component analysis of the covariance matrix: ● = *Bassus tegularis*, ○ = *B. tumidulus* Shape variation along the axes is depicted by superimposed extremes in shape of the wing. Grey lines illustrate the negative sides of the axes, black lines the positive ones.
margin, associated with shorter 2-R1 endings, and small pterostigmas with longer 2-R1 endings.

The second axis allowed for a partial discrimination of the two groups (Fig. 3). Wing deformations along this axis mainly affected the proximal posterior area, the distal anterior area being more stable. In specimens attributed to *B. tegularis* (Fig. 3, dots and black lines), the relative length of 2-CU1 and of m-cu increased. Veins m-cu, 2-SR+M, and the second submarginal cell were all shifted toward the wing apex, while SR1 and 1-M veins were shifted toward the base of the wing. Veins 3-CU1 and cu-a were roughly parallel in *B. tumidulus* (Fig. 3, circles and grey lines), while they diverged in *B. tegularis*. No relationship could be found between the variability in the first PCA plane and the geographical origin of the samples. Allometry was not significant within each species at the 0.05 level (not shown; multiple correlation test using progressive subsets of shape components). Therefore, although the two species differed in both size and shape, no allometric component seemed to be implied in this difference.

Kernel density analyses showed that two main unequal density peaks could be found within the first PCA plane (Fig. 4A, to be compared with Fig. 3), which correspond to two ‘natural’ groups, largely overlapping with the a priori defined species. Nevertheless, specimens at the top, i.e. those which are attributable to *B. tegularis* appeared more heterogeneous with many outliers. KDEs using subsequent PCA axes did not provide clear additional information. Gaussian mixtures analyses provided more detailed insights. In order to allow for possible within-group non-gaussian heterogeneities, we calculated mixture parameters for two groups and more, and checked the congruence with the two species assignments. A two-groups solution emerged with the highest BIC (Fig. 4B). These two groups, although not fully congruent, remained close to the species assignments. Three *tumidulus* specimens were ill-classified, two of which exhibited ambiguous head character states. Of the five *tegularis* ill-classified specimens, two were of French origin and also exhibited uncertain character states. The three remaining exhibited *tegularis* head characters, but two came from Bulgaria and one from Macedonia (out of five and one, respectively, see Table 1). Alternative solutions had BIC values lower by at least two points, a difference which may be considered as significant following the empirical criteria of Fraley & Raftery (1998). These results were obtained using the first two principal components of shape. With three (53.9%) and four components (62.6%) the main solution was a single *tegularis + tumidulus* group, of which four to six peripheral *tumidulus* were excluded. This result may be due to the dominance of the within-group variability which is equally structured in both species. When more components were used, increasing numbers of subgroups were found, but as a rule the misclassification percentages were always higher for *tegularis* specimens than for *tumidulus* ones. These results again highlighted the higher shape heterogeneity of the *B. tegularis* specimens. No further relationships between Gaussian mixtures and the geographical origin of the specimens could be established. Similarly, specimens obtained from *Sparganothis pilleriana* never keyed apart.

**Figure 4.** Exploratory analyses of the patterns of points onto the first PCA plane (compare with Fig. 3). (A) Kernel density estimates visualized by isocontours. (B) Limits of the two groups identified by unconstrained Gaussian mixtures analyses.
deformation along the PCA axes. Again, there was no allometric component involved in the discrimination.

The misclassification percentages reached 10.78% for the ‘leave-one-out’ cross-validation without size (two tumidulus and four tegularis), and 5.39% (one tumidulus and two tegularis) with size included. Neural network cross-validated misclassification percentages were equal to 6.6% (two tegularis and two tumidulus), and to 8.2% (four tegularis and one tumidulus) with and without size, respectively. All specimens misclassified by discriminant functions and neural networks were equally misclassified by Gaussian mixtures. Among the three ill-classified individuals identified by the discriminant function including size, one tegularis was in fact difficult to assign due to the bad preservation of the area between antennal insertions. The two remaining specimens had ambiguous qualitative character states.

DISCUSSION

Exploratory analyses highlighted the presence of two groups within the sample, one of which was congruent with the tumidulus species assignments. In contrast, specimens attributed to B. tegularis appeared more heterogeneous and were split into subgroups. It is worth noting that within-group variability was particularly high and represented a dominant part of the total wing shape variability. Despite this lack of homogeneity, discrimination rates between a priori species were higher than 94%. This result was obtained using both linear discriminant methods as well as neural network approaches. Discrimination implied mostly shape parameters, and the inclusion of size increased the discrimination rates by less than 2%. Cases of non-congruence between Gaussian groups and species involved three tumidulus and five tegularis specimens, of which two and three, respectively, exhibited ambiguous qualitative characters. The same conclusion applied to two of the three specimens misclassified by discriminant functions. It follows that most misidentifications as well as many non-congruence cases clearly resulted because the qualitative head and antennal morphological characters were not fully usable. The fact that no quantitative descriptions and analyses were used to assess their taxonomic value could have resulted in an over-simplification of the patterning, perhaps by focusing on two extreme forms. Finally, three specimens out of eight were wrongly attributed by Gaussian mixtures, although qualitative characters were apparently unambiguous. Such a discrepancy may result either because qualitative characters overlap, or alternatively because some specimens did not fit the parametric conditions of Gaussian mixtures. Due to the relatively low sample sizes such a conclusion is not unlikely.

Four main conclusions were drawn from morphometric analyses of wing shapes: (1) two probable taxa may be effectively found within this dataset, that partly recover the assignments made using qualitative characters; (2) morphological variability within species has been largely underestimated; (3) the fact that three out of six B. tegularis from Bulgaria and Macedonia were misclassified indicates the need for further investigations using specimens of known geographical origins and hosts; and (4) the fact that most misclassified specimens, both by Gaussian mixtures and by discrimination, showed ambiguous qualitative character states may be interpreted in two ways: either qualitative head characters do not allow the definition of non-overlapping taxa, or ambiguous states may correspond to intermediate specimens. In which case the specific status of B. tegularis and B. tumidulus would be debatable. However, simple visual inspection of ill-classified specimens favours of the first explanation.

Figure 5. Size and shape discriminant function between the two a priori species: histograms of the discriminant function and corresponding wing transformations. Black lines = Bassus tumidulus-like shapes; Grey lines = B. tegularis-like shapes.
Traditional and geometric morphometric investigations have demonstrated the limits of simple qualitative approaches in the resolution of complex taxonomic cases. Nevertheless, even the most careful morphometric analyses may be inconclusive, due to the lack of clear patterns given by multivariate investigations. This may happen when groups are not well separated and when no biological information, such as prey, hosts or host-plants, which would define potential groups are available. In such cases, the patterning process still remains a subjective task, as the limits of the human eye apply equally well at distinguishing patterns of biological forms or patterns of points onto multivariate projection planes. This is why classifications and unsupervised learning approaches are of prime importance in order to objectively analyse the results of morphometric investigations.

KDEs and Gaussian mixtures are only a fraction of the still-growing corpus of innovative pattern recognition approaches (Ripley, 1996; Dunn & Everitt, 2001). Both are now widely available and have been extensively tested, although rarely in a morphometric context. Some reported tests (see Ripley, 1996 for examples and references) even used classical morphometric datasets. Most tests used Anderson’s Iris or blue and orange Leptograpsus rock crabs (Campbell & Mahon, 1974). However, most were not conducted within a strict morphometric framework, i.e. using log-transformed distances, and with size and shape partitioned out. Therefore, published results are difficult to compare with traditional ones. Like the Bassus example, they nevertheless clearly demonstrated the potential interest of modern pattern-recognition approaches applied to (geometric) morphometric data. Such a combination could constitute a real alternative to the search for discrete morphological characters. Gaussian mixtures should receive particular attention. They are particularly well adapted for preliminary investigations of ill-defined species, or defined using unassessed morphological characters, highly variable species with a large host range and/or with large geographical distribution or fragmented populations. But their value should be assessed in various taxonomic conditions, particularly with regard to the effects of small sample sizes.

Obviously, molecular approaches should also be used in order to gain insights into the genetic structure of the taxa. However, they call for new samples which cannot – even partly – be destroyed for analytical purposes. It should also be noted that, within the framework of allometry and size/shape partitioning, subtle patterns revealed by geometric morphometrics can be considered to have a clear genetic basis (Auffray et al., 1996; Baylac & Daufresne, 1996; Adams & Funk, 1997). Morphometrics also offers the additional benefit of directly assisting in the investigation and definition of potential morphological characters which could be used for species identifications.

Geometric morphometrics is not restricted to landmark data. Outline methods, (Rohlf & Archie, 1984; Bookstein, 1996; Monti et al., 1998) could also be used as they have been repeatedly shown to supersede simple landmark approaches in various situations (Loy et al., 2000; Friess & Baylac, in press). Finally, analyses should not be restricted to organs or body parts for which differences have already been reported. Geometric morphometrics is powerful enough to reveal subtle shape differences otherwise invisible. It could be applied, as in our Bassus example, to organs which are known to evolve quite rapidly, therefore defining what could be called morphometric markers, as we define molecular ones. The choice of wing markers appears particularly suitable in insects, as indicated by the growing evidence in recent literature (see references in the Introduction). The fact that many parasitoid species have reduced venations is not a real a priori objection: complex phylogeographical patterns in a single dipteran species were successfully resolved using just six wing landmarks (Baylac & Daufresne, 1996), a result later amplified and confirmed by microsatellite studies (M. Baylac & Sarrassat, unpubl. data). It would therefore be particularly interesting to evaluate the taxonomic interest of wing morphometric markers using similar approaches in Hymenoptera families which, like Chalcidoidea, exhibit highly reduced venations and contain a potentially high number of cryptic species.

ACKNOWLEDGEMENTS

The MNHN ‘Morphométrie et Analyses de Formes’ programme and the CNRS GDR 2474 ‘Morphométrie et évolution des formes’ contributed to this study. The authors are grateful to Dr Kees van Achterberg for the loan of part of the tested specimens. A kind thanks also goes to Gilles Sentenac (ITV, Beune) who provided the French specimens obtained from Sparganothis pilleriana. Special thanks to L. Deharveng, G. Dobigny and M. Friess who commented on earlier versions of the draft and to Mrs Palczewski for the correction of the manuscript in English.

REFERENCES
