Infection, Genetics and Evolution 70 (2019) 197–207

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

A template-dependent semilandmarks treatment and its use in medical entomology

ARTICLE INFO

Keywords:
Medical entomology
Landmarks III
Semilandmarks
Template
Digitisation

ABSTRACT

In medical entomology, as well as in many other groups of arthropods, geometric morphometrics has become a powerful tool for species identification and population characterization. The approach lies on the relative position of some anatomical points (landmarks) or, more recently, of curved features (semilandmarks). Landmarks are described by coordinates of points easy to recognize from one individual to another. According to this criterion decreasing levels of homology have been recognized, going from strong (type I) to weak (type III landmarks). Semilandmarks (or sliding landmarks) are points having poor homology like landmarks III, but making it possible to capture curves or surfaces where landmarks are sparse. Their use is becoming increasingly routine.

Superimposition of semilandmarks differ from what is currently applied to landmarks, ways and tools for collecting them may also differ from collecting landmarks. They can be collected by simply digitizing points along a curve or a surface but can also be collected in a more systematic way by the use of a template. In the CLIC package (https://xyom-clic.eu), as well as in the XYOM software (https://xyom.io), we created an algorithm-based template to both collect and align semilandmarks or landmarks III. The use of such template for the final alignment of these special points represents an original approach, so that a comprehensive explanation is required.

Using a published example, we compare in details the results of our method with the ones produced by the currently applied approaches. A close parallelism of information is found. The specificities and limitations of our method are discussed.

1. Introduction

Geometric morphometrics makes use of specific points on a biological form or image of a form located according to some rule: the landmarks. These points are assumed to have biological homology, which is supposed to be expressed by a geometric correspondance among individuals. Here, the word “homology” refers to the positional equivalence of a small biological structure, as small as a point at the requisite scale. Landmarks are homologous in the sense that they are relocatable points, and according to this criterion various levels of homology have been recognized [see type I, II and III landmarks, (Bookstein, 1991)]. Briefly, type I landmarks may be considered as anatomical points or patches recognizable from one individual to another (juxtaposition of tissue types), type II landmarks need some geometric criterion to be applied on a larger, homologous anatomical feature (point of maximum curvature along a boundary), and type III landmarks are designed relative to the position of other landmarks (Zelditch et al., 2004; MacLeod, 2013).

Landmarks I, II and III are processed by Procrustes methods, i.e., each landmark is aligned to its corresponding consensus landmark according to some optimization criterion (most commonly least square criterion, (Rohlf, 1990)). They are sometimes opposed to “pseudo-landmarks” which are the points used to capture curves or close contours “without landmarks”. Pseudolandmarks are not supposed to be homologous, curves or contours do, and they are processed statistically in a different way, focusing on the curve they are representing instead of on each pseudolandmark separately (Kuhl and Giardina, 1982; Rohlf and Archie, 1984; Lohmann, 1983).

The term “semilandmarks” has been introduced after the concept of landmark III, referring to a series of points that are located relative to one another and collectively describing the shape of a curve (Bookstein, 1997a; Bookstein, 1997b) or a surface (Gunz et al., 2005). However, landmarks III and semilandmarks are conceptually similar, even if there are minor differences between them. While landmarks III can focus on a single point, semilandmarks are exclusively about capturing curves (or surfaces) between landmarks. While the position of the semilandmarks along a curve is more or less arbitrary, depending on other landmarks and semilandmark, the position of most landmarks III along the curve is defined according to some geometric construction rule, or template: it must be at the intersection of the template and the curve. In spite of these apparent differences, both kinds of landmarks can be treated the same way (Yee et al., 2011) because they suffer from the same problem, the loss of some level of independence.

The solution to that problem is not unique. Currently, the methods specifically developed for semilandmarks make use of a minimization criterion (Gunz and Mitteroecker, 2012), either minimum bending energy (Bookstein, 1991) or minimum Procrustes distance (Rohlf and Slice, 1990). As they imply the sliding of points along the tangent to the curve (or tangent plane to the surface), the term “sliding landmarks” is also used to refer to semilandmarks. Sheets et al. (Sheets et al., 2006) suggested to slide the semilandmarks projecting them on the perpendicular of the line connecting two adjacent points (Sampson et al., 1996). A technique also using “projection” had been described for 3D curves by Reddy et al. (Reddy et al., 2005). All these methods for semilandmarks processing do not produce exactly the same results (Gunz and Mitteroecker, 2012; Perez et al., 2006).

https://doi.org/10.1016/j.meegid.2019.03.002
Received 21 January 2019; Received in revised form 2 March 2019; Accepted 4 March 2019
Available online 07 March 2019

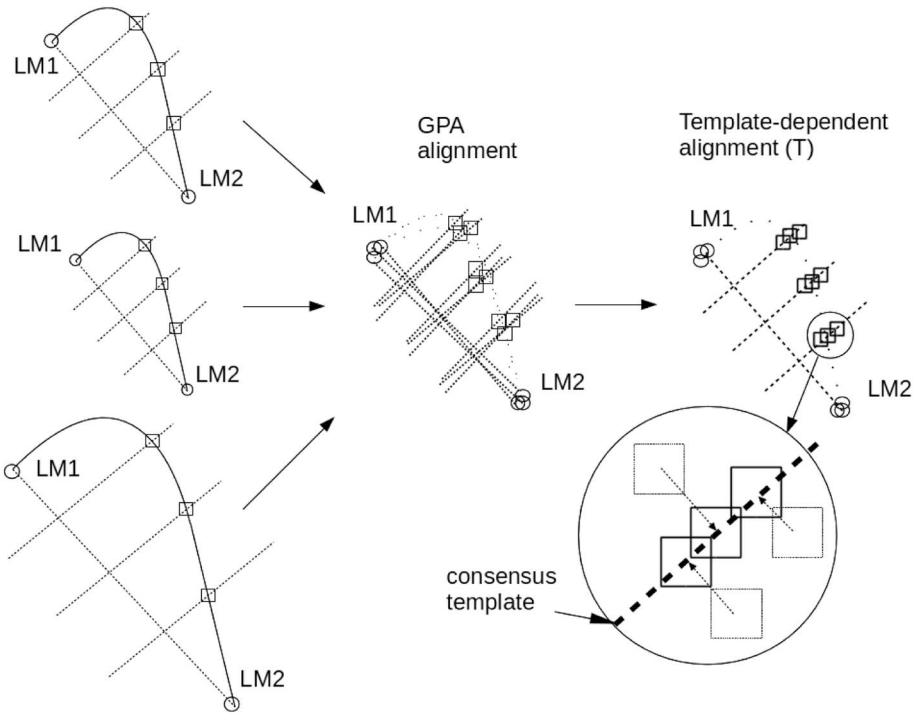


Fig. 1. Visual summary of the T method. There are three objects (left part) with two landmarks (LM1 and LM2) and three semilandmarks (squares) on the curve between landmarks. The template to capture these semilandmarks appears in dashed lines; it is computed by the digitizing software as perpendiculars to a chord between two successive landmarks (see also Fig. 6). After Procrustes alignment of all points (see mid part of the figure), the template is recomputed for the consensus shape (bold dashed line, see right part of the figure); semilandmarks are then orthogonally projected on this consensus template.

We suggest a strategy using perpendicular projections related to the template helping the collection of the semilandmarks. Our proposition involves two main steps, (i) during the digitizing session, the computation of a template for semilandmarks acquisition adapted to each object, and (2) at the final alignment step, after the Procrustes superposition of all points, the computation of a consensus template to which to project the semilandmarks.

2. Method

2.1. The template

Our template used to create semilandmarks consists of evenly spaced lines perpendicular to an interlandmark connecting line. In our algorithm, only successive landmarks are connected by a template, with the possibility to select a given number of evenly spaced perpendiculars. Between two successive landmarks, the algorithm-based template always raises the same number of perpendicular lines. The intersections of these perpendiculars with a curve (see Fig. 1) are the "template-dependent semilandmarks" susceptible to be treated by our method. The algorithm was made in such a way that only one intersection was allowed by perpendicular, and always in the sense of increasing landmarks order (ordinal numbers).

Our approach includes the digitization step, where each semilandmark is precisely located using a geometrically determined protocol (as for landmarks type III, typically). Thanks to the algorithm used by us (https://xyom.io; https://xyom-clic.eu) the template adapts to the size, position and orientation of each new individual. Its architecture is derived from the location of flanking landmarks, not from points primarily sampled on the curve as in Sheets et al. (Sheets et al., 2006) or Reddy et al. (Reddy et al., 2005).

The points do not need to be equidistant, they are just required to be at the crossing of the template with an anatomical feature, as for example a curved vein of an insect wing (Dujardin et al., 2015a; Kaba et al., 2017) (see Figs. 2, 3) or the operculum of eggs of Triatominae (Santillán-Guayasamín et al., 2017) (Fig. 4). Thus, the semilandmarks are scored like landmarks III, which allows the algorithm to be used also for landmarks III capture, as illustrated in Fig. 5.

2.2. The template-dependent semilandmark superimposition method (T)

The statistical method starts with a generalized Procrustes alignment for all points (all being first considered as landmarks) onto a consensus configuration. Thus, each landmark, including semilandmark, is first "aligned" to a consensus "point", i.e. the consensus corresponding landmark. After this first step, a consensus template is computed, i.e. the template of the consensus specimen, and each semilandmark is aligned to this "consensus" template. This final alignment, restricted to the semilandmarks, is achieved by projecting the aligned semilandmark orthogonally to the corresponding perpendicular of the consensus template. This way, all corresponding semilandmarks of various configurations are strictly aligned in a single direction (depending on the consensus template) and do not present differences between them in any other direction (Fig. 1). This procedure could evoke the "perpendicular projections" as described by Reddy et al. (Reddy et al., 2005) or by Sheets et al. (Sheets et al., 2006). The differences with the latter are illustrated Fig. 6.

2.3. Application to wings of predatory and haematophagous hemipterans

Our detailed statistical comparisons of semilandmark treatments were performed on the venation of hemipteran wings (Fig. 2) as studied by Dujardin et al. (Dujardin et al., 2015a). Semilandmarks were collected using three templates, one between landmarks 3 and 4, one between landmarks 4 and 5, and one between landmarks 6 and 7 (see Fig. 2). This sequence was summarized in the following code: {3 6} {4 6} {6 7}, where for each pair between brackets the first digit is a landmark number, the second digit is the number of semilandmarks (Figs. 2,4,5). We compared five different methods, (i) no special treatment given to semilandmarks (N, for no alignment of semilandmarks) (ii) the minimum bending energy method (E) (iii) the minimum Procrustes distance method (D), (iv) the projection method (P) and (v) the template dependent method (T) (see Figs. 7,8,9).

2.3.1. Materials

The landmarks and semilandmarks data were collected on the wings of insects belonging to different subfamilies, genera and species of

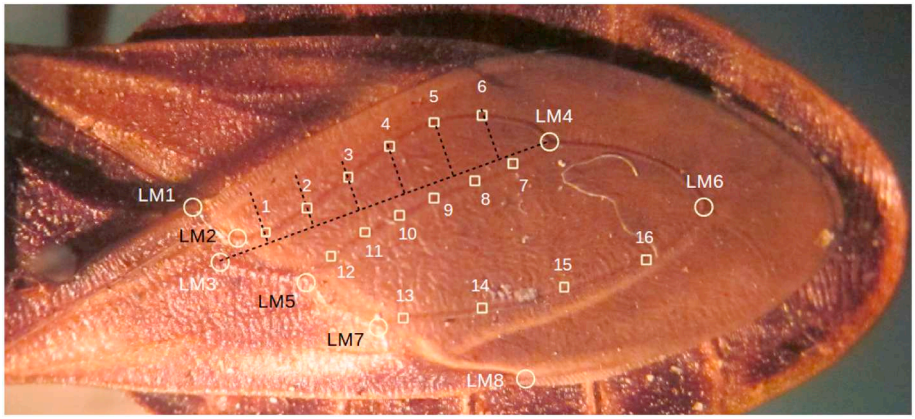


Fig. 2. Wing of Heteroptera, picture extracted from Dujardin et al. (Dujardin et al., 2015a). Circles are landmarks (LM1 to LM8), squares are semilandmarks (numerated separately from 1 to 16). The template for the first six semilandmarks of this wing is shown (dashed black lines). The T protocol code (or semi-landmarks code) for distinguishing landmarks and semilandmarks only needs two numbers by curve: the starting LM and the number of semilandmarks after it. Here, it is {3 6} {4 6} {5 4} as formatted for CLIC (https://xyom-clic.eu); for XYOM (https://xyom.io) this code may be entered as 3,6,4,6,5,4. The last four semilandmarks (13, 14, 15, and 16) were built relative to the positions of the two preceding landmarks (LM5, LM6).

Reduviidae. A total of 138 specimens were analyzed, subdivided into 7 groups having the following sample sizes: 8, 10, 24, 15, 28, 45, 8 (Dujardin et al., 2015a). We used only two subfamilies out of the 22 existing ones: the predatory Ectrychodiinae (genus *Ectrychotes*, $N = 8$), and the haematophagous Triatominae. In a comprehensive phylogenetic tree assembling 18 of the 22 Reduviidae subfamilies, Ectrychodiinae and Triatominae were found very distant taxa (Hwang and Weirauch, 2012). Our sample of Triatominae contained two tribes (out of 5): the Rhodniini and the Triatomini. The Rhodniini ($N = 10$) are composed of two genera that have been shown to be assimilable to a unique one (Monteiro et al., 2000): both of them were included here, namely *Rhodnius* ($n = 7$) and *Psammolestes* ($n = 3$). Our sample of Triatomini was mainly composed of species belonging to the *Triatoma* genus, and subdivided according to the geographic origin: South America ($N = 24$), with 4 species [*Triatoma sordida* ($n = 11$), *T. eratyrusiformis* ($n = 2$), *T. maculata* ($n = 8$) and *T. pseudomaculata* (n = 3)], North America ($N = 15$), with 4 species [*Triatoma sanguisuga* (n = 7), *T. lecticularia* (n = 3), *T. barberi* (n = 3) and *T. rubida* (n = 2)], and Asia ($N = 28$) [*T. bouvieri* (n = 11), *T. cavernicola* (n = 3), *T. leopoldi* (n = 3), *T. migrans* (n = 11)]. Also found in Asia, but also in other continents, we included a sample of *T. rubrofasciata* ($N = 45$), the only species of Triatominae that can be found in both the New World and the Old World (Dujardin et al., 2015b). We finally added 8 specimens of the genus *Linshcosteus*, the only genus of Triatominae other than *Triatoma* which is found in Asia. It has been shown that *Linshcosteus* and *Triatoma*, in spite of their clear-cut body and head morphological differences, are probably the same *Triatoma* genus (Hypsa et al., 2002), and that North American and Asiatic *Triatoma* conform a unique clade, different from the South American one (Dujardin et al., 2015a).

All insects were pinned dry, museum specimens belonging to the British Museum of Natural History (London, UK) and to the Royal Belgian Institute of Natural Sciences (Brussels, Belgium).

2.3.2. Statistical comparisons

Shape variables were the orthogonal projections on a flat space tangent to the consensus of the residual coordinates, the ones obtained after partial Procrustes superimposition (aligned specimens).

To verify whether the final semilandmark sliding had produced shape distortion of the initial objects, we used the Euclidean distance between a configuration (tangent space variables) before and after semilandmark sliding. This was computed for the 138 specimens and compared with the same measures for methods T, E, D and P (see Table 1).

To perform further comparisons, we computed for each method (N, T, E, D and P) the principal component analysis of shape (Table 2), and on that base the relationships among groups as illustrated by an UPGMA-based dendrogram (Fig. 9). The PCA was based on the variance-covariance matrix of the tangent space variables. The rank of the matrix of variance-covariance of the aligned specimens was computed for each method.

In addition, the covariation patterns were compared between methods by the distribution-free S statistics (Table 3). The S statistics is

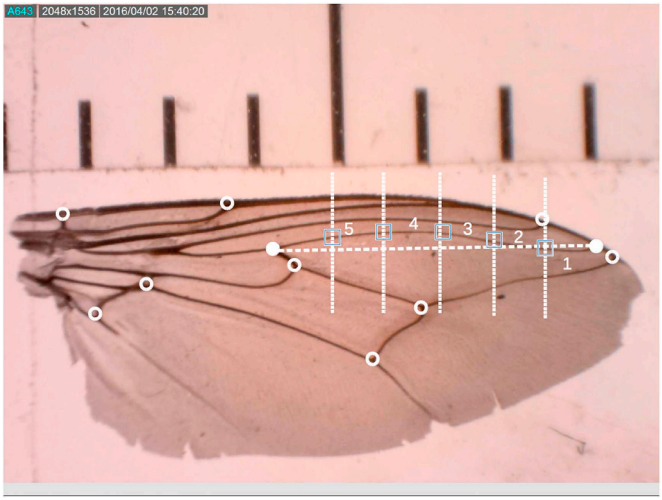


Fig. 3. Wing of tsetse fly (Diptera, Glossinidae). The adjunction of 5 semilandmarks (squares) between two successive landmarks (white circles) out of the 11 landmarks shown on the picture allowed to slightly improve by the recognition of 7 species: 94% if based on 11 landmarks only, 96% if helped by semilandmarks (Kaba et al., 2017). The template for digitizing the semilandmarks of this wing is shown (dashed white lines).

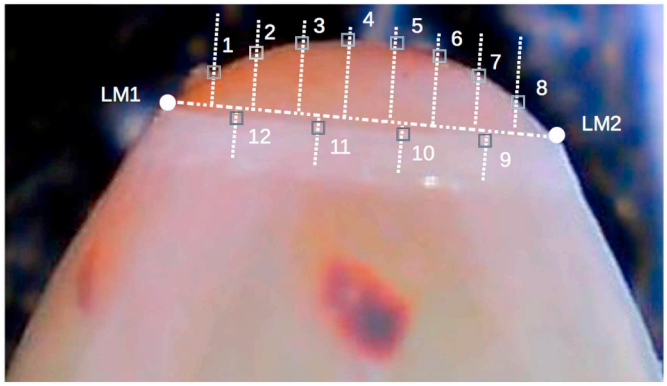


Fig. 4. Egg of kissing bug (Hemiptera, Reduviidae, Triatominii). The shape of the operculum captured by two landmarks (LM1, LM2) and 12 semilandmarks (Santillán-Guayasamín et al., 2017). With only two anatomical landmarks in the tribe Triatominii, the shape of the operculum needs to be captured by a combination of them with semilandmarks. The template for digitizing the semilandmarks of this operculum is shown (dashed white lines). The semilandmarks code is {1 8} {2 4} (or 1,8,2,4), i.e., 8 semilandmarks after LM1, and 4 semilandmarks after LM2.

a recent procedure to compare covariance matrices: it provides information about the nature of possible divergence (S1) in terms of the contributions of differences in matrix orientation (S2) and shape (S3) (Garcia, 2012). The confidence interval of the S1 statistics was computed by bootstrapping 1000 times the set of aligned specimens involved in the pairwise comparisons.

We also estimated the morphological disparity of each group according to each method (Table 5). Morphological disparity was computed as the trace of the variance-covariance matrix of the aligned specimens (Zelditch et al., 2004).

Finally, we computed the Mahalanobis distances between groups and, based on them, we produced validated reclassification scores of the main groups, namely *Ectrychotes* ($N = 8$), *Linshcosteus* (N = 8), *T. rubrofasciata* ($N = 45$), North American *Triatoma* ($N = 15$), South American *Triatoma* ($N = 24$), Rhodniini ($N = 10$) and Asiatic Triatoma ($N = 28$) (Table 4). Since the smallest sample was N = 8, we used the first 7 PCs as input for the Mahalanobis distances computation: they represented from 88% to 92% of the total shape variance, according to the method used (see Table 2).

2.3.3. Software

Both the CLIC package (http://xyom-clic.eu) or the recent XYOM online software (https://xyom.io, see Dujardin & Dujardin, this issue) may be used to digitize and compute template-based semilandmarks (Fig. 1).

However, with the idea to homogenize data prior to inter-methods comparisons, the initial alignment of all points, i.e., the one before the final semilandmarks alignment, was realized with the *Morpho* package, http://sourceforge.net/projects/morpho-rpackage of the R software (https://cran.r-project.org/). The same package was used to obtain the minimum bending energy (E) or distance (D). For the P method, the initial alignment of all points and the sliding process were performed through the IMP software (http://www3.canisius.edu/ sheets/morphsoft.html). Thus, except for the P method, comparative statistics (Tables 1 to 4, Figs. 6 and 7) used specimens primarily aligned (previous to the final semilandmarks alignment) according to the same software (R *Morpho* package). The trees were computed according to the UPGMA algorithm (the *hclust* function of R), from Euclidean distances based on the principal components of shape.

3. Results

The shape distortion brought by the sliding technique T was comparable or even lower than for the E, D or P methods (Table 1).

A striking difference between the T method and the remaining ones appeared in the ranks of the variance-covariance 48 × 48 matrices (8 landmarks plus 16 semilandmarks) of the aligned specimens. The T

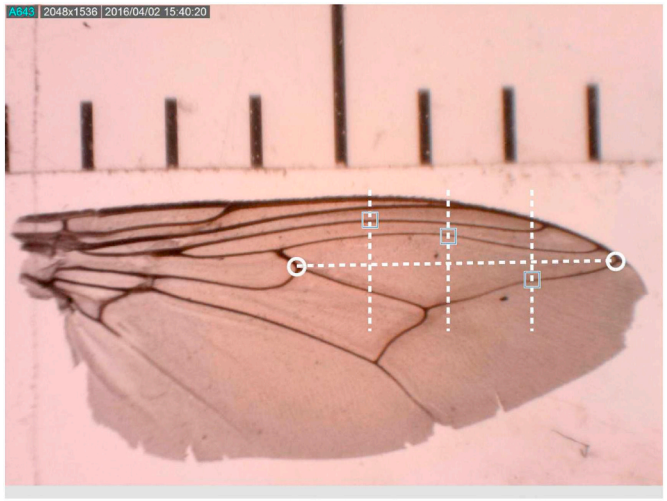


Fig. 5. Wing of tsetse fly (Diptera, Glossinidae). Between two landmarks type I (circles) three landmarks type III (squares) are collected, providing partial information about the relative position of three different veins. The T protocol code is here {1 3} (or 1,3), i.e. 3 points after the first landmark. No other system could reliably perform this kind of data capture if not using an algorithm-based template.

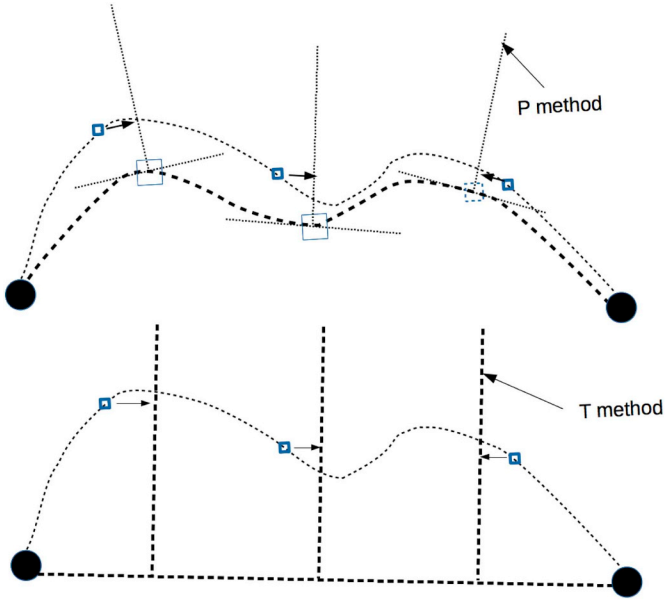


Fig. 6. Visual comparison of some differences between the P (top) and T (bottom) methods. Top: Semi-landmark alignment based on the P method using the SemiLand6 program (http://www3.canisius.edu/ sheets/morphsoft.html): there is a target and a reference form, the semi-landmarks on the target form are projected along lines perpendicular to the curve passing through corresponding semi-landmarks on the reference form. During digitization, there is not necessarily a guide or template to put the semilandmarks. Bottom: for the T method, there is no target and reference form, but a final template computed between successive consensus landmarks on which semilandmarks are projected. During digitization, this template is mandatory, it adapts to each configuration.

method produced a loss of 16 degrees of freedom, versus the maximum of 4 degrees of freedom lost using the other methods. The eigenvalues of the principal components (PCs) of shape presented an apparently similar pattern from one method to another, with, however, a slightly decreasing total contribution of the seven first PCs to the total variance, going from no specific semilandmark treatment (N, 92%) to the other methods (T, 91%; E and P, 89% and D, 88%), clearly visible for the first PC (see Table 2).

All pairwise comparisons of covariance matrices produced very low values of S1 (ranging from 0.0008 to 0.0126, on a scale going from 0 to 1), always with a major contribution of orientation (S2 ranging from 61% to 98%) (Table 3). The closest covariance matrices, with values under the lowest one of the 99% confidence interval, were obtained when comparing the T, D and P methods (Table 3).

Reclassification scores were comparable between methods, with the best average scores obtained with the P method (84.25%), and the lowest one obtained (78.9%) with the D method (Table 4). Except for one group (*Linshcosteus*), there was a perfect agreement between T and D reclassification scores.

Considering the morphological disparities of each sample and their total (Table 5), a different pattern appeared between the E method on one side, and the remaining semilandmarks methods on the other side. The E method slightly increased the values of morphological disparity obtained when no specific semilandmark treatment was applied to the aligned specimens, while the T, D and P methods produced a clear drop in shape disparity.

Whatever the special semilandmark processing - T, E, D or P -, its use was apparently necessary to produce a better fit to the known phylogenetic relationships among the taxa: the N method (no semi-landmarks processing) indeed wrongly clustered the predatory Reduviidae within the haematophagous ones (Fig. 9). Note that the classification trees disclosed a slightly different topology between E and other methods, but an identical one between P, D and T treatments (Fig. 9).

4. Discussion

4.1. Application to medical entomology

As far as we know, four published works on medically important insects used semilandmarks as a complementary capture of shape (Dujardin et al., 2015a; Kaba et al., 2017; Santillán-Guayasamín et al., 2017; Stephens and Juliano, 2012). Of them, three used the template-based digitization described here, on wings of either *Glossina sp.* (Kaba et al., 2017) (Fig. 3) or Triatominae (Dujardin et al., 2015a) (Fig. 2), as well as on eggs of Triatominae (Santillán-Guayasamín et al., 2017) (Fig. 4). One of them (Kaba et al., 2017) examined the possible benefit brought by adding semilandmarks to landmarks, observing a slightly improved classification power (Fig. 3).

4.2. Assisted digitization process

The lines perpendicular to the connection between selected successive landmarks do not need to be equidistant. Their being equidistant was a feature of our current algorithm, and does not mean that semilandmarks themselves were or needed to be evenly spaced. The algorithm used to perform the digitization adapts the template to each new individual, so that it is also able to assist the digitizing process itself by applying the necessary correction to ensure a perfect contact of each digitized point with the template (Fig. 6). This is obtained by projecting the manually digitized coordinates orthogonally to the corresponding perpendicular; such an assistance is not mandatory, but it helps to increase the digitization accuracy.

On the other hand, a template-assisted digitization ensures a reliable repeatability of the digitized points, which is sometimes a concern in the semilandmarks acquisition (Sheets et al., 2006; De Groote et al., 2010; Monnier and Mcnulty, 2010).

4.3. Semilandmarks and landmarks III

Strictly speaking, in spite of using here the word “semilandmarks”, our study considers the type III landmarks as defined by (Bookstein, 1991): they are “intersections of a contour with perpendiculars to a

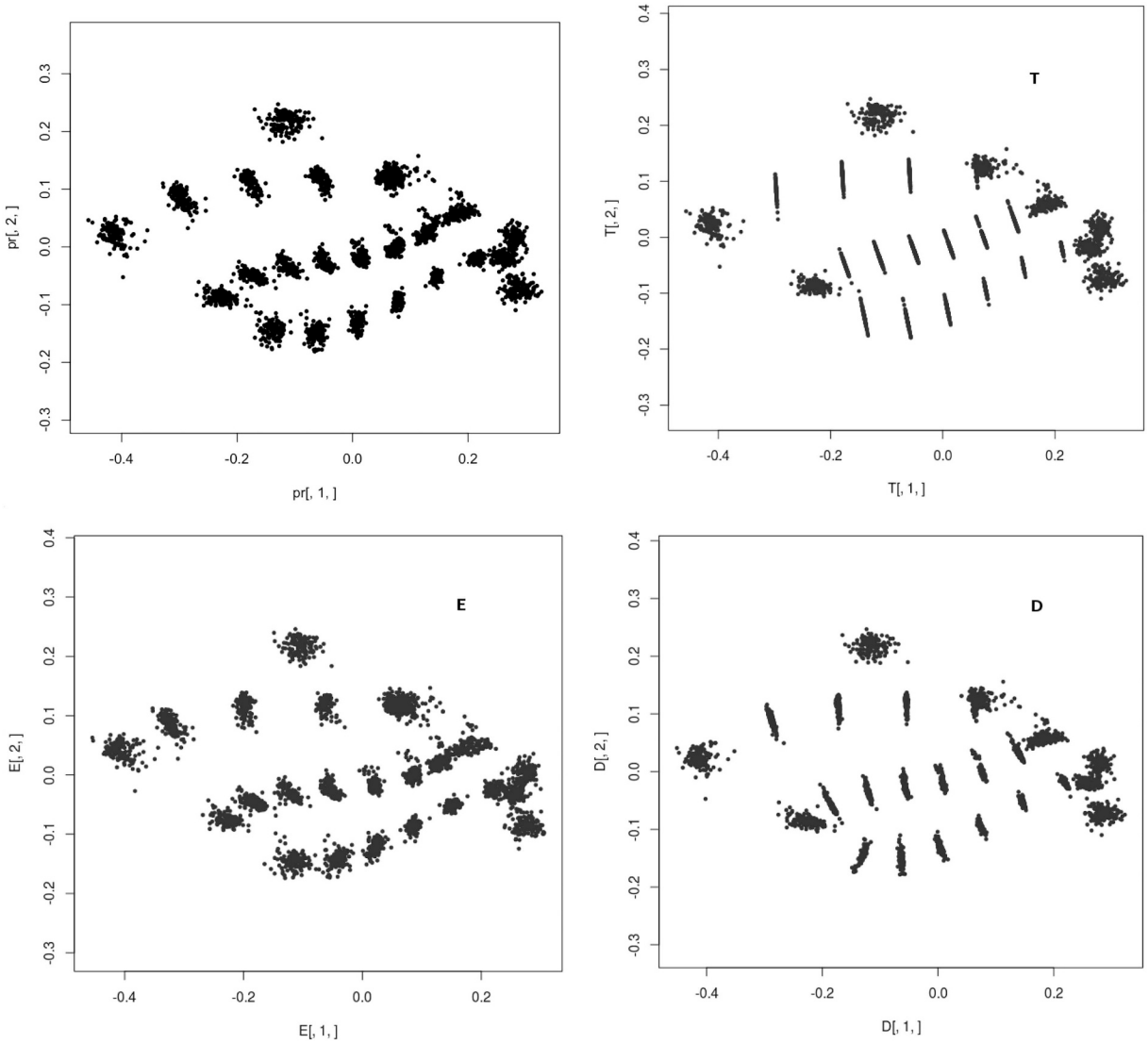


Fig. 7. Aligned specimens according to methods. pr: Aligned specimens, all points superimposed to the consensus configuration according to the least squares criterion, without semilandmarks alignment (N). T: Aligned specimens, with the semilandmarks also aligned along their corresponding consensus perpendicular (cfr. The T method illustrated Fig. 1). E: Aligned specimens according to the minimum bending energy (E) method. D: Aligned specimens according to the minimum Procrustes distance (D) method. Graphics produced with R software (Claude, 2008).

chord (...)". Such kind of landmarks are "deficient" in geometric information because of their dependency on the placement of other landmarks, so that "their displacement is meaningful principally in a single direction" (all quotes taken from page 65 of (Bookstein, 1991)). The special development of type III landmarks, that we call "template-dependent semi-landmarks" also allows, as are doing "sliding landmarks", the description of curved lines between two classical landmarks (Bookstein, 1997a; Bookstein, 1997b), and as such, they can receive the statistical treatment currently given to semilandmarks.

4.4. Comparisons using the example of Hemiptera

In spite of quite divergent venation shapes between two subfamilies (Ectrychodiinae and Triatominae), there was no evidence for shape

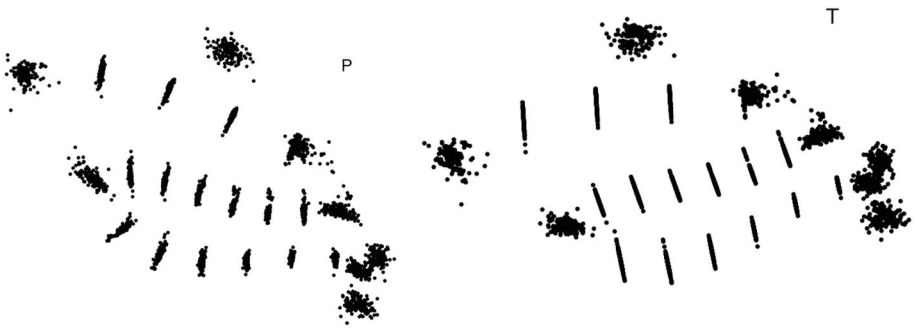
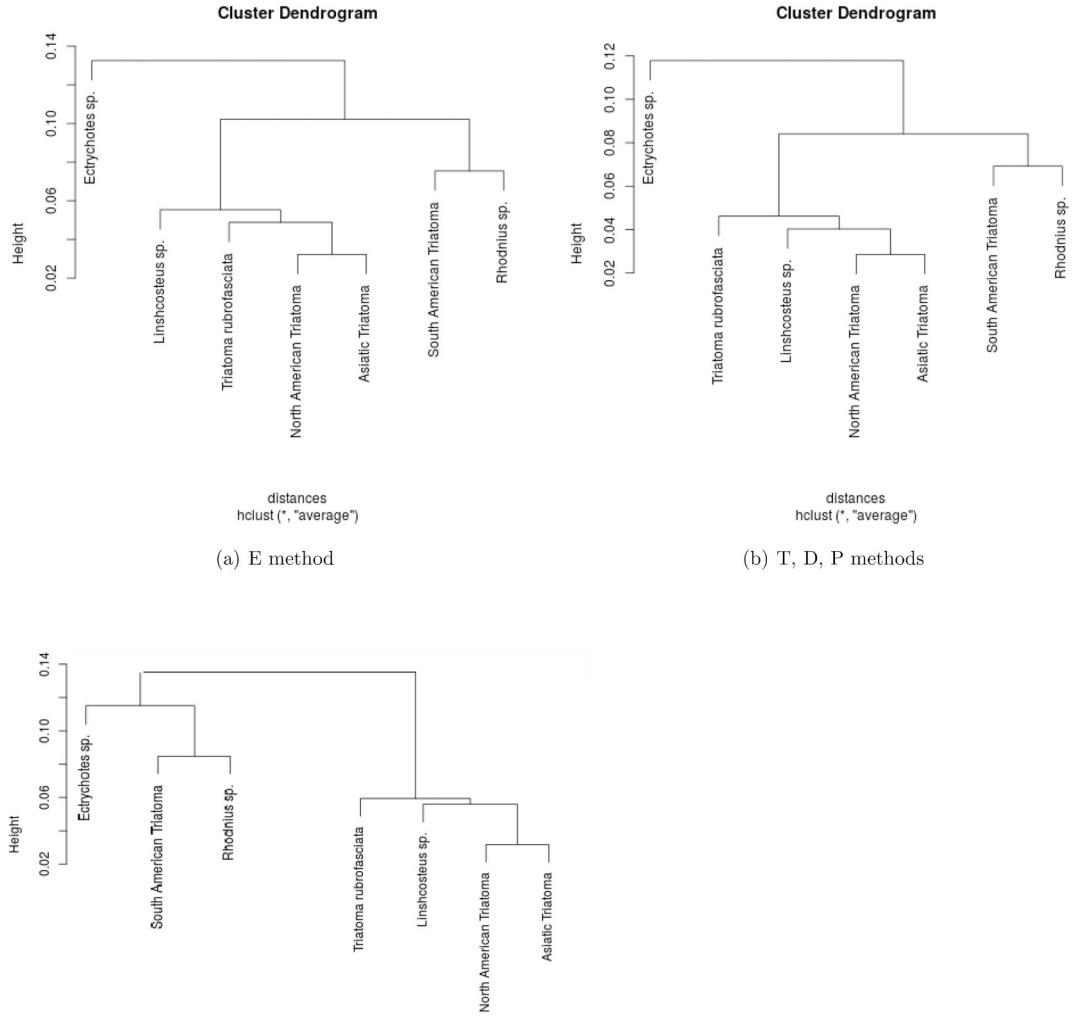


Fig. 8. Aligned specimens according to the P and the T methods. Left: Aligned specimens, with semilandmarks processed according to the perpendicular projection method (here labeled as the P method, using the IMP software from http://www3.canisius.edu/ sheets/morphsoft.html). Right: Aligned specimens, with the semilandmarks aligned along corresponding perpendicular lines of the consensus template (cfr. The T method, as illustrated Fig. 1). Graphics produced with R software (Claude, 2008).



(a) E method

(b) T, D, P methods

(c) No semilandmarks alignment

Fig. 9. UPGMA trees based on Euclidean distances between shapes. Input variables were the principal components of the tangent space variables. (a): semilandmarks processed through the minimum bending energy (E) method; (b): an almost identical topology after semilandmarks processed according to three methods: the template-dependent (T) method, the minimum distance (D) method and the projection method (P); and (c): no semilandmarks processing (N). Graphics produced with the "hclust" function of R software.

Table 1
Euclidean distances (multiplied by 1000) between tangent space variables before and after sliding semilandmarks. T; template-dependent method; E, minimum bending energy criterion; D, minimum distance criterion and P, perpendicular projection method.

Euclidean distance	T	E	D	P
Minimum	2.03	2.36	1.87	1.84
Maximum	56.37	64.58	56.12	52.16
Mean	10.66	12.83	10.31	10.94
Stdev	9.11	11.13	8.86	8.79

distortion induced by the T method, at least no evidence that possible distortion would be more important that the ones scored for the other methods (Table 1).

The most obvious statistical change induced by the T method was the ranking of variance-covariance matrix of aligned specimens. The number of degrees of freedom lost was predicted by the number of template-dependent semilandmarks. Because of the orthogonal projection of each homologous semilandmark onto the corresponding line of the consensus template, each semilandmark could differ from other homologous ones in one direction only, and therefore it lost one degree of freedom.

Table 2
The seven first principal components of shape (PC1 to PC7), shape as orthogonal projections of the aligned specimens on the plane tangent to the consensus. Decimal values are eigenvalues (multiplied by 1000) of each component, and percent values are the respective contributions of each component to the total variance. The total contribution of the seven components are on the first row, beside the method symbol. N, no semilandmark processing at all; T template dependent method; E, minimum bending energy criterion, D minimum distance criterion and P, perpendicular projection method.

	N	92%	T	91%	E	89%	D	88%	P	89%
PC1	4.826	57%	3.086	50%	2.682	42%	1.613	35%	1.613	36%
PC2	1.005	12%	0.957	15%	1.075	17%	0.881	19%	0.880	19%
PC3	0.864	10%	0.567	9%	0.724	11%	0.544	12%	0.523	12%
PC4	0.451	5%	0.313	5%	0.377	6%	0.328	7%	0.319	7%
PC5	0.303	4%	0.281	5%	0.283	4%	0.236	5%	0.240	5%
PC6	0.244	3%	0.232	4%	0.268	4%	0.229	5%	0.229	5%
PC7	0.185	2%	0.183	3%	0.211	3%	0.176	4%	0.177	4%

Due to the different natures of statistical treatments T, E, D and P,

Table 3
S1 refers to the general shape differentiation, S2 to the component of S1 attributable to the orientation of the matrix, S3 to the component attributable to the difference in shape (Garcia, 2012). Since S1 = S2 + S3, we represented S2 and S3 as percents of S1. The S statistics were computed on the eigenvalues expressed as proportions of the total variance, and divided by 8 (the maximum for S1) so that they could vary between zero and 1. The two first columns show the S1 range of variation after 1000 bootstraps on corresponding aligned matrices. All observed S1 values were very low, comparisons with an asterisk indicate an observed S1 lower than the 99% lower limit of the confidence interval. N, no semilandmark processing at all; T, template dependent method; E, minimum bending energy criterion, D minimum distance criterion and P, perpendicular projection method.

		S1:Upper 99%	S1:Lower 99%	S1:Observed	S2	S3
N	T	0.02393	0.00928	0.01097	82%	18%
	E	0.00900	0.00228	0.00242	61%	39%
	D	0.02200	0.00792	0.00939	78%	22%
	P	0.03684	0.01661	0.02112	93%	7%
T	E	0.02505	0.01057	0.01261	80%	20%
	D*	0.00925	0.00126	0.00079*	91%	9%
	P*	0.01347	0.00638	0.00600*	98%	2%
E	D	0.02296	0.00997	0.01178	76%	24%
	P	0.03386	0.01593	0.01963	90%	10%
D	P*	0.01209	0.00358	0.00300*	98%	3%

Table 4
Validated reclassification scores, using as input the seven first principal components of shape. N, no semilandmark processing; T, template dependent method; E, minimum bending energy criterion, D minimum distance criterion and P, perpendicular projection method; N. Am., North and S. Am., South.

Group	N	T	E	D	P
Ectrychotes	100.0%	100.00%	100.00%	100.00%	100.00%
Linshcosteus	50.0%	50.00%	25.00%	25.00%	44.44%
T.rubrofasciata	95.6%	95.56%	91.11%	95.56%	95.65%
N. Am. *Triatoma*	66.7%	53.33%	73.33%	53.33%	60.00%
S. Am. *Triatoma*	95.8%	100.00%	100.00%	100.00%	100.00%
Rhodniini	100.0%	100.00%	100.00%	100.00%	100.00%
Asiatic *Triatoma*	78.6%	78.57%	82.14%	78.57%	89.66%
Average	83.8%	82.5%	81.7%	78.9%	84.25%
Standard deviation	19.45%	22.40%	29.59%	29.36%	22.64%

Table 5
Morphological disparity, for each group of insects and for the total sample. It was computed as the trace of the variance-covariance matrices of the aligned specimens. N, no semilandmark processing at all, T; template dependent method; E, minimum bending energy criterion; D minimum distance criterion and P, perpendicular projection method, N. North and S. South.

Group	N	T	E	D	P
Ectrychotes	0.00259	0.00224	0.00292	0.00217	0.00249
Linshcosteus	0.00291	0.00257	0.00316	0.00244	0.00277
T. rubrofasciata	0.00255	0.00206	0.00265	0.00191	0.00195
N. American *Triatoma*	0.00234	0.00189	0.00272	0.00180	0.00191
S. American *Triatoma*	0.00289	0.00243	0.00294	0.00232	0.00241
Rhodniini	0.00459	0.00400	0.00442	0.00375	0.00416
Asiatic *Triatoma*	0.00285	0.00244	0.00309	0.00227	0.00234
Total	0.00615	0.00483	0.00620	0.00453	0.00452

minor variation may be expected also, like for instance the clear difference in morphological disparity between the T and the E methods (Table 5). A lower metric disparity had already been observed for the D method relative to the E one, and attributed to the D method removing "all the tangential variation along outlines" (Perez et al., 2006). We would interpret this as a reduction, not a removing, of the tangential variation along outlines: a complete removing is obtained by the T method, and witnessed by the corresponding loss of degrees of freedom. Such removing (T) and reduction (D, P) are obvious on the figures showing the aligned specimens (Figs. 7, 8), where the T, D and P methods show the semilandmarks plotted along an imaginary line, perfectly (T) or closely (D, P). All methods (including N, the lack of semilandmark treatment) agreed, however, in their relative estimations of group variability: they indicated the highest disparity for the Rhodniini sample, in accordance with the known morphological differences of the two genera composing this sample, and the lowest variability for the North American group, composed indeed of very close species (Dujardin et al., 2015a).

The four semilandmark treatments (T, E, D and P) showed a clear improvement of the phylogenetic signal relative to no treatment at all (Fig. 9). Such discrepancy in respecting the phylogenetic signal might just be of a contingent nature, depending on the selection of semilandmarks in our sample (it disappeared for instance when removing the last four semilandmarks, details not shown).

The comparisons of the covariance matrices of aligned specimens between methods (Table 3) disclosed very low values of shape differentiation (from S1 = 0.0008 to 0.0126, on a scale with maximum S1 = 1). Maybe which such low values it is risky to make comparisons, however it is worth noting that the T, D and P covariance matrices were almost identical, with most of their slight divergence due to orientation.

The T method of semilandmarks treatment was closer to the D and P methods than to the E one, at least after the following analyses: (i) visually, the effects of D, P and T methods on images superpositions were very similar, aligning the semilandmarks along a single line (Fig. 7, right); however, they differed in the directions of these alignments: they were parallels for the T method because of the template built that way, they were more or less convergent to a point for the D and P methods because their direction must be perpendicular to the tangent of the estimated curve (Figs. 7, 8), (ii) the comparisons of covariance matrices showed a nice agreement with these visual patterns: the T covariance matrix was almost identical to the D and P ones, with most of their very low divergence due to orientation (see Table 3), (iii) the tree topology based on Euclidean distances between principal components of shape, as well as (iv) the reclassification scores based on Mahalanobis distances, were similar between the T, D and P methods. Thus, the T method produced an information parallel to the one of the D and P methods, themselves almost identical.

In this example, the various methods (T, E, D and P) produced very similar results. They were also congruent based on other data comparisons, although not detailed ones (Kaba et al., 2017; Santillán-Guayasamín et al., 2017).

4.5. Specificities of the T method

Contrary to the current techniques, and importantly, the template architecture is derived from the location of flanking landmarks, not from points primarily sampled on the curve (Sheets et al., 2006; Reddy et al., 2005). The T method identifies each semilandmarks through a very compact code. In the example above (Fig. 2), it had 6 symbols to examine three curves defined by a total of 19 points, i.e., {3 6} {4 6} {6 7}; the same points needed 48 symbols for the E or D method, and 120 for the P method.

Because it is linked to an algorithmically built template, the T method allows a computer-assisted semilandmarks digitization, which improves both accuracy and repeatability of semilandmarks acquisition. But even with computer assistance, the technique imposes a careful manual digitization of semilandmarks, inviting the user not to score a high number of them. This should not be an obstacle since curves do not necessarily need many semilandmarks to be satisfactorily represented (MacLeod, 2013), and the sampling of a very high number of semilandmarks is not necessarily a desirable practice (De Groote et al., 2010; Webster and Sheets, 2010) (although it could be useful for estimating missing data (Mitteroecker and Gunz, 2009)).

Our technique ensures a complete control for each semilandmark location, and allows to sample a very small number of semilandmarks,

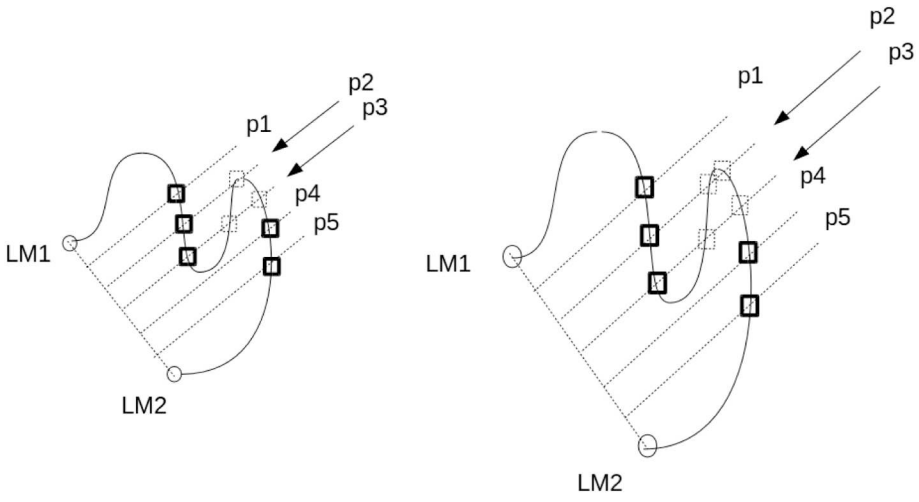


Fig. 10. One limitation about our algorithm: the reentrant curves. Left and right sketch figures are two different objects where a complex curve is studied between homologous landmarks LM1 and LM2. p1 to p5 are the perpendiculars of the template (dashed lines), where p2 and p3 are intercepting the curve more than once. Only one intercept must be kept, the same one for each object, severely limiting the capture of shape.

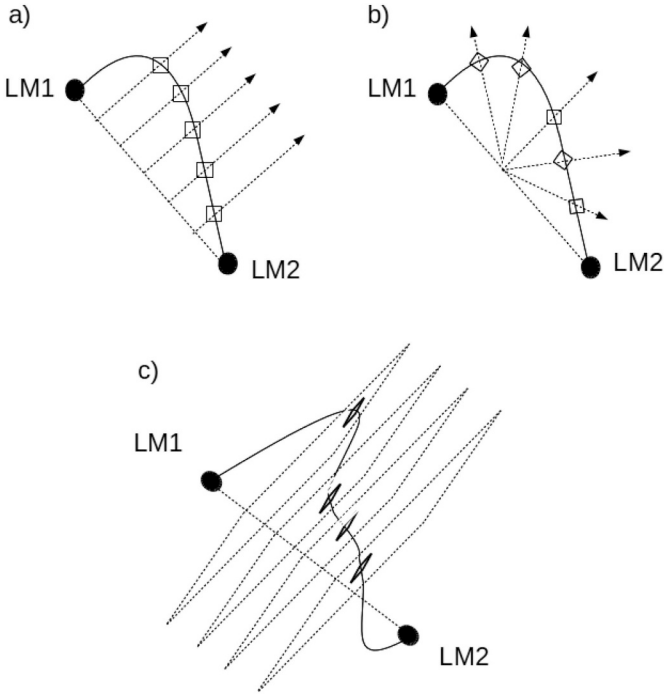


Fig. 11. Some possible templates (dashed lines). a) the template used in this study, having the "comb" shape as in IMP software (http://www3.canisius.edu/ sheets/morphsoft.html); b) another template suitable for 2D curves; c) a possible template for 3D curve, made of plans raised from perpendiculars crossing (horizontally, or with the help of a third landmark) the line connecting two landmarks. Squares are semilandmarks. In all cases, the template architecture must be derived from the location of flanking landmarks (LM1, LM2), not from points sampled on the curve.

even a single one (landmark III, see Fig. 5). Thus, curves and any other anatomical feature can be sampled economically on the same figure.

4.6. Limitations of the T method

Our method cannot be applied to data collected in another way, i.e., without the use of an algorithm-based template. Moreover, the digitizing algorithm itself may represent a limitation. Our own algorithm is among the simplest ones, limited to sampling between two successive landmarks. Any other possible arrangement of semilandmarks construction relative to a reference system is valid (see for instance Fig. 11 b), as long as the same constructing method is used (after Procrustes superposition) to align the semilandmarks to the consensus template.

Because it allows only one interception to be scored by perpendicular, our algorithm would capture reentrant curves very partially (Fig. 10). As for the current techniques of semilandmarks sampling, no missing data are allowed, i.e., in our algorithm each perpendicular of the template must be intercepted by the interlandmark form. These limitations might be overcome by developing algorithms for the construction of more adapted templates.

4.7. From 2D to 3D curves and surfaces

It seems possible to create a template adapted for 3D curves using planes instead of perpendiculars (Fig. 11). However, for 3D surfaces, the perpendicular planes would be intersected by a curving line, which would itself be digitized according to an adapted template. With such an approach, the template would be constructed between pairs of type III landmarks (Fig. 12). Neither the 3D curve (Fig. 11) nor the 3D surface templates (Fig. 12) have been tested yet.

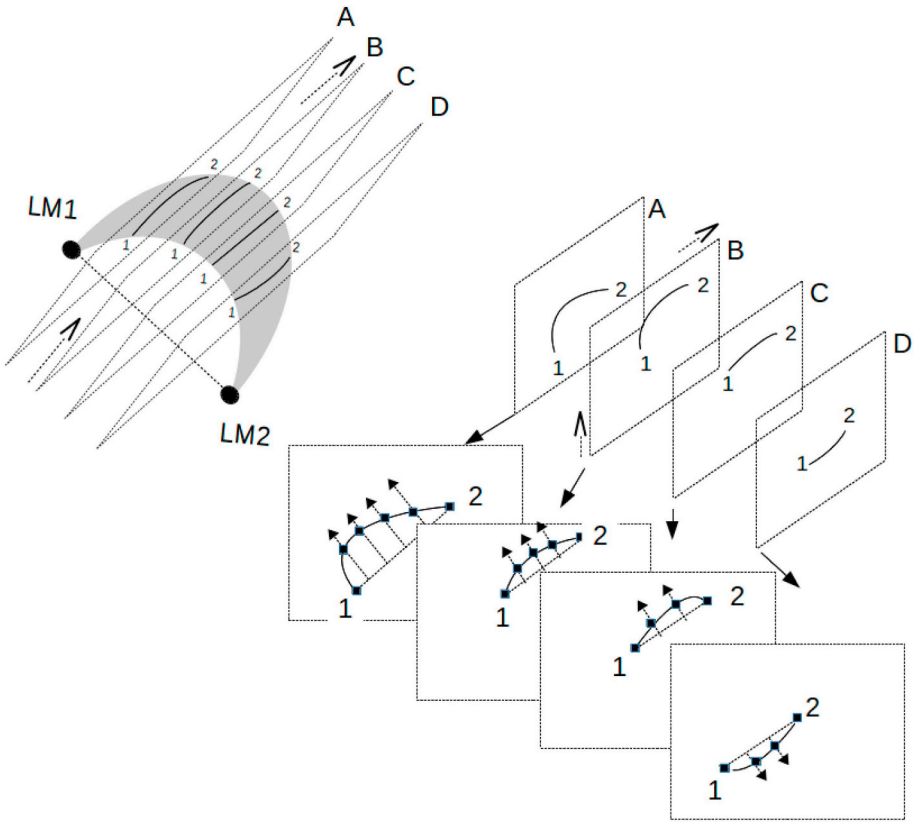


Fig. 12. A possible template for 3D surface. The grey shape (left part) is figuring a 3D form in contact with two landmarks (LM1, LM2). A succession of parallel plans (A, B, C, D) may be raised from a plan containing the connecting line between LM1 and LM2. Each plan captures a part of the grey form as a curve. Each curve can then be sampled by constructing a template between its beginning and its end. In this case, the flanking landmarks ("1" and "2") are of type III.

5. Conclusions

Currently, the methods specifically developed for semilandmarks alignment do not converge to exactly the same results, they make use of a minimization criterion, either minimum bending energy (E) or minimum Procrustes distance (D). Another criterion than minimization, using "projection" to a reference feature (P), represents a third group of methods. Our method belongs to this third group, except that the "reference feature" is used for both digitization and alignment of semilandmarks. Our study showed the close parallelism of information provided by our method (T) and the main other ones (E, D, P). The interest of our method lies in its adaptation to semilandmarks and to landmarks III data, thus covering the study of both curves and special features of interest, even limited to one point. Its simplicity ensures a fast computation and an intuitive understanding. It is however limited to 2D simple curves between successive landmarks, and would benefit from a more versatile template construction, allowing more variety in the "template-dependent semilandmarks" acquisition.

Acknowledgment

I am sincerely grateful to Julien Claude for his kind help and advice, which greatly improved the quality of the manuscript. F. James Rohlf provided helpful and encouraging comments. Any errors that remain are the author's own responsibility. This work benefited from international collaboration through the ECLAT network, and received support from the SYNTHESYS project http://synthesys3.myspecies.info which is financed by the European Community Research Infrastructure Action under the FP7.

References

Bookstein, F.L., 1991. Morphometric tools for landmark data. In: Geometry and Biology. Cambridge University Press, NY.
Bookstein, F.L., 1997a. Landmark methods for forms without landmarks: localizing group differences in outline shape. Med. Image Anal. 1, 225–243.
Bookstein, F.L., 1997b. Shape and the information in medical images: a decade of the morphometric synthesis. Comput. Vis. Image Underst. 66, 97–118.
Claude, J., 2008. Morphometrics With R. ISBN 978-0-387-77789-4. Springer Science +Business Media LLC 330pp.
De Groote, I., Lockwood, C.A., Aiello, L.C., 2010. Technical note: a new method for measuring long bone curvature using 3D landmarks and semi-landmarks. Am. J. Phys. Anthropol. 141, 658–664.
Dujardin, J.P., Pham Thi, K., Truong Xuan, L., Panzera, F., Pita, S., Schofield, C.J., 2015a. Epidemiological status of kissing-bugs in South East Asia: a preliminary assessment. Acta Trop. https://doi.org/10.1016/j.actatropica.2015.06.022.
Dujardin, J.P., Truong Xuan, L., Pham Thi, K., Schofield, C.J., 2015b. The rising importance of *Triatoma rubrofasciata*. Memorias Oswaldo Cruz 110 (3), 319–323.
Garcia, C., 2012. A simple procedure for the comparison of covariance matrices. BMC Evol. Biol. 12, 222.
Gunz, P., Mitteroecker, P., 2012. Semilandmarks: a method for quantifying curves and surfaces. Hystrix. https://doi.org/10.4404/hystrix-24.1-6292.
Gunz, P., Mitteroeker, P., Bookstein, F.L., 2005. Semilandmarks in three dimensions. In: Slice, D.E. (Ed.), Modern Morphometrics in Physical Anthropology. Kluwer Academic/Plenum Publishers, New York.
Hwang, W.S., Weirauch, C., 2012. Evolutionary history of assassin bugs (Insecta: Hemiptera: Reduviidae): insights from divergence dating and ancestral state reconstruction. PLoS One 7 (9), e45523. https://doi.org/10.1371/journal.pone.0045523.
Hypsa, V., Tietz, D.F., Zrzavy, J., Rego, R.O.M., Galvão, C., Jurberg, J., 2002. Phylogeny and biogeography of Triatominae (Hemiptera : Reduviidae): molecular evidence of a New World origin of the Asiatic clade. Mol. Phylogenet. Evol. 23 (3), 447–457.
Kaba, D., Berté, D., Ta, B.T.D., Tellería, J., Solano, P., Dujardin, J.P., 2017. The wing venation patterns to identify single tsetse flies. IGE 47, 132–139.
Kuhl, F.P., Giardina, C.R., 1982. Elliptic fourier features of a closed contour. Comp. Graph. Image Process. 18, 236–258.
Lohmann, G.P., 1983. Eigenshape analysis of microfossils: a general morphometric procedure for describing changes in shape. Math. Geol. 15, 659–672.
MacLeod, N., 2013. Palaeomath 101: landmarks and semilandmarks: differences without meaning and meaning without difference. Palaeontol. Assoc. Newslett. 82, 32–43.
Mitteroecker, P., Gunz, P., 2009. Advances in geometric morphometrics. Evol. Biol. 36, 235–247.
Monnier, G., Mcnulty, K., 2010. Questioning the link between stone tool standardization and behavioral modernity: Analytical approaches to paleolithic technologies, pp. 61-81. In: Lycett, S., Chauhan, P. (Eds.), New Perspectives on Old Stones. Springer, New York, NY ISBN 9781441968609.
Monteiro, F.A., Wesson, D.M., Dotson, E.M., Schofield, C.J., Beard, C.B., 2000. Phylogeny and molecular taxonomy of the Rhodniini derived from mitochondrial and nuclear DNA sequences. Am. J. Trop. Med. Hyg. 62 (4), 460–465.
Perez, S.Y., Bernal, V., Gonzalez, P.N., 2006. Differences between sliding semi-landmark

methods in geometric morphometrics, with an application to human craniofacial and dental variation. J. Anat. 208, 769–784.
Reddy, D.P., Harvati, K., Kim, J., 2005. An alternative approach to space curve analysis using the example of the Neanderthal occipital bun. In: Modern Morphometrics in Physical Anthropology, pp. 99–115 ed. dennis e. slice. pp383, ISBN: 978-1-4419-3467-3.
Rohlf, F.J., 1990. Rotational fit (Procrustes) methods. In: Rohlf, F., Bookstein, F. (Eds.), Proceedings of the Michigan Morphometrics Workshop. Special Publication Number 2. The University of Michigan Museums, Ann Arbor, MI, pp. 227–236 pp380.
Rohlf, F.J., Archie, J.W., 1984. A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera:Culicidae). Syst. Zool. 33, 302–317.
Rohlf, F.J., Slice, D.E., 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst. Zool. 39, 40–59.
Sampson, P.D., Bookstein, F.L., Sheehan, F.H., Bolson, E.L., Marcus, L.F., Corti, M., Loy, A., et al., 1996. Eigenshape analysis of left ventricular outlines from contrast ventriculograms. In: Advances in Morphometrics. Plenum, pp. 211233.
Santillán-Guayasamín, S., Villacís, A.G., Grijalva, M.J., Dujardin, J., 2017. The modern morphometric approach to identify eggs of Triatominae. Parasit. Vectors 10, 55. https://doi.org/10.1186/s13071-017-1982-2.
Sheets, H.D., Covino, K.M., Panasiewicz, J.M., Morris, S.M., 2006. Comparison of geometric morphometric outline methods in the discrimination of age-related differences in feather shape. Front. Zool. 3 (15).
Stephens, C.R., Juliano, S.A., 2012. Wing shape as an indicator of larval rearing conditions for *Aedes albopictus* and *Ae. aegypti* (Diptera: Culicidae). J. Med. Entomol. 49 (4), 927938.
Webster, M., Sheets, H.D., 2010. A practical introduction to landmark-based geometric morphometrics. In: Alroy, John, Hunts, Gene (Eds.), Quantitative Methods in Paleobiology. 16. Paleontological Society Short Course, The paleontological Society Papers, pp. 163–188.
Yee, W.L., Sheets, H.D., Chapman, P.S., 2011. Analysis of surstylus and aculeus shape and size using geometric morphometrics to discriminate *Rhagoletis pomonella* and *Rhagoletis zephyria* (Diptera: Tephritidae). Ann. Entomol. Soc. Am. 104 (2), 105–114. https://doi.org/10.1603/AN10029.
Zelditch, M.L., Swiderski, D.L., Sheets, H.D., Fink, W.L., 2004. Geometric Morphometrics for Biologists: A Primer. Elsevier, Academic Press, New-York.

Jean-Pierre Dujardin
UMR 2724177, INTERTRYP, CIRAD-IRD, Baillarguet, France
E-mail address: dujjepi@gmail.com.