Infection, Genetics and Evolution 70 (2019) 189–196

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Geometric morphometrics in the cloud

ARTICLE INFO

Keywords:
Geometric morphometrics
Online software

ABSTRACT

XYOM, for XY Online Morphometrics, is an online implementation of the geometric morphometric (GM) approach. It is a platform-independent product, and is presented here as an optional alternative software to client side morphometrics software. From the point of view of the user, the interesting features of a web application are: no download, no installation, no configuration, and automatic updating. Because XYOM is accessible through a standard web interface, it is expected to allow an easier and faster learning process. Additional benefits are that users will have their own highly secured cloud storage, with a 24/7 access from any device, allowing users to share their data, export/download them into their device. Ideally, there would be a permanent, anywhere anytime access on any device (computer, tablet, smartphone, etc). Using modern web browsers, XYOM allows online 2D images digitization of either landmarks, semilandmarks or pseudolandmarks (contours), and develops corresponding statistical analyses. In its present configuration, XYOM is dedicated to the identification and characterization of organismal forms.

1. Introduction

Geometric morphometric (GM) tools require specialized algorithms realizing tasks as diverse as image digitization, Procrustes superposition, semilandmarks sliding, Elliptic Fourier analysis, uni- and multivariate analyses, etc. Most of these features are provided by freely available software (see for instance http://life.bio.sunysb.edu/morph/, maintained by J. Rohlf).

These softwares are high quality products, frequently developed by the best specialists in the field. However, they are client side softwares, i.e., desktop/traditional softwares, as opposed to online softwares, i.e., web application/cloud softwares. The user has to download and install the software on his/her computer hard disk, with the corresponding tasks to configure, maintain and update the successive versions.

We present an alternative product, XYOM, as a solution to the problems generally linked to client side software. It circumvents most of the desktop software inconveniences: no download, no installation, no configuration are necessary. The user benefits from an automatic updating, and needs not worry about platforms compatibilities. At this stage of development, the XYOM software implements the main analyses required to apply modern morphometrics in medical entomology, i.e., characterization and identification (Dujardin, 2008).

2. Materials & methods

Presently, the XYOM interface follows classical features commonly found on the web. The Web application is running on Google Cloud Platforms (https://cloud.google.com/). The user can authenticate with his free Google account allowing a secure session during his work.

2.1. Web application architecture and stack

The application uses recent technologies (Table 1). It interacts with various backend services such as Google APIs (Google Oauth2, Google Drive, Google Datastore) and a web server running under NodeJS and Express (Fig. 1).

2.2. Digitization

XYOM allows the user to digitize and process landmarks, semilandmarks and pseudolandmarks (contours). The first image must be scaled. This is true even when all images have been collected in exactly the same way. In case some images have been obtained under a different optical configuration, the user is invited to scale each image.

Digitization data are stored in the user's Google Drive. Each image and its landmarks can be retrieved later for further examination. Raw landmarks, semilandmarks and pseudolandmarks, as well as size and shape variables computed from them, are stored in the user's Google Drive or in the computer hard disk as text, CSV or TPS formats. They may be exported in formats allowing to use them with other statistical software.

2.3. Coordinates files format

Coordinates of landmarks are stored in files as data matrices for subsequent statistical analyses. The typical format of such file is a data matrix starting at the second row of the file, as for the CLIC package (http://www.xyom-clic.eu). The first row allows any free comment describing the data and their possible subdivision. The subdivision is a sequence of numbers corresponding to the size of each con-

https://doi.org/10.1016/j.meegid.2019.02.018
Received 16 December 2018; Received in revised form 14 February 2019; Accepted 16 February 2019
Available online 19 February 2019

Table 1
On the client side we use Angular, a framework develop by Google (https://angular.io/). On the backend side, we use NodeJS (https://nodejs.org/en/) and the framework Express (https://expressjs.com/). We also use Google APIs such as Google Oauth2 (https://developers.google.com/identity/protocols/OAuth2) and Google Drive (https://developers.google.com/drive/). For the database we use Google Datastore (https://cloud.google.com/datastore/docs/concepts/overview).

Backend	NodeJS express
Frontend	Angular JQuery
Database	Googe datastore
APIs	Googe Drive, Google OAuth2
Libraries	Plotly, Numeric javascript, Multilayer Perceptron

stitutive group in the file. There is no column(s) describing each individual (row), so that if the data contain individuals belonging to different species or geographic areas, they are supposed to be assembled as successive groups. Although not necessarily required for some analyses to run, the subdivision information is very useful to interpret the graphical output.

2.4. Morphometrics analyses

Most statistical algorithms were adapted from the Tcl source of the CLIC package (http://www.xyom-clic.eu), replacing the Hume's linear algebra Tcl library (http://www.hume.com/la/la.html) by the javascript one: numerics.js (http://www.numericjs.com/). A machine learning module is proposed, based on a multilayer perceptron library available at https://www.npmjs.com/package/mlp. The different analyses currently developed by XYOM are listed in Table 2.

The graphical illustrations presented here were derived from data published in (Santillán-Guayasamín et al., 2017). They are available as example data from the "help" button of the XYOM menu (https://xyom-clic.eu/xyom-analyzes/).

2.4.1. Size

In case of landmarks (and semilandmarks) digitization, size is computed as centroid size (Bookstein, 1991). For outlines defined by pseudolandmarks, the size internally used for normalization is half the major axis of the first ellipse (first harmonic), but for ease of visual representation, the output files of size contain either the perimeter or the square-root area of the contour. Statistical comparisons of size among groups are performed by a oneway ANOVA, and illustrated by quantile boxes. Classification based on global size variation is possible using a maximum likelihood based validated reclassification approach (Dujardin et al., 2017), as well as a multilayer perceptron (https://www.npmjs.com/package/mlp).

2.4.2. Shape

For anatomical landmarks, a simple Procrustes superimposition to the consensus configuration is computed according to the Generalized Procrustes Analysis (Rohlf, 1990). If semilandmarks were digitized in addition to landmarks, they are slided according to the perpendicular projection to the template used to collect them (see Dujardin, this issue). The resulting aligned specimens are projected orthogonally onto the Euclidean space tangent to the consensus form, they are called here ORP (orthogonal projections).

For outlines (pseudolandmarks), shape variables are represented by the normalized elliptic Fourier coefficients (NEF) (Kuhl and Giardina, 1982; Lestrel, 1997).

Shapes variables, either ORP or NEF, are the input variables for the principal component analysis (PCA). Actually, the resulting principal components (PC) are considered as final shape variables for subsequent analyses such as the discriminant analysis (DA), the maximum likelihood-based and the Mahalanobis distances-based validated reclassification tests (CCCMLi and CCCMaha, respectively, see Table 2). The use of PC as final shape variables allows to reduce multidimensionality where necessary by adapting the number of variables to the sample sizes, as generally the number of specimens of the smallest group less one (Dujardin et al., 2014).

A separate section of XYOM is devoted to the bilateral symmetry analyses, restricted to objects having homologous left and right landmarks and based on the Procrustes superimposition (Klingenberg and McIntyre, 1998).

2.5. Graphical outputs

Graphics are built using the Plotly library: it provides open source tools for composing, editing, and sharing interactive data visualization via the Web (https://plot.ly/). The user may easily customize the graphics, and download them if needed.

Size variation is illustrated by quantile boxes corresponding to the subdivision given by the user (Fig. 2).

More graphical outputs are presented for shape. First the various steps of the general Procrustes analysis (GPA) are illustrated by corresponding graphics: translation, scaling and rotation. The final graphic represents the superposed mean objects: this figure allows to quickly visualize the areas showing the most important changes between groups (Fig. 3).

An example of the corresponding graphic for the outlines analyses is illustrated (Fig. 4).

For the PCA the plot of the first two PC is produced. According to the subdivision which was entered, a different color may be given a posteriori to each group, as well as convex hulls. For the DA, for which the subdivision information is mandatory, the same kind of graphical output is produced (Fig. 5). Using both the keyboard and the mouse, individual identification is possible in the plot.

To illustrate allometric content of shape, or the contribution of size variation to the shape-based discrimination among groups, a plot is shown of the first PC on size (or any other PC), with the expected shape position after regression of shape on size (Fig. 6).

A simple dendrogram may be obtained using as input the Euclidean distances between either PC of DF (Mahalanobis distances). The hierarchical clustering analysis performed by XYOM currently uses the single linkage algorithm (Fig. 7). The report of the analysis includes the Newick format of the tree. A bootstrap of the tree may be realized according to the procedure recommended by (Couette et al., 2005).

3. Results

After logging in with a Google account, a dashboard is presented where the user can create a new project, manage his projects, select/explore data in Google Drive or access some video tutorials (later, a community via a forum will be set up).

3.1. Digitization

XYOM currently allows landmarks, semilandmarks, and pseudolandmarks digitization.

Scaling has been made mandatory. Thus, images submitted to digitization should have a size scale apparent. The user is invited to scale each image. Only the first image is and must be scaled in case all images have been collected in exactly the same way.

Before digitizing an image, it can be zoomed, flipped horizontally and/or vertically.

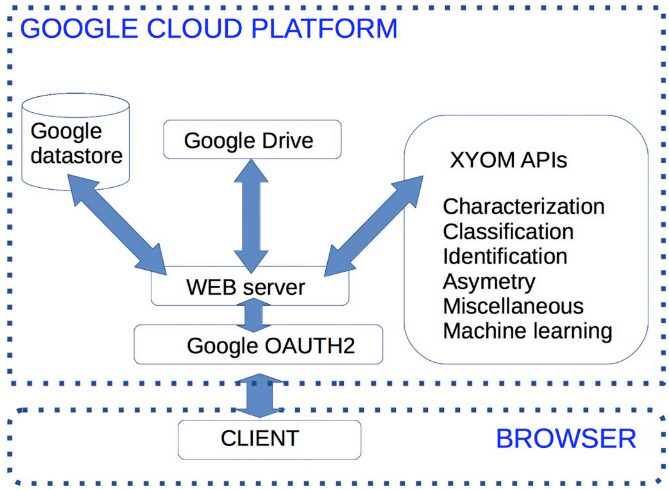

Fig. 1. Xyom architecture. On the client side the application is managed by Angular on modern browsers. All the requests must be validated by Google Oauth2 that states whether the user is connected to his/her Google account. The requests (authentication, analyses, and others) are sent to the web server on the Google Cloud Platform servers. They are then handled accordingly and rerouted to other web services such as Google Datastore for storing users profile, Google Drive to store users images and XYOM API (Application Programming Interface) to handle the geometric morphometrics analyses requests.

Table 2
Main analyses.

Input data	Analyses	Graphics
	Characterization	
Landmarks (LM)	GPA	Yes
LM and semiLM	GPA + TD	Yes
PseudoLM	EFA	Yes
	Classification	
Principal components	DA	Plot DF1/DF2
	CCC MLi, Maha, HC bootstrap	No
Distances	HC	Dendrogram
Global size	Oneway ANOVA	Quantile boxes
Linear measurements	Oneway ANOVA	No
Other	PCA	Plot PC1/PC2
	Identification	
Raw LM	GPA, then MLi, Mahalanobis	No
Raw pseudoLM	EFA, then MLi, Mahalanobis	No
Variables	ANN (MLP)	No
Global size	MLi	No
Linear distances	MLi, Mahalanobis	No
	Symmetry	
Raw LM	GPA	No
	FA size, shape, individual FA scores	No
	Miscellaneous	
Raw LM	MD after GPA	No
Raw pseudoLM	MD after EFA	No
Shape	Shape Escoufies coefficient	No
Size	Shape Linear regression analysis	Scatter plot
ASCII files	Extracting rows / columns	No
ASCII files	Concatenation by rows/columns	No
ASCII files	Converting to/from TPS format	No
ASCII files	Deleting rows	No

ANN (MLP), artificial neural network (multilayer perceptron); ANOVA, oneway analysis of variance; CCC Validated classification, based on Mahalanobis distances (Maha) or Maximum Likelihood (MLi); DA, Discriminant Analysis (F, factor); EFA, Elliptic Fourier Analysis; FA, Fluctuating asymmetry; GPA, Generalized Procrustes Analysis; HC, Hierarchical clustering; MD, metric disparity; PCA, Principal Component (PC) Analysis; TD, Template dependent semilandmarks analysis.

Each digitized image is saved so that the user can check previously digitized images and see the landmarks that had been digitized on it.

3.2. Analyses

Saved data on the cloud or on the computer hard disk may be retrieved for further analyses. The analyzes are mainly oriented to species or population characterization and identification. They are currently subdivided into five items: characterization, classification, identification, symmetry, and miscellaneous. There is an additional section devoted to the machine learning tools, which is (conceptually) connected to the classification and identification sections.

3.2.1. Characterization

This topic refers to the generation of size and shape variables from landmarks or pseudolandmarks (outlines) coordinates. The available analyses are the Generalized Procrustes Analysis (GPA) (Rohlf, 1990) and the Elliptic Fourier Analysis (EFA) (Kuhl and Giardina, 1982; Lestrel, 1997). The user selects the input file from either the computer hard disk or from their Google Drive. An additional input is required: the subdivision. The input file is frequently composed of various groups, the members of which are sequentially listed. If there is no known subdivision, or if there is only a single group, the total of rows must be entered instead.

In this section, XYOM anticipates the need of the user by proposing two classification analyses (see next paragraph): a principal component analysis (PCA) and a discriminant analysis (DA). If these analyses are selected in addition to the GPA, the PCA automatically uses as input the Procrustes residuals, i.e., the rotated individuals after their orthogonal projection to the Kendall space (ORP). When it is in addition to the EFA, the PCA automatically uses as input the NEF. The latter are saved restricting their number to the quantity providing a power of 99%.

3.2.2. Classification

Under this term XYOM gathers analyzes trying to present the structure possibly present in known data after their characterization, as well as the relationship they could draw between individuals and groups.

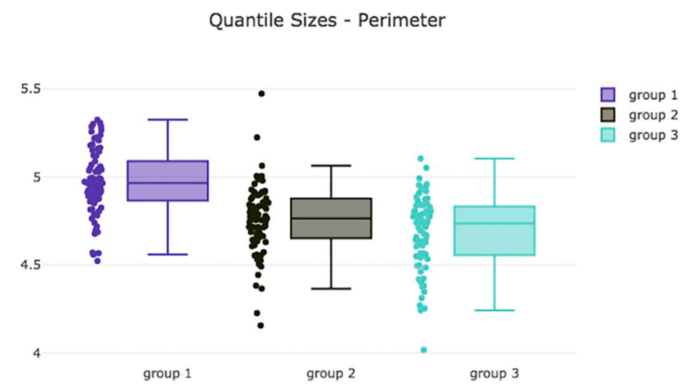

Fig. 2. Size quantiles. Quantiles of egg perimeter variation among two *Panstrongylus* species (*P. chinai* and *P. howardi*, as groups 1 and 2) and *Triatoma carrioni* as the third group (data from Santillán-Guayasamín et al. (Santillán-Guayasamín et al., 2017)).

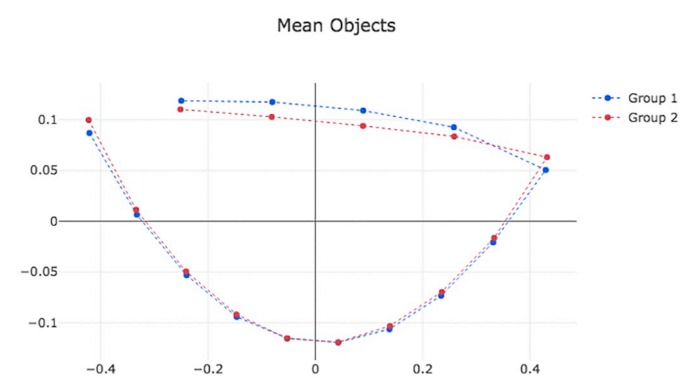

Fig. 3. Average landmarks and semilandmarks positions partially defining the shape of egg operculum for two genera of Triatominae. Group 1 *Panstrongylus chinai*. Group 2 *Triatoma carrioni* (data from Santillán-Guayasamín et al. (Santillán-Guayasamín et al., 2017)).

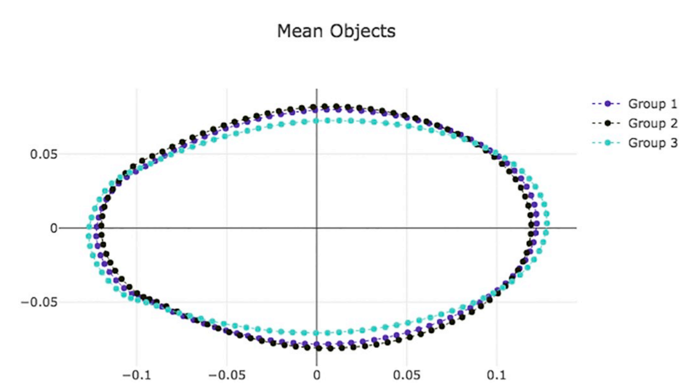

Fig. 4. Mean egg contours of three species: two *Panstrongylus* species (*P. chinai* and *P. howardi*, as groups 1 and 2) and *Triatoma carrioni* as the third group (data from Santillán-Guayasamín et al. (Santillán-Guayasamín et al., 2017)).

Visualizing the factor maps of the two first multivariate factors (Fig. 5) of either a principal component analysis or a discriminant analysis, helps to classify the groups according to their morphological affinities (Albrecht, 1980). A more complete illustration, using the information provided by all the multivariate factors, is given by building a classification tree (Fig. 7).

To estimate the potential of the data to identify external, unknown individuals (see next paragraph), a validated classification (also called cross-checked classification, or leave-one-out classification) is one of the most common approach (Manly, 1986). XYOM provides the possibility to base the cross-checked classification (CCC, Table 2) upon Mahalanobis distances and upon the maximum likelihood method (CCCMaha and CCCMLi, respectively, see Table 2).

3.2.3. Identification

Any tentative identification of unidentified or doubtful specimens requires their comparison with reference ones. Generally, the tentative identity of an unknown specimen is given by selecting the reference group closest to it.

XYOM proposes three statistical techniques using unknown and reference specimens; the Mahalanobis distance method, the maximum likelihood method (Dujardin et al., 2017; Polly et al., 2004) and the artificial neural network (Fedor et al., 2008; Lorenz et al., 2015). For the two first methods, the user is invited to separately upload the file with unknown data and the file with reference data. If input data are raw coordinates of either landmarks or pseudolandmarks, XYOM internally computes the corresponding shape variables (ORP, or NEF). For

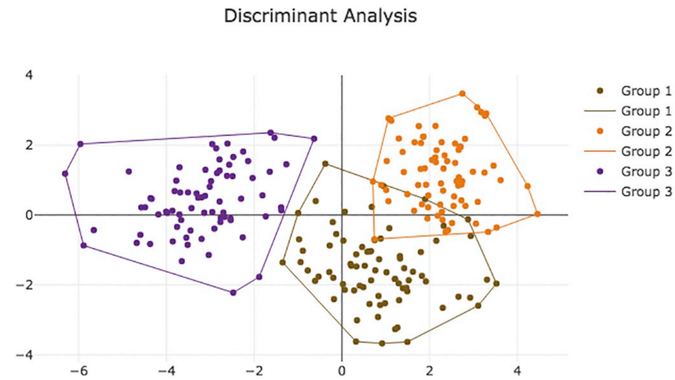

Fig. 5. Factor map of the discriminant factors describing shape divergence among three species: *Panstrongylus chinai* as group 1, *P. howardi*, as group 2 and *Triatoma carrioni* as the third group (data from Santillán-Guayasamín et al. (Santillán-Guayasamín et al., 2017)).

raw coordinates of landmarks, shape variables are computed on concatenated unknown and reference data. The subdivision which must be given is of course the one of the reference file. For the ANN method, the kind of variables corresponding to the unknown data must be of the same nature as the variables used for training the reference data. The training, after configuration of various parameters (Fig. 8), provides a weight file which is the one requested for identifying unknown individuals.

3.2.4. Symmetry

The analyses related to the symmetry of the bilateral structures of the body are currently restricted to anatomical landmark data. The user is invited to enter four files respecting the following order: the raw coordinates of one side, first digitization, the raw coordinates of the same side, second digitization, the raw coordinates of the other side, first digitization, and finally the raw coordinates of the other side, second digitization.

XYOM delivers a short report about size and shape symmetry, with the possibility to save individual fluctuating asymmetry (FA) scores for size and shape (Klingenberg and McIntyre, 1998).

3.2.5. Miscellaneous

This section gathers various analyses specific to modern morphometrics as well as some utilities for files management.

3.2.5.1. Metric disparity. Metric disparity refers to the variance of shape, it is computed as the trace of the variance-covariance matrix of shape variables. However, input data must be raw coordinates (landmarks, or pseudolandmarks): XYOM internally computes first the shape variables (ORP, NEF), then their variance-covariance matrix.

3.2.5.2. Covariation of shape between parts of the same body. In this analysis, the input variables are two distinct files containing shape variables of two parts of the same body. The Escoufiers coefficient (Escoufier, 1973) is computed; this statistics is analogous to a correlation coefficient between two sets of shape variables. It is the percentage of covariance of one set predicted by the other set (Claude, 2008).

3.2.5.3. Comparison of two variance-covariance matrices. The input variables are two distinct files containing shape variables, they must have the same number of columns. S1, S2 and S3 statistics are 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 (Garcia, 2012). Statistical significance is computed after 1000 bootstraps.

3.2.5.4. Allometry. Here the user is invited to upload a first file containing size, and a second file containing shape. The linear regression coefficient is computed with size as independent variable, after the user specify the shape component which will be the dependent variable. Both size and shape data must refer to exactly the same individuals in exactly the same order.

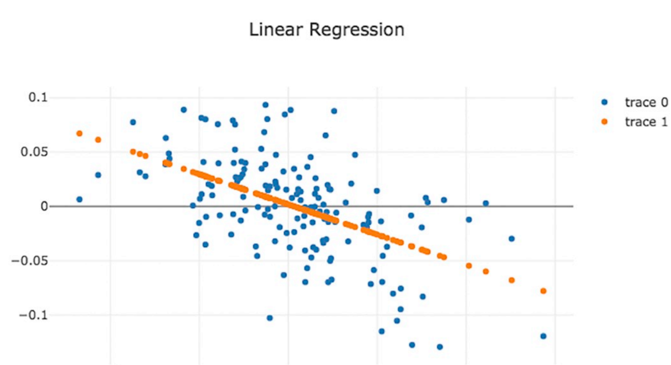

Fig. 6. Allometry. Linear regression prediction (orange dots line) of the first principal component (Y axis, derived from the operculum shape of *Panstrongylus chinai* and *T. carrioni*), according to the centroid size variation. Data are from Santillán-Guayasamín et al. (Santillán-Guayasamín et al., 2017).

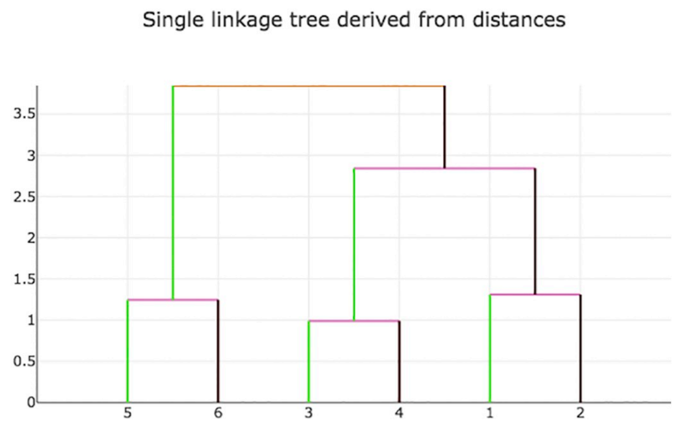

Fig. 7. Hierarchical clustering tree based on egg contour variation among three species of Triatominae. The dendrogram depicts morphological similarities between two subsamples of *Panstrongylus chinai* (numbers 1, 2), two subsamples of *P. howardi* (numbers 3, 4) and two subsamples of *Triatoma carrioni* (numbers 5, 6). The distances used here for the tree building are the Mahalanobis distances. Data from Santillán-Guayasamín et al. (Santillán-Guayasamín et al., 2017).

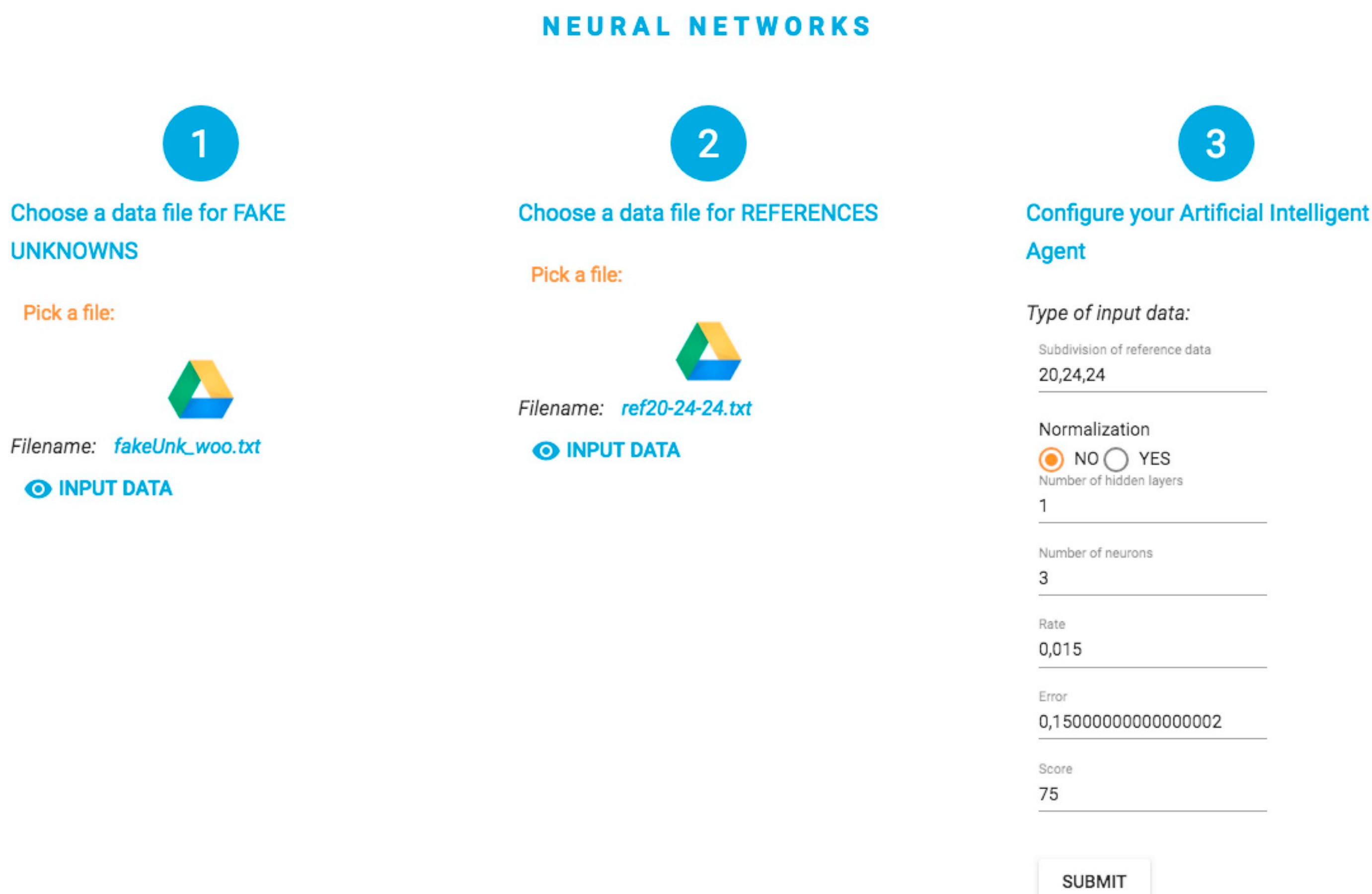

Fig. 8. Artificial neural network. The XYOM interface for using the artificial neural network dedicated to group identification. Three arguments must be entered as initial input data: (i) the unknown data, preferably fake unknown to allow the user to check for possible overfitting, but can be true ones also (ii) the reference data which will be subdivided into training and testing data, and (iii) the subdivision of the reference data. XYOM does not normalize the data, it is the first parameter that the user is allowed to modify (by default set to 0). Other parameters that the user can modify are: the number of hidden layers (default is 1), the number of neurons by layers (default is the output dimension, which is deduced from the subdivision argument), the learning rate, default is the product: " 0.015 x output dimension ", and finally the correctly classified objects required in the testing set (default is 75%). The user cannot suggest a number of neurons which would be different from one hidden layer to another. Training and testing sets of the reference data are automatically separated by the program in such a way each reference group is included in the training set.

3.2.5.5. File extraction. This utility allows to create a new file containing part of an input file, a submatrix, either a bloc of rows or a bloc of columns. After calling the input file, the user enters the first row or column, and the last one, to create the required submatrix. XYOM detects the situation where an outline file is submitted, allowing to extract rows from a table of data with unequal number of columns according to the individuals.

3.2.5.6. File concatenation. This utility allows to create a new file merging rows (if the two input files have the same number of columns) or columns (if the two input files have the same number of rows). XYOM detects the situation where two outline files are submitted, allowing their merging by rows in spite of presenting different columns. After merging (by rows), it makes the concatenated matrix a rectangular one by adding NaN symbols where necessary.

3.2.5.7. File conversion. The format currently used by XYOM is the one of CLIC (https://xyom-clic.eu), i.e., a data matrix after a very first row containing a free comment. This format can be converted to the TPS format, or a TPS file can be converted to the XYOM format. There is a special conversion needed by XYOM in case of entering a data matrix containing exclusively pseudolandmarks (outline data): the conversion adds three columns, the first one is the number of points (including the centroid point), the next two ones are the coordinates of the centroid point. This XYOM format allows an internal checking of outline data with error detection. The conversion makes the matrix rectangular by filling missing data with NaN symbol.

4. Discussion

Because of the mathematical complexity underlying the study of forms, the use of specialized statistical software to perform geometric morphometric analyses is mandatory. In the last two decades a number of excellent propositions has been made freely available by the best specialists in the field (see http://life.bio.sunysb.edu/morph/).

All of them need to be previously installed on the user's computer. This very first step may cause troubles due to platform compatibility (MS windows, OS X, Linux, Chrome OS, Unix, Android, iOS, etc.). Moreover, the installation process itself may be delicate. Users may be reluctant to spend time in updating or securing the current version of the software they succeeded once to install on their computer. The result is, they find themselves using outdated tools. Small bugs, security issues, new analyses, interface changes, and so on, are sometimes missed because of a non-updated software.

Frequently, client side software performing digitization and/or statistical analyses are developed in separate package. Frequently also, the landmark-based and the outline-based approaches are implemented in separate software. Thus, the user has to play with various software to perform morphometric studies, each software with its own specific interface.

A completely online solution could solve any of the problems mentioned above. There would be no platform compatibility problem, the user interface would be homogenized across different analyzes (digitization, outlines, anatomical landmarks), updates and security issues would be addressed without the need of user's intervention. In addition, a personal cloud space would help each user to organize their data, often split into many files on ones hard disk after statistical analyses.

Last but not least, an online solution would allow scientists to share their own images and analysis with other coworkers all around the world.

Nowadays we are more used to work and learn with web interfaces. As far as we know, two web initiatives related to modern morphometrics already took place: PhyloNimbus (https://www.phylonimbus.com/morphometrics/) and morpho-tools (http://morpho-tools.net/). PhyloNimbus is an online tool for collecting 2D and 3D morphometric data, while morpho-tools provides online tools to make morphometric analysis more widely accessible and quicker. XYOM comes into one bundle combining both needs: digitization (although 2D limited) and statistical analyses.

5. Conclusion

XYOM, as an online implementation of the geometric morphometric approach, is a platform-independent product, and represents an optional alternative software to client side morphometrics software. Because XYOM is aware of the Web standards, it gives the users a familiar experience and can be considered an easy tool for relatively complex analyses. The only requirement, also a limitation, is an internet connection, preferably with a modern browser.

6. Perspectives

New features will be added progressively thus enriching the XYOM application and environment. Soon to be published features are: semi automatic capture of outlines, machine learning capacity including SVM techniques, an addon to allow the user to work offline and synchronize its digitization data when back online. A long term ambition is to interactively connect XYOM with a bank of reference images. Community support would allow publication of images that have been validated by its users and, ideally, referenced in scientific publications. As already wished by various authors (Dujardin et al., 2010; Sonnenschein et al., 2015; Kitthawee and Dujardin, 2016; Lorenz et al., 2017; Kaba et al., 2017), such a bank would attempt to provide reliable references against which users could compare, classify and identify their own images.

Acknowledgments

To Philippe Solano (Intertryp, IRD) for his unconditional support, and to the following reviewers of the very first version of XYOM: Soledad Santillan-Guayasamín (Medical entomology, Universidad Catolica de Quito, Ecuador), Marcia Gumiel (INLASA, La Paz, Bolivia), Claude-Aimée Diaha-Kouame (IP, Abidjan, CI). The following institution provided direct partial financial support: IRD, UMR 177 IRD-CIRAD INTERTRYP, Campus international de Baillarguet, Montpellier, France Ref. 1D030000 IRD MONTPELLIER 17349000020. Indirect support was provided by the following two institutions: (i) the University of Mahidol, Department of Medical Entomology, Bangkok, Thailand, and (ii) ETNA, Ecole des Technologies Numériques Appliquées, & Rue Maurice Grandoin, 94200 Ivry-Sur-Seine.

References

Albrecht, G.H., 1980. Multivariate analysis of the study of form with special reference to canonical variate analysis. Am. Zool. 20, 679–693.
Bookstein, F.L., 1991. Morphometric tools for landmark data. Geometry and Biology. Cambridge University Press, NY.
Claude, J., 2008. Morphometrics with R, ISBN 978-0-387-77789-4. Springer Science + Business Media LLC 330pp.
Couette, S., Escarguel, G., Montuire, S., 2005. Constructing, bootstrapping, and comparing morphometric and phylogenetic trees: a case study on new world monkeys (Platyrrhini, Primates). J. Mammal. 85 (4), 773–781.
Dujardin, J.P., 2008. Morphometrics applied to medical entomology. Infect. Genet. Evol. 8, 875–890.
Dujardin, J.P., Kaba, D., Henry, A.B., 2010. The exchangeability of shape. BMC Res. Notes 3, 266.
Dujardin, J.P., Kaba, D., Solano, P., Dupraz, M., McCoy, K.D., Jaramillo-O, N., 2014. Outline-based morphometrics, an overlooked method in arthropod studies? IGE 28, 704–714.
Dujardin, J.P., Dujardin, S., Kaba, D., Santillán-Guayasamín, S., Villacís, A.G., Piyaselakul, S., Sumruayphol, S., Samung, Y., Morales-Vargas, R., 2017. The maximum likelihood identification method applied to insect morphometric data. Zool. Syst. 42 (1), 46–58.
Escoufier, Y., 1973. La dépendance de deux aléas vectoriels critères et visualisation, Revue de statistique appliquée. vol. 21. pp. 5–16.
Fedor, P., Malenovsky, I., Vanhara, J., Havel, J., 2008. Thrips (thysanoptera) identification using artificial neural networks. Bull. Entomol. Res. 98 (5), 437–447.
Garcia, C., 2012. A simple procedure for the comparison of covariance matrices. BMC Evol. Biol. 12, 222.
Kaba, D., Berte, D., Ta, B.T.D., Telleria, J., Solano, P., Dujardin, J.P., 2017. The wing

venation patterns to identify single tsetse flies. IGE 47, 132–139.
Kitthawee, S., Dujardin, J.P., 2016. The *Diachasmimorpha longicaudata* complex in Thailand discriminated by its wing venation. Zoomorphology 135. https://doi.org/10.1007/s00435-016-0307-x.
Klingenberg, C.P., McIntyre, G., 1998. Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. Evolution 52 (5), 1363–1375.
Kuhl, F.P., Giardina, C.R., 1982. Elliptic Fourier features of a closed contour. Comp. Graph. Image Process. 18, 236–258.
Lestrel, P.E., 1997. Introduction and overview of Fourier descriptors. In: Chap 2: Fourier Descriptors and their Applications in Biology. Cambridge University Press, Cambridge.
Lorenz, C., Ferraudo, A.S., Suesdek, L., 2015. Artificial neural network applied as a methodology of mosquito species identification. Acta Trop. 152, 165–169.
Lorenz, C., Almeida, F., Almeida-Lopes, F., Louise, C., Pereira, S.N., Petersen, V., Vidal, P.O., Virginio, F., Suesdek, L., 2017. Geometric morphometrics in mosquitoes: what has been measured? Infect. Genet. Evol. 54, 205–215.
Manly, B.F.J., 1986. Multivariate Statistical Methods. A Primer, Chapman and Hall, London, pp. 154.
Polly, P.D., Head, J.J., Ashraf, M., Elewa, T., 2004. Maximum-likelihood identification of fossils: Taxonomic identification of quaternary marmots (Rodentia, Mammalia) and identification of vertebral position in the pipesnake Cylindrophis (Serpentes, Reptilia). In: Morphometrics Applications in Biology and Paleontology, pp. 197–221.
Rohlf, F.J., 1990. Rotational fit (Procrustes) methods. In: Rohlf, F.J., Bookstein, F.L. (Eds.), Proceedings of the Michigan Morphometrics Workshop, ISBN 0-9628499-01. University of Michigan Museums, Ann Arbor, pp. 227–236.
Santillán-Guayasamín, S., Villacís, A.G., Grijalva, M.J., Dujardin, J.P., 2017. The modern morphometric approach to identify eggs of Triatominae. Parasit. Vectors 10, 55. https://doi.org/10.1186/s13071-017-1982-2.
Sonnenschein, A., VanderZee, D., Pitchers, W.R., Chari, S., Dworkin, I., 2015. An image database of *Drosophila melanogaster* wings for phenomic and biometric analysis. GigaScience 4, 25.

Sebastien Dujardin, Jean-Pierre Dujardin*
IRD, UMR INTERTRYP IRD, CIRAD, University of Montpellier, F-34398 Montpellier, France
E-mail address: jean-pierre@xyom.io (J.-P. Dujardin).

* Corresponding author.