Shobnom Ferdous and Jonathan W. Armbruster

Department of Biological Sciences
331 Funchess
Auburn University, AL 36849
USA

ferdosh@tigermail.auburn.edu
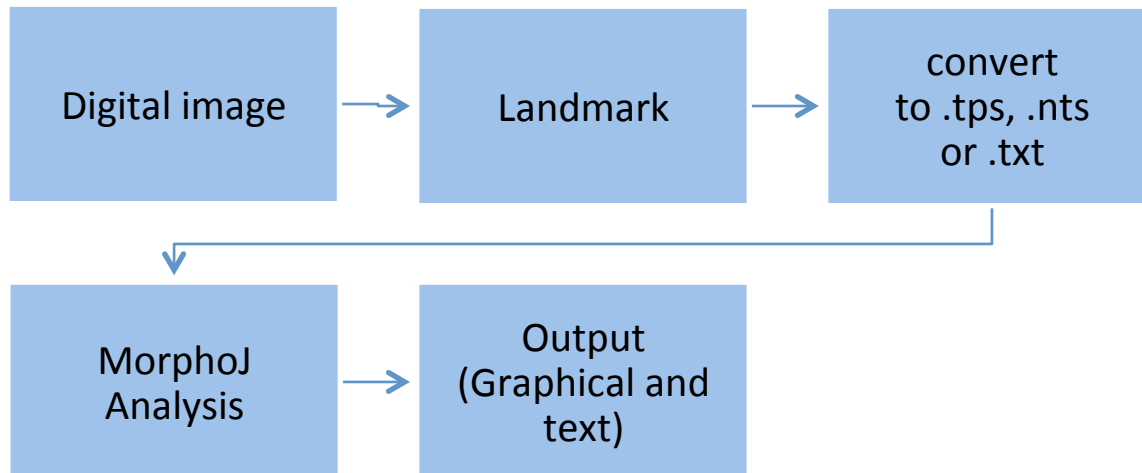
armbrjw@auburn.edu

Contents

1   Introduction

MorphoJ is an integrated software package to perform geometric morphometric analysis. The main purpose of MorphoJ is to provide a user-friendly environment to include the set of analysis commonly done in Geometric morphometrics. This is free software that can be run on Windows, Macintosh and the linux/unix operating system. MorphoJ can be downloaded from http://www.flywings.org.uk/MorphoJ_page.htm.

MorphoJ provides a single, integrated environment for geometric morphometrics so the user can concentrate on biological and statistical aspects of analysis (Klingenberg, 2011).

The complete sequence of preparation, analysis and presentation of result involves several steps. This is shown below:

```
┌──────────────┐     ┌──────────────┐     ┌──────────────┐
│              │ ──▶ │              │ ──▶ │   convert    │
│ Digital image│     │  Landmark    │     │ to .tps, .nts│
│              │     │              │     │   or .txt    │
└──────────────┘     └──────────────┘     └──────────────┘
        │
        ▼
┌──────────────┐     ┌──────────────┐
│              │     │   Output     │
│  MorphoJ     │ ──▶ │ (Graphical and│
│  Analysis    │     │    text)     │
└──────────────┘     └──────────────┘
```

Throughout this user guide we shall use two datasets. One of them is Mormyrid.tps. This dataset includes 25 individuals of mormyrid fishes that have a great amount of shape variation within the group. The other one is Lythrurus_Lateral.NTS, for a genus of North American minnows with comparatively less shape variation.

2   Preparation

The input to MorphoJ analysis is a data file that can be text (.txt), from Rohlf's statistical package (.tps), or NTSPC format (.nts). The final output of MorphoJ analysis is a set of statistical results and visualizations. But as we can see from the flow chart above, there are several steps involved before starting an analysis with MorphoJ.

In this user guide, we shall put the digital images in a dedicated folder and show each step with examples and explanations.

1. **Digital Image**: Take digital images and put all the images in a designated common folder. Download tpsutil and tpsDIG2 from The Morphometrics at SUNY Stony Brook website (http://life.bio.sunysb.edu/morph/).
2. **Digitize**: Build tps file

a. Open tpsutil.
b. From 'Select operation' choose 'build tpsfile from images'
c. Click input, find your images and click any image from this folder
d. Click output, Create a name for the tps file that will be created.
e. Click setup actions window. All images are checked, uncheck any that you don't want in the file 'Include'.
f. Make sure that 'Include path' is unchecked. This will include only the image name in the tps file. This is helpful if digitizing file will be saved in the same folder as images, and will result in a short name for the images.
g. Click 'Create' and the tps file is ready to be digitized.

3. **Digitize Images using tpsdig:**.
a. Choose: File>Input source>File and select the tps file you created.
b. First image should appear (may take a few moments). Explore all tabs, find out how to zoom in and out, Delete a specimen, Set scale, Mark landmarks or outlines, change size and color of symbols and labels.
For each images you need to
1. Set the scale:
a. Options>Image tools (or hammer icon). The last tab is 'Measure'
b. Click set scale, measure ruler, change Reference length and units if needed, click OK.
2. Digitize landmarks
a. Change size or color of dot: Image tools, the first tab is 'Color'.
b. Label landmarks: In options – Label landmarks (or # icon).Change size and color of number if desired in Tools>Color.
3. Save file. When you save the file choose: File>save data>Overwrite. It is important to choose 'Overwrite' instead of 'Append'. Use 'Append' if you want to combine two different datasets (this can also be done with tpsutil.
4. Examine tps file with text-reading software if desired.

A landmark file with 18 landmarks created using tpsdig will look like this:

LM=18
11.00000 173.00000
26.00000 131.00000
47.00000 122.00000
………………………
………………………
18 rows

IMAGE= Lythrurus_Lateral. 211L.JPG
ID=0
SCALE=0.17
 In tpsDig the digitized file could be saves as .tps or .NTS

The following link also gives a brief overview of how to place landmarks in digital images.
http://www.morpho-tools.net/softwareguide/GM%20guide%20v4%20OLs.pdf


3   MorphoJ User Interface

MorphoJ does not create a data file for geometric morphometrics. MorphoJ also can not analyze outline data. MorphoJ can import raw landmark data from tab- or comma-delimited text files or from other morphometric software: the TPS series (http://life.bio.sunysb.edu/morph/soft-tps.html), NTSYSpc (Rohlf 2008) and Morphologika (O'Higgins & Jones 1998). The MorphoJ web page has an excellent user guide which is available online at

 http://www.flywings.org.uk/MorphoJ_guide/frameset.htm?index.htm

What follows is a step by step simple user guide for the beginners. For advanced analysis the online MorphoJ user guide is a good reference.

3.1   MorphoJ Main Window

MorphoJ has a graphical user interface from which all analyses can be done. There can be multiple analyses for one or multiple data sets. The datasets and analyses are organized into a project.
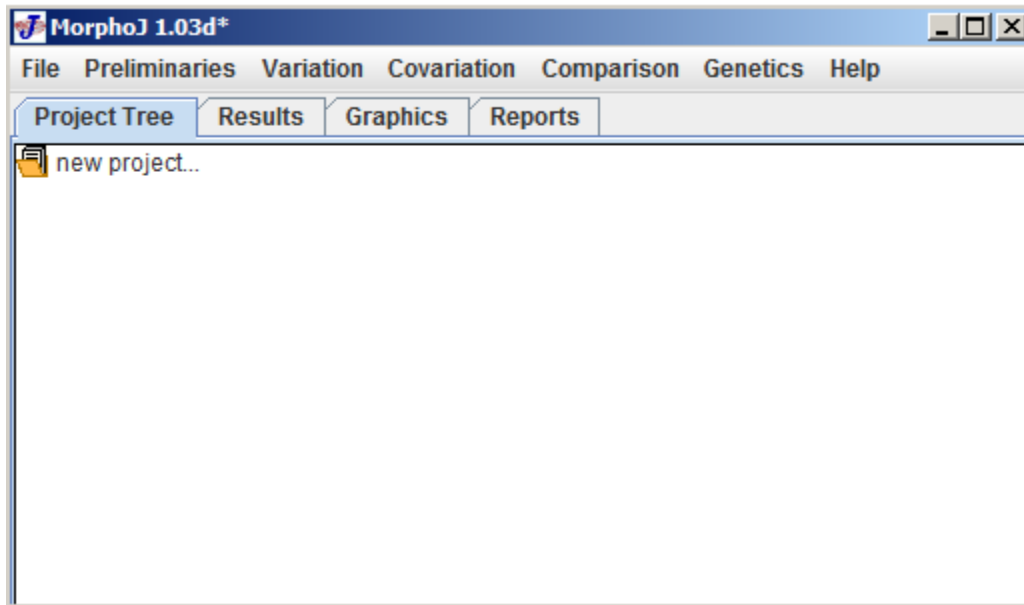

At the top of the MorphoJ window there is a menu bar with seven menus named as follows: **File**, **Preliminaries**, **Variation**, **Covariation**, **Comparison**, **Genetics** and **Help**. These are the menus used to complete an analysis using MorphoJ.

There are four tabs under the menu bar: **Project Tree**, **Results**, **Graphics** and **Reports**. Each of these menu bars allows you to view different aspects of the analyses.

Create new project: To create a new project click on the file menu and select **Create new project** which opens the following dialog box:
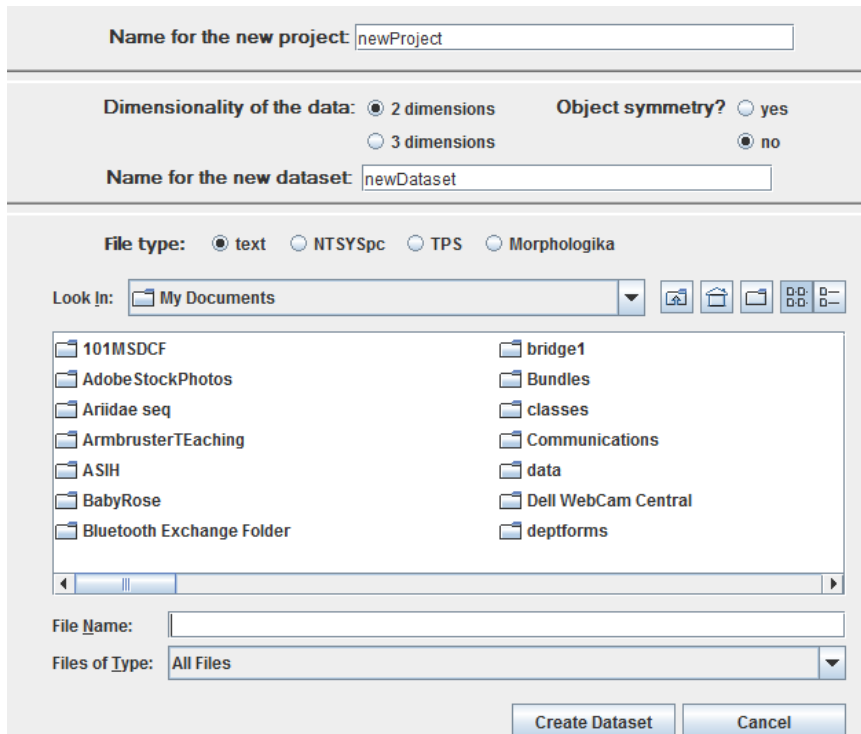
1. At the very top of the dialog box is the text box for naming the new project. Select a project name and choose the **Dimensionality of data**, if it is a 2D or 3D dataset (usually 2D). 'Object symmetry' is if all of your landmarks have bilateral symmetry, for a lateral view, this would be unchecked. The 2nd text box is **Name for the new dataset** and choose the **File type. File type** is what kind of data file is going to be imported in MorphoJ**.** For our analysis we are using Mormyrid.tps as data file and it is a 2D data set.
2. MorphoJ always opens in 'My Documents' in Windows and 'Documents' in MacOS. Navigate to the file. After selecting the correct data file select **Create Dataset,** which adds the data file in MorphoJ.
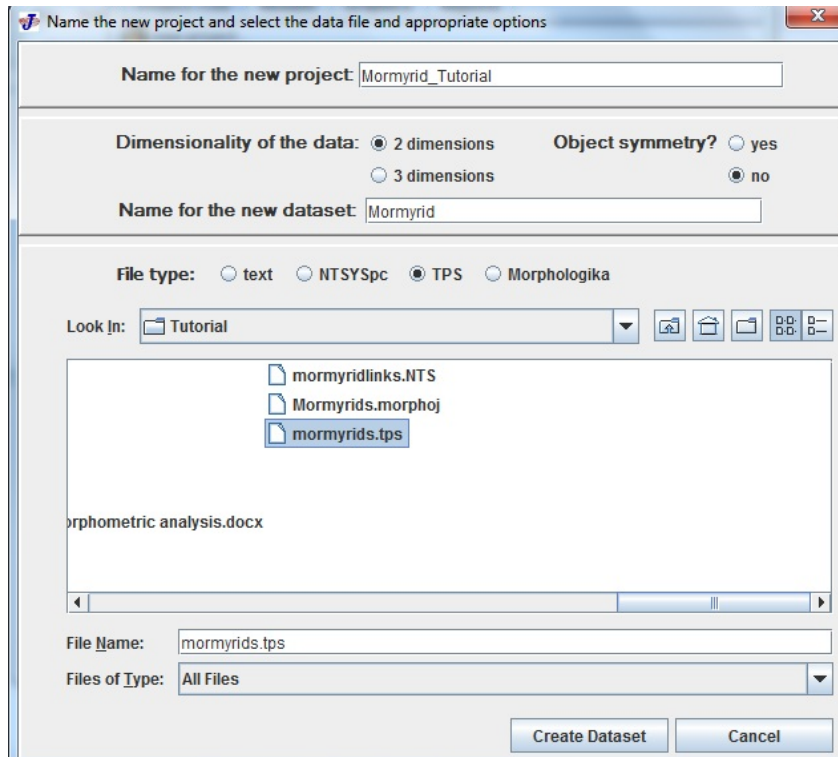
3. Once the data set has been created, click on the report tab which gives a message "the project **'Mormyrid_Tutorial'** has been loaded successfully".


The following screen shot shows the elements of the user interface and how it looks as we create the new project and data set.
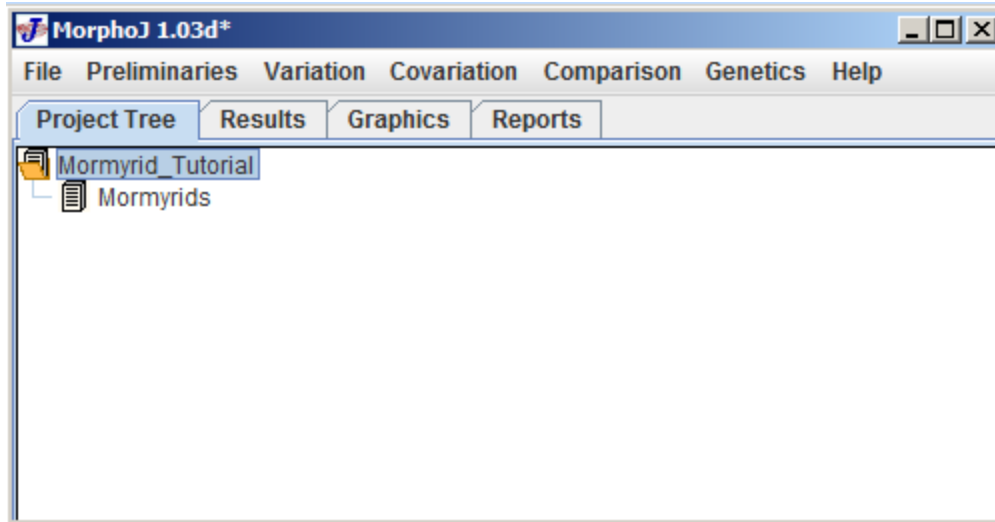
4. To create another data set under the same project select **File** menu and click on the **Create New Data** set and do the selection as appropriate with the data set.
5. To save a project select **File** menu and select **Save** or **Save Project As**. When you save a project it saves as given projectname.morphoJ. For our example here the project name is Mormyrid_Tutorial.morphoJ
6. If there is a project already saved it could be always opened by selecting – **File**>**Open project** or **Open Recent Project,** which will open the most recent projects.

4   Procrustes Fit

The first step in a morphometric analysis is to extract shape information by Procrustes superimposition on the data set. MorphoJ runs a full Procrustes fit and projects the data onto tangent space. To run a Procrustes fit select the **Mormyrid** data set under the **Project Tree** window, select **Preliminaries** menu and click on **New Procrustes Fit** which opens the following dialog box**:**
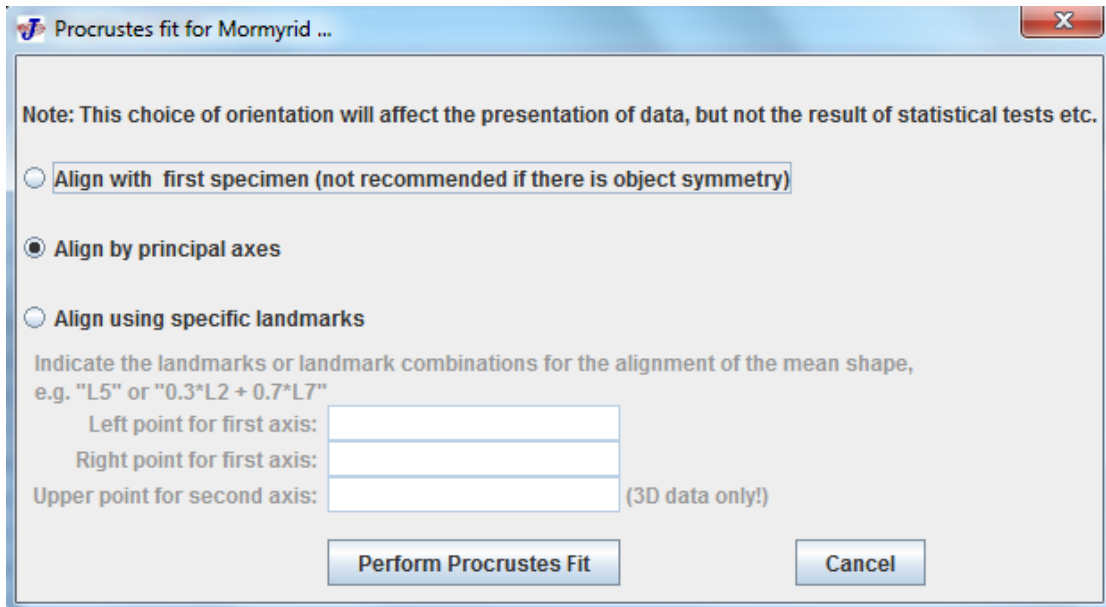
There are 3 choices in this dialog box and the default is aligning by principal axes. After selecting the default click Perform **Procrustes Fit** and it does give a consensus for all the landmarks. It also shows a graphical representation of the data clouds for each individual landmark.

After the Procrustes fit the graphical representation of the dataset looks like the following:
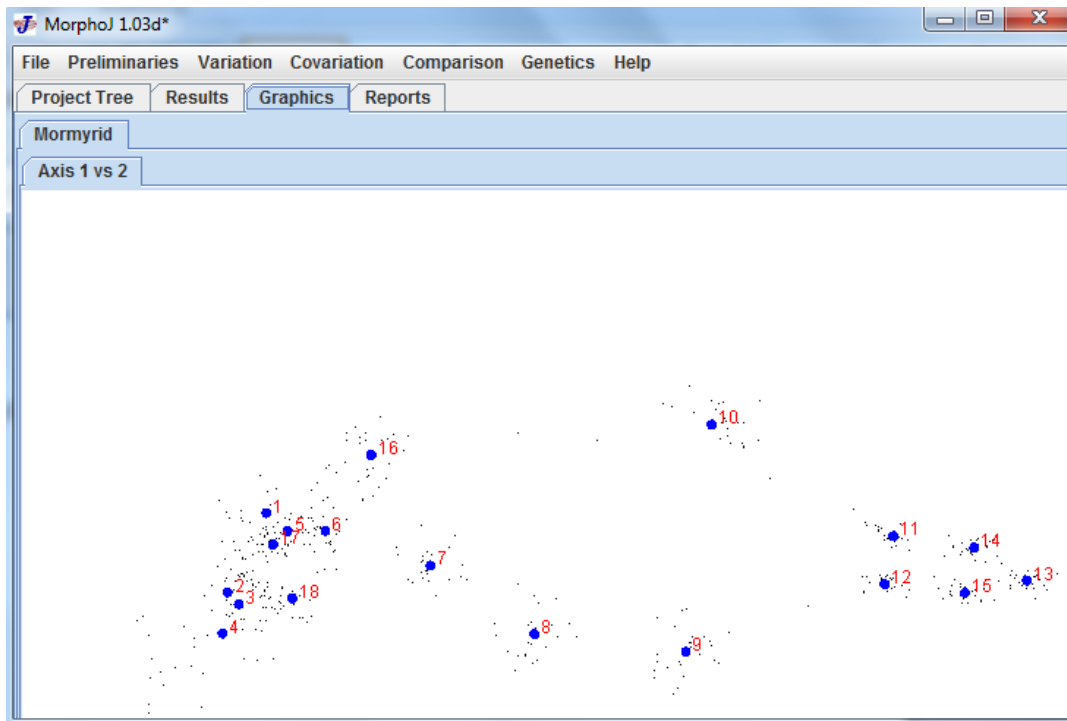
5   Data Outliers

It is always very important to check the data set to determine if there are any data outliers to determine if there are any problems in the data set. To check for outliers in a dataset, click the Mormyrid Project Tree, select the Mormyrids dataset and then select **Find Outliers** in the **Preliminaries** menu.
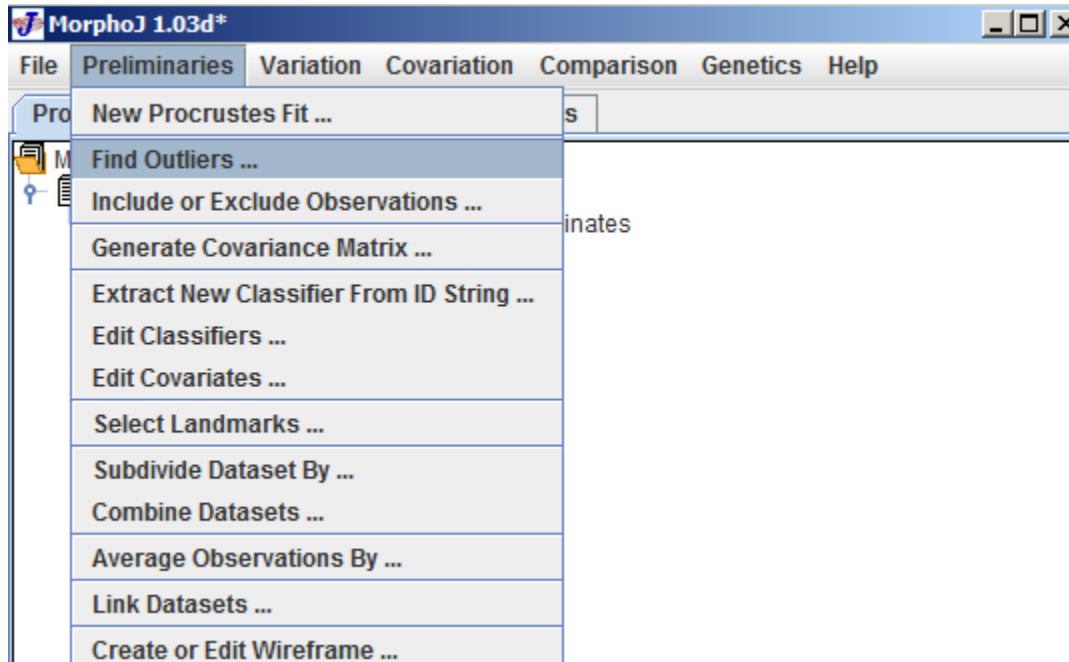

After clicking find outliers the output window shows an interface like the following in the graphics tab:
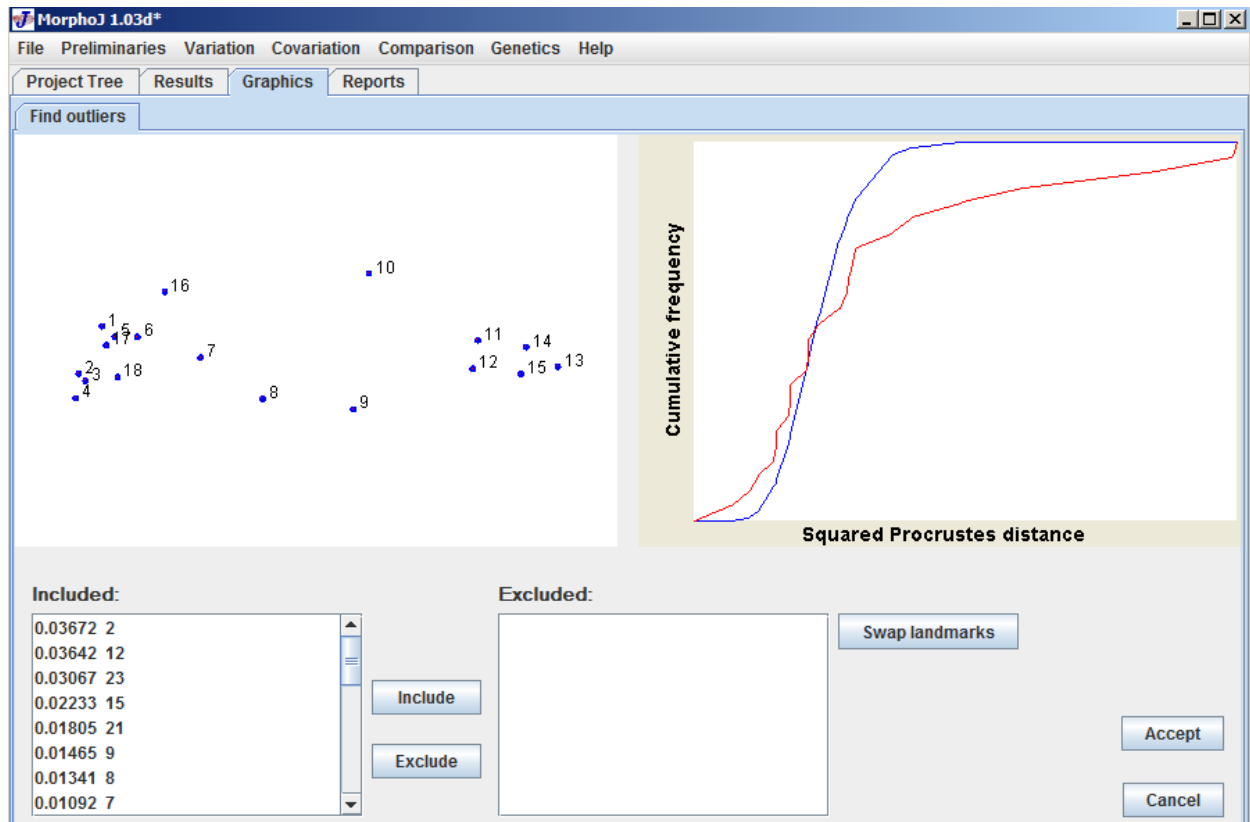
On the upper left side the diagram shows the average shape in blue dots and deviation from the average in red line. The panel on the lower left corner "**Included**" shows each individual specimen with their Id and the number of their diviation from the average.By selecting any of them shows which landmark deviated most. The upper-right panel shows how much each individual specimen vary from the average shape of the entire sample for Mormyrid data set. This diagram is just an approximate guide to the quality of the data. If any individual specimen is not desired in the data set we can exclude that specimen and get a new dataset for further analysis. To get the excluded individual someone can always get that from "**Excluded**" panel and select the specific Identifier need to be included.

6   Exclude Observations

Some time it is important to use some of the individuals (i.e. groups, male or female) of a data set for particular analysis. The easiest way to do this is going under project tree tab, select **preliminaries** menu and then select **include or exclude observations**. After selection the following dialog box appears:
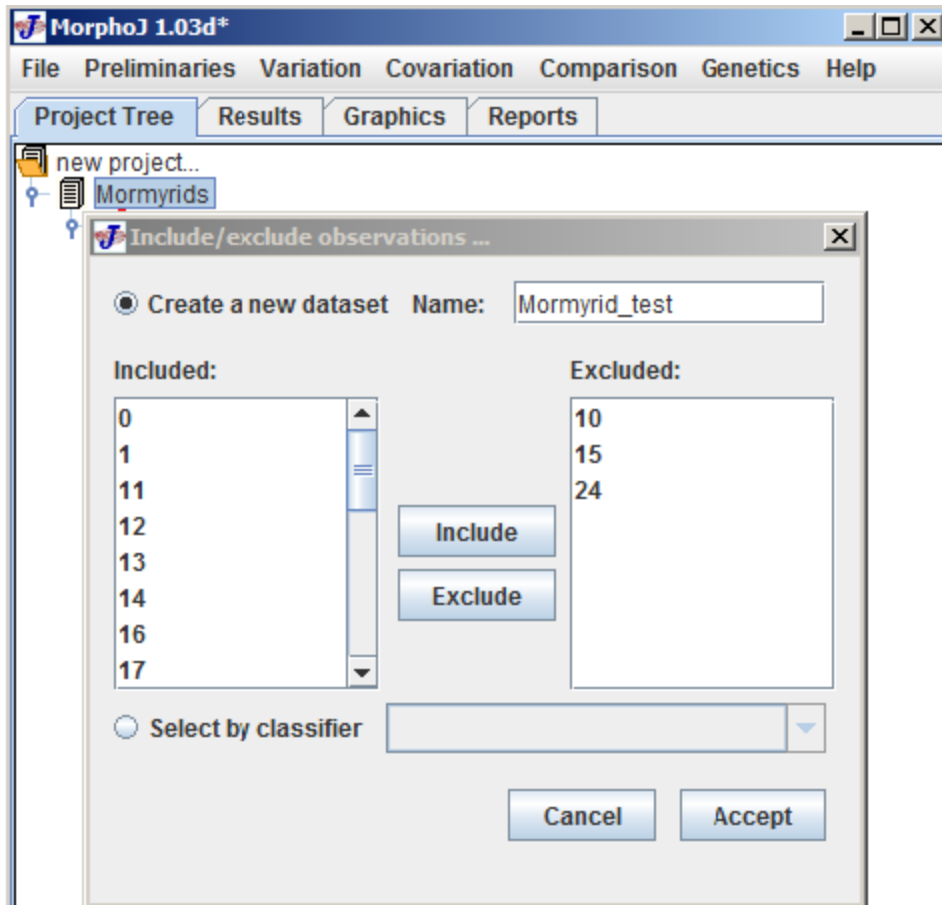
In this dialog box the top right corner has **Create a new dataset** button. If the included or excluded observations need to be a in a separate new dataset, this button should be selected which enables "Name" box to have a new name for the dataset. Here is an example which shows a separate dataset has been created named as "Mormyrid_test" and where 3 observations have been excluded. At the end click **Accept** tab and it creates a new dataset under the existing project.

The user interface looks like the following as we create a new dataset:
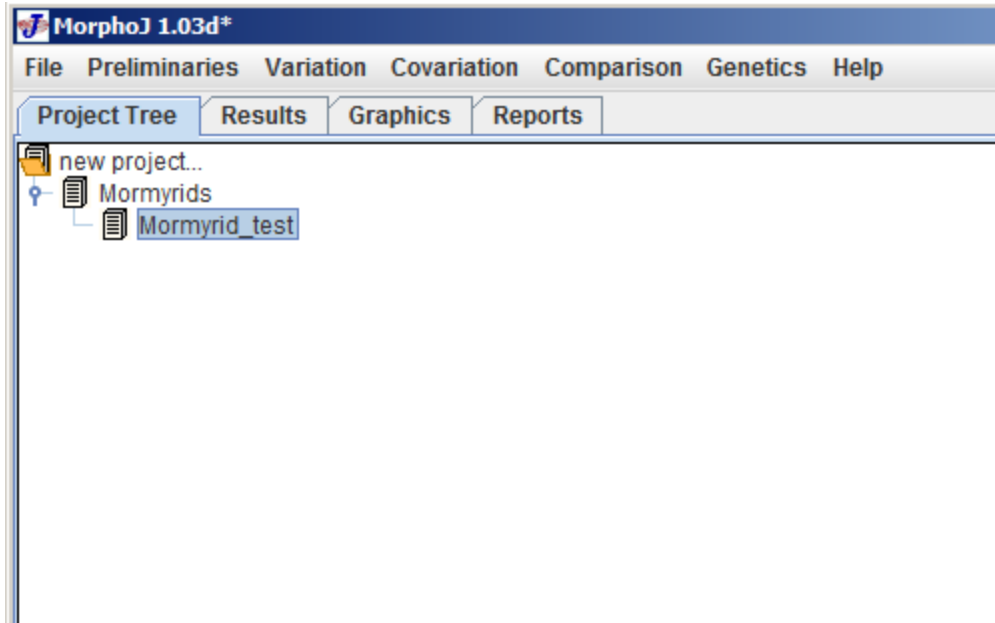
Note:

There might be more than one data set in an existing project. MorphoJ runs an analysis on the data set highlighted in the **Project Tree** tab, so it is improtant to select the correct data set to run an analysis.
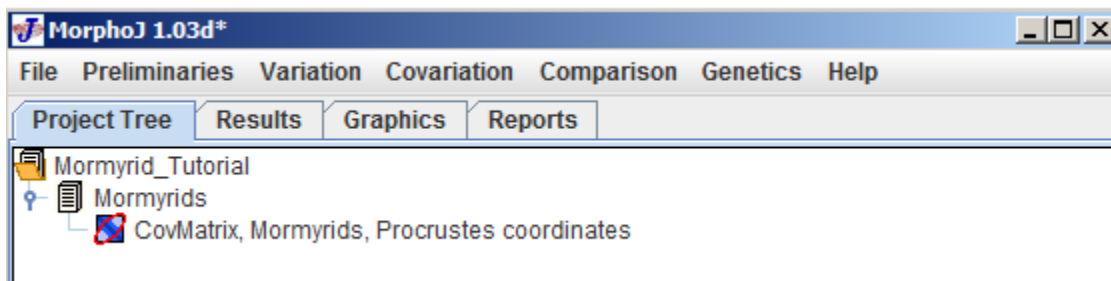
7   Covariance Matrix

After procrustes super impositon its very important to run the covariance matrix. Covariance matrices are used for a lot of the morphometric analyses like Principal component analysis. MorphoJ can generate Covariance matrix from various types of data i.e. shape data set after the procurstes superimposition has been done or regression residuals.

 First select the **Mormyrid dataset** in the main project tree and select covaraince matrix from **preliminaries** menu. Then the following dialog box appears:
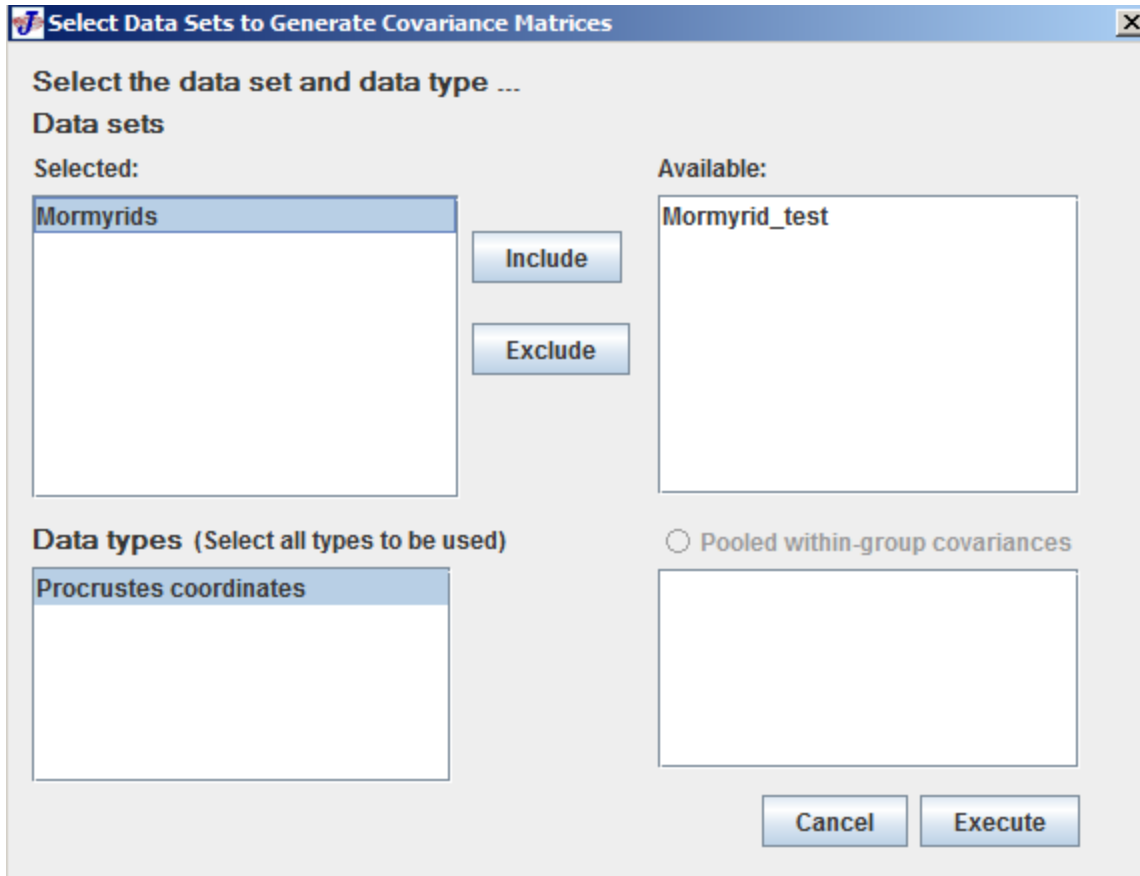
There are two datasets in this project. At the top left corner "**Selected**" has the data set selected to generate covariance matrix and Mormyrid test is excluded which is under Available box. Data types shows what kind of data are present in the current data set. After selecting data set and data types click **Execute** which opens the following window:


In this window we can see the covariance matrix under Mormyrids data set.

8 Classifiers

For this part of the tutorial we will be using Lythrurus data set as the morymrid photos did not have names. After importing the data file and running the Covaraince matrix the project window should look like the following:
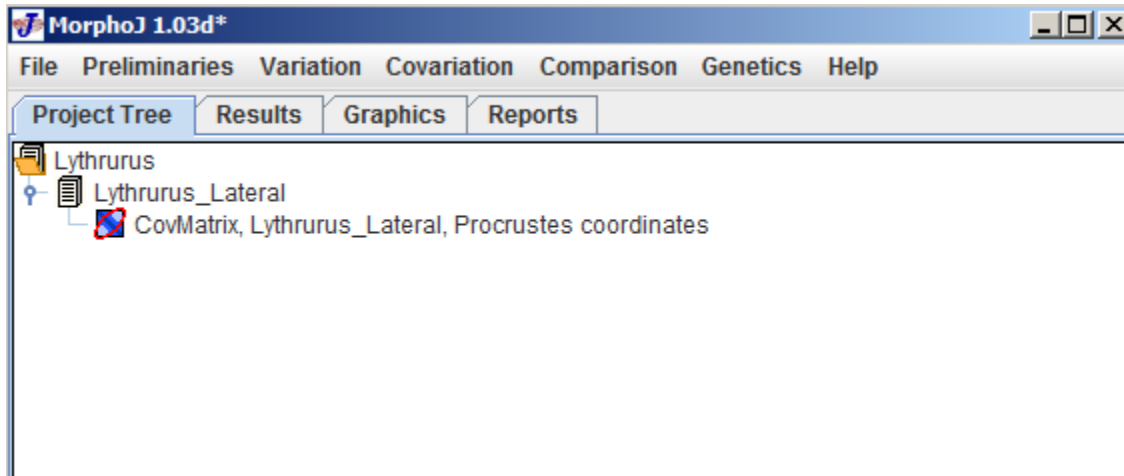

Select the Lythrurus_Lateral data set. Under **Preliminaries** menu click **Extract New Classifier from Id String** and the following dialog box appears:
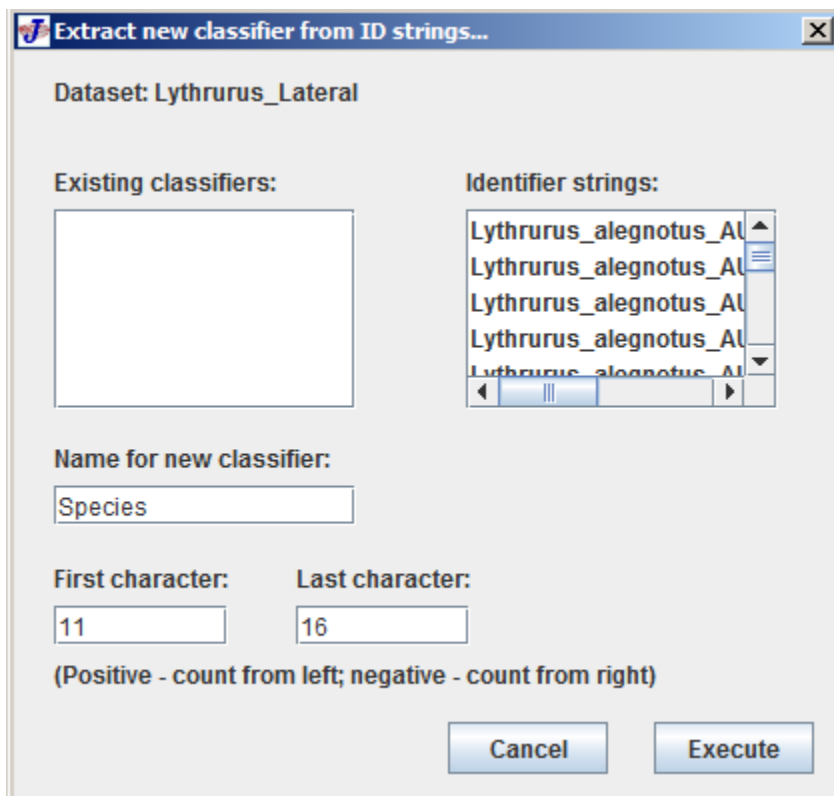
On the right side of the dialong box **Identifier Strings** list shows the file name for the studied individuals. On the left side the list shows if there are any existing identifiers. Here we name Species for the **Name for new classifier.** To extract a classifier choose the first letter that you will want to use. In this case, we want species. All species are in Lythrurus, which is 9 letters long, there is then an underline, so the first letter of the species is 11. Choosing 6 letters will ensure that you get enough to determine species for all photos. Select **Execute**.

It is always a good to check the results after extracting classifiers by selecting **Edit Classifier** under **Preliminaries** menu and the user interface opens the following dialog box
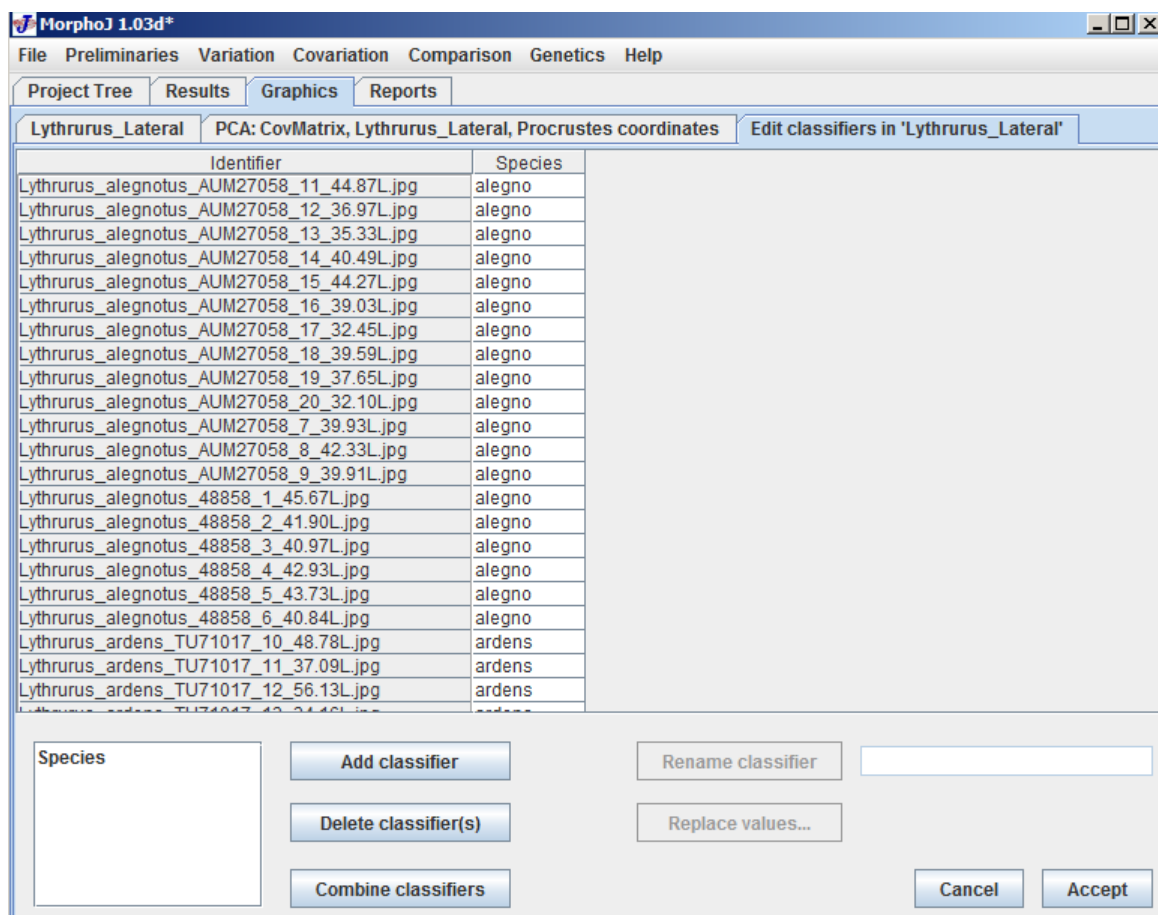

The list on the left of this dialog box shows the new classifier and if no editing is needed, just select **Accept** button**.**
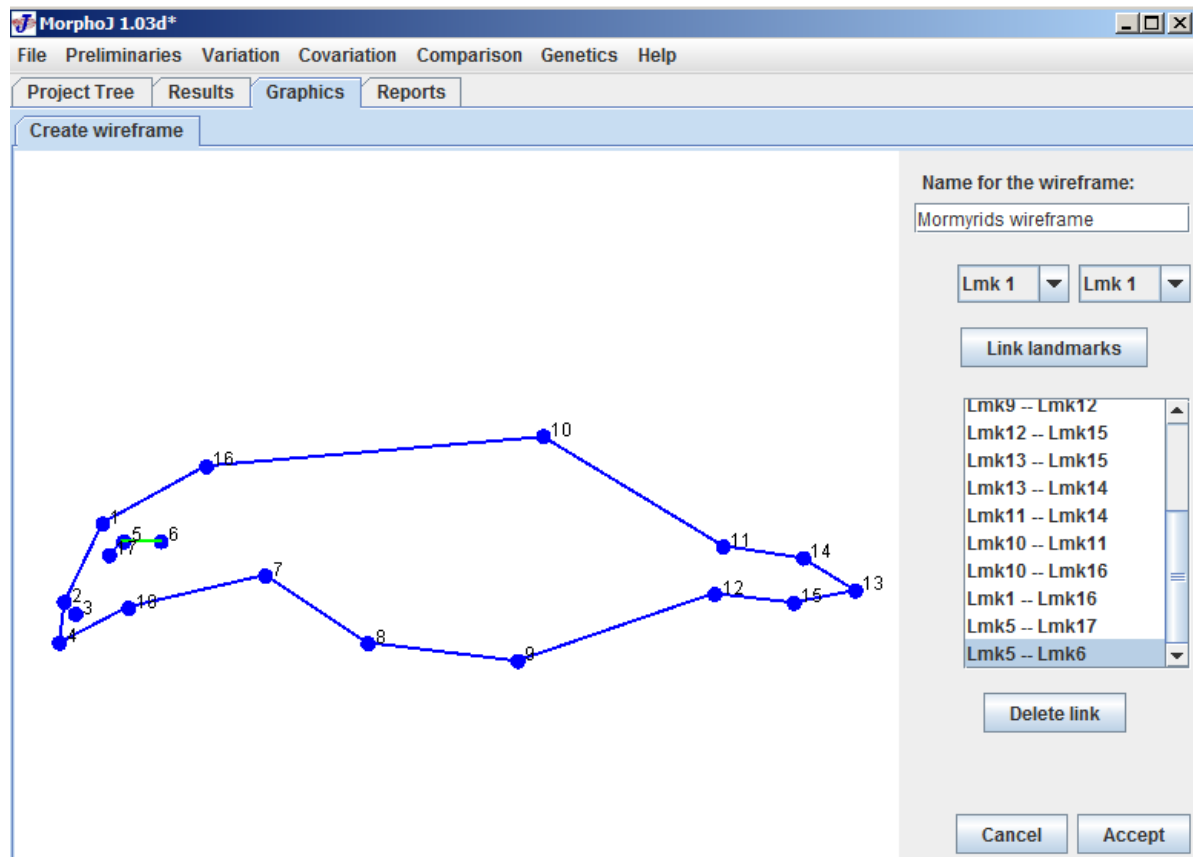

9   Wireframe

A wireframe is a set of lines which connect the land marks in a graph. It makes it easier for the viewer to check the confuguration of the landmarks and to visualize the results.

At first select the Mormyrid dataset under project tree window and click on
Preliminaries menu bar and then click **Select or edit wireframe**. The new graphics
tab appears:
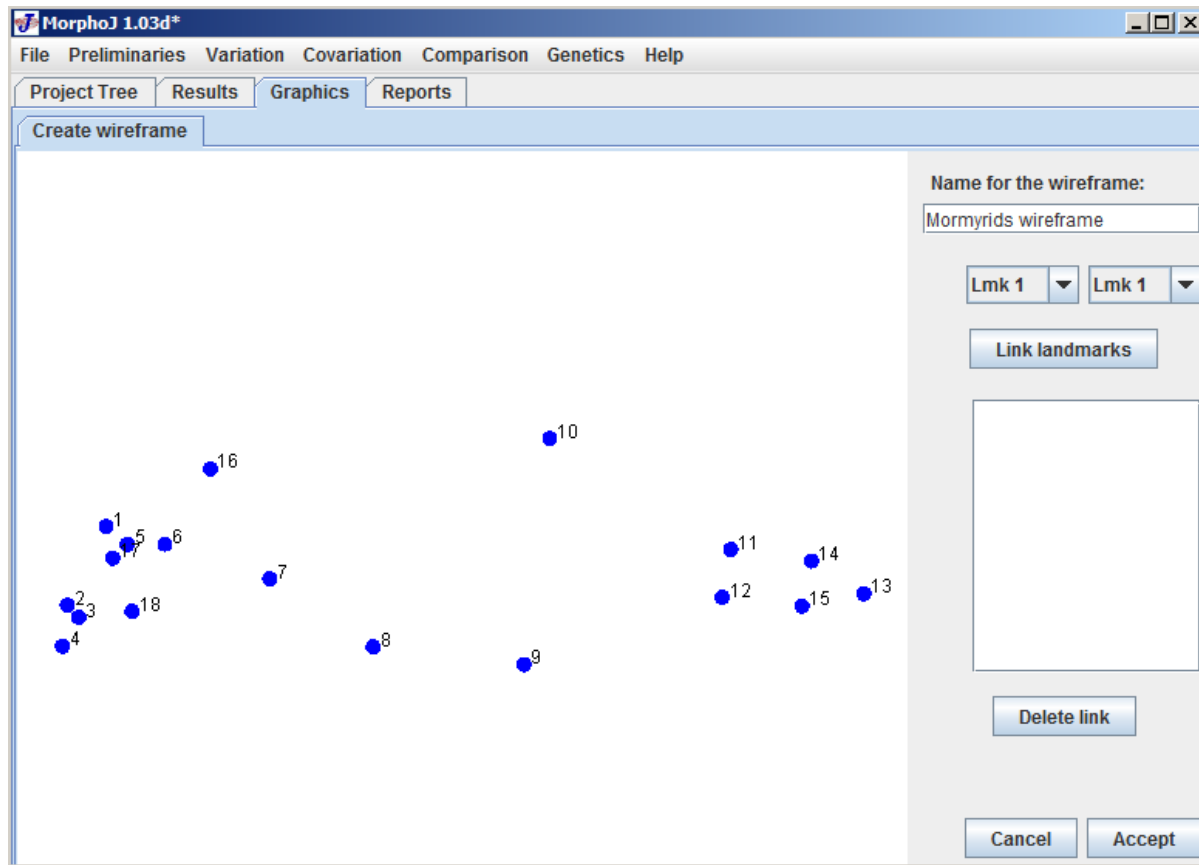
It is very important to have an idea about the landmarks and the organism has been
used in the study. The next step is to link each land mark where a solid line is desirable.
And the easiest way to do is it to click on one landmark and then drag to the next and at
the end click Accept button. As the wireframe is created the user interface widnows

looks like the following:

There are options that a wireframe could be hard or soft. As the wireframe is created the project window shows a tab with wireframe.
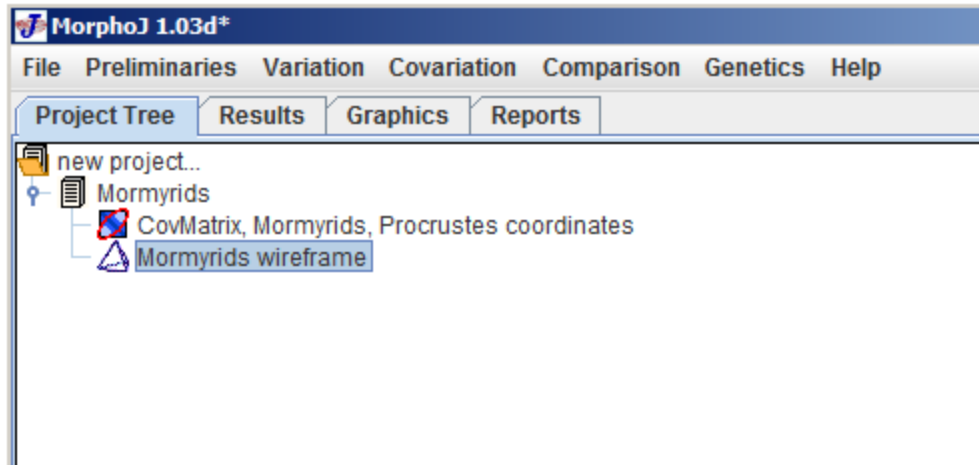
To edit an existing wireframe, click on the the wireframe in the project tree and then under **Preliminaries** menu select **Create or Edit Wireframe** option. A new graphics tab shows as **Edit wireframe**. Here in the link landmarks table it shows all the links of the existing wireframe. Any link could be selected and deleted and then press the **Accept** button.

There is an option in MorphoJ where you can choose not to include all the landmarks in a data set. This a great oppoortunity to test if some of the landmarks are necessary or you can exclude those. So to create a new dataset with some of the landmarks select **Mormyrids** data set in project tree window and**, then** click **preliminaries** menu and click on **Select Landmarks**.

The dialog box looks like the following:
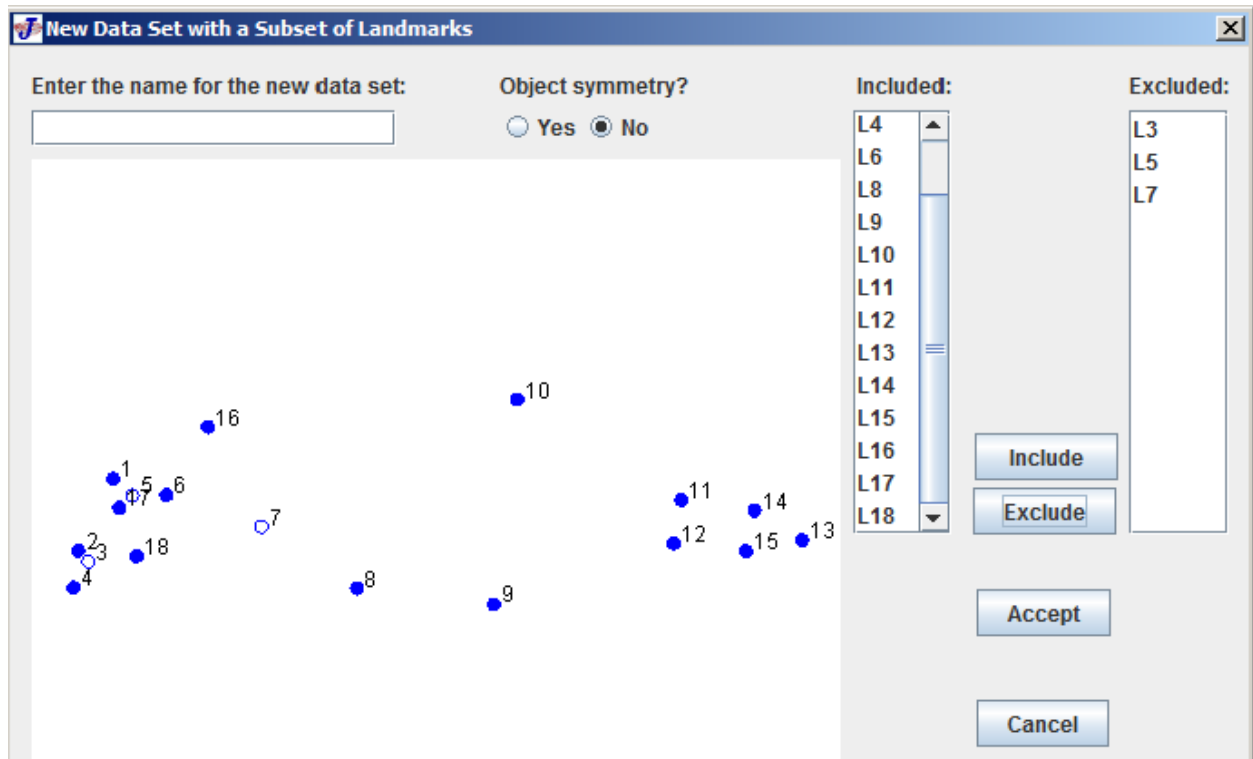
On the upper right corner there is a dialog box **Included** has the list of the landmarks that would create the new data set. The next dialog box contains the landmarks which have been selected for deletion. In this example 3 landmarks have choses to be deleted and it is shown in the left diagram as hollow circle. As this is 2D data set, by default it chooses **No** object symmetry.

10 Principal Component Analysis

Principal Component Analysis (PCA) is very important for a lot of studies as it shows major shape variation. It is one of the most widely used methods for multivariate analysis. In MorphoJ, PCA is calculated from the covariance matrix. To run a PCA select **CovMatrix** from the **Project Tree** window and click on **Principal component Analysis** (PCA) from the **Variation** menu. During selection the project tree window looks like the following:
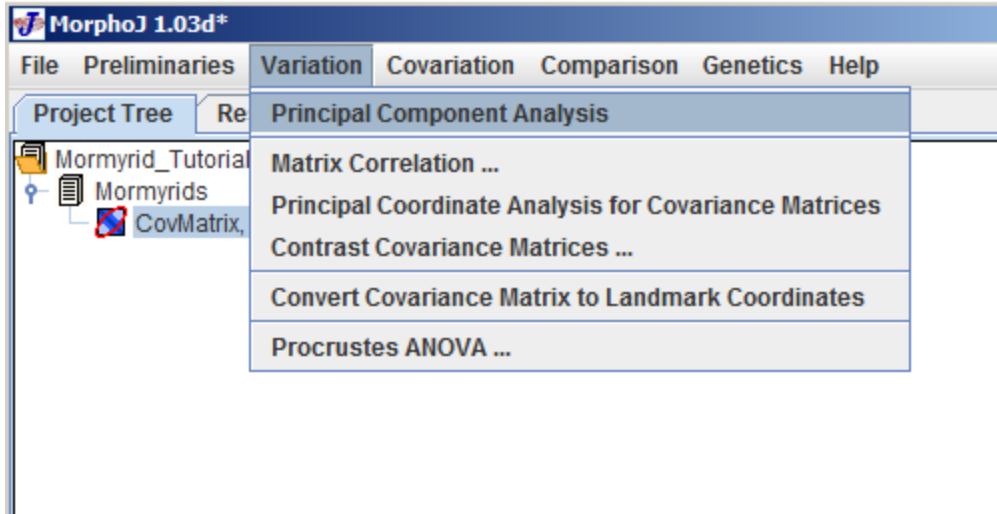
As we click on the PCA it gives two types of output, graphical and printed output. Graphical output has three tabs which are **PC shape Changes**, **Eigenvalues** and **PC Scores**. Clicking on any of these tabs show the graphical output. If some one selects the the **PC shape changes** tab the following window appears:
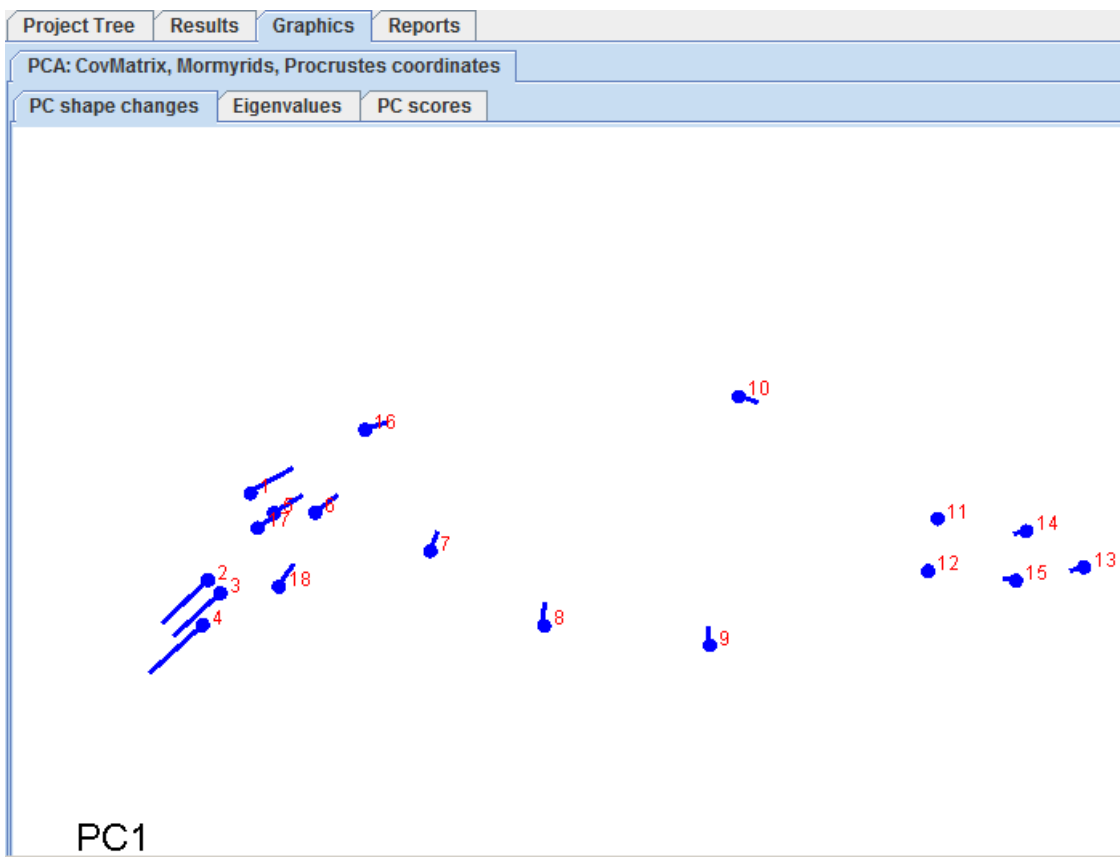
There are options that the PC could be changed and other modifications to the graph is possible as well. If we right click, the text box shows up with the options:

Transformation Grid, Warped Outline Drawing, and Wireframe Graph. These will give you different ways of visualizing changes along the different PCA axes. For a **Warped Outline Drawing,** an outline will have to be created in a program like the tps package and entered under **Preliminaries**.
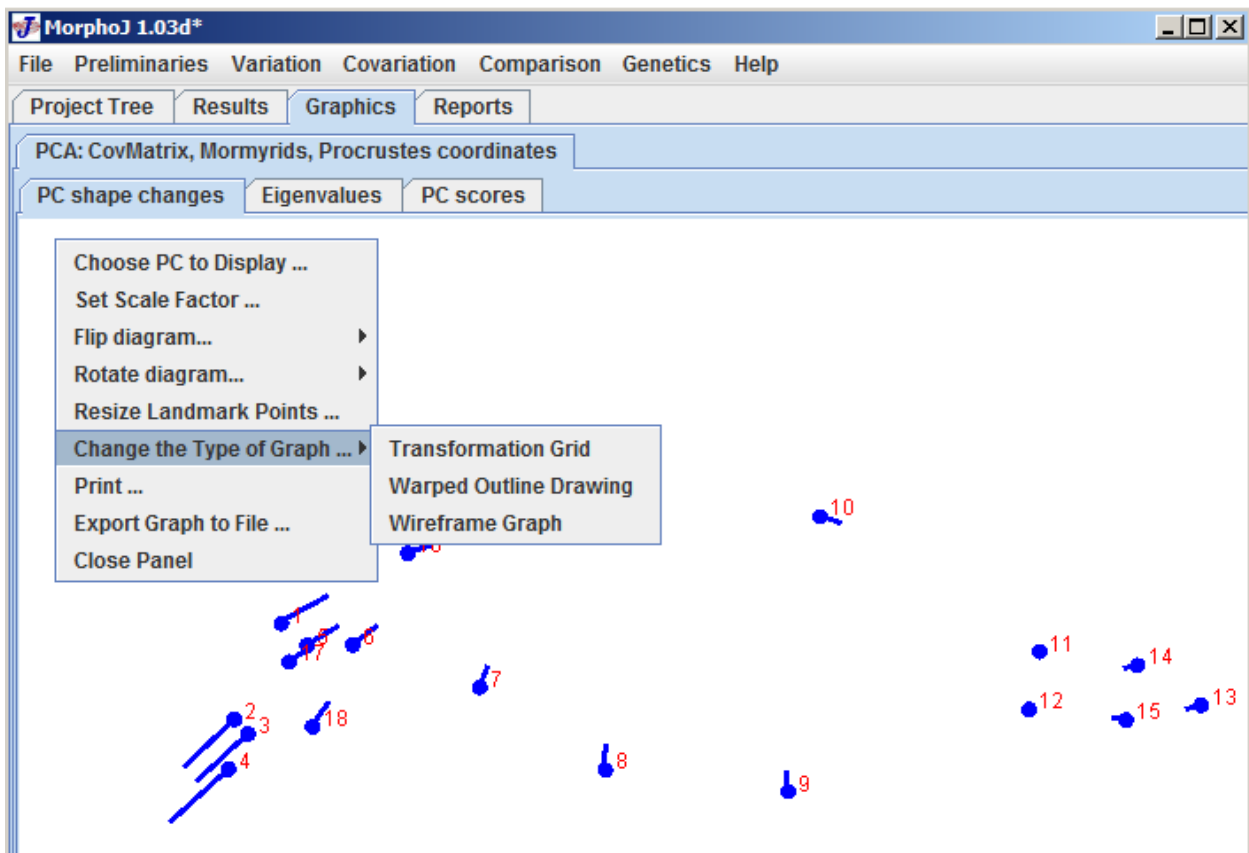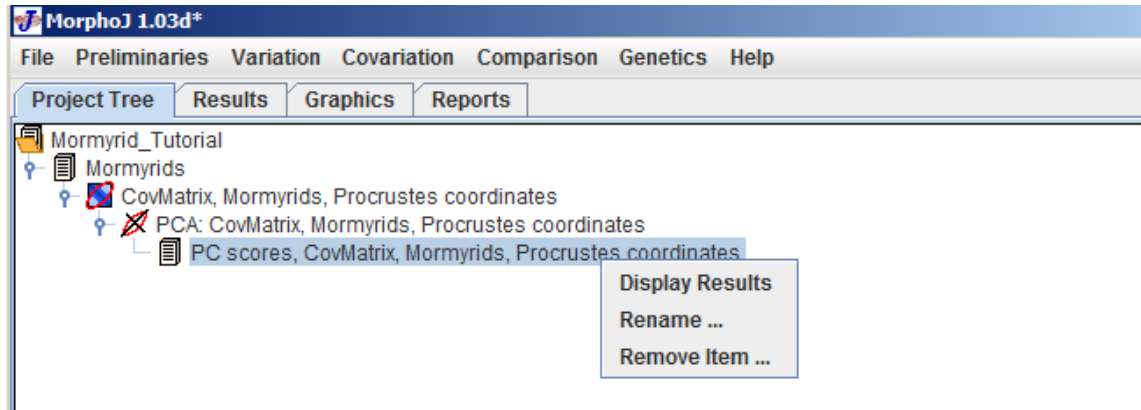
The Eigenvalue graphical window shows the relative percentage of principal components, where all PCs account for 100%. In shape analyses, the number of axes is the number of landmarks times the number of dimensions minus the degrees of

freedom (4 for 2D). If the **Eigenvalues** are chosen under graphical tab the following window shows up. In general, no more than the first 5 axes show enough variation to be useful.
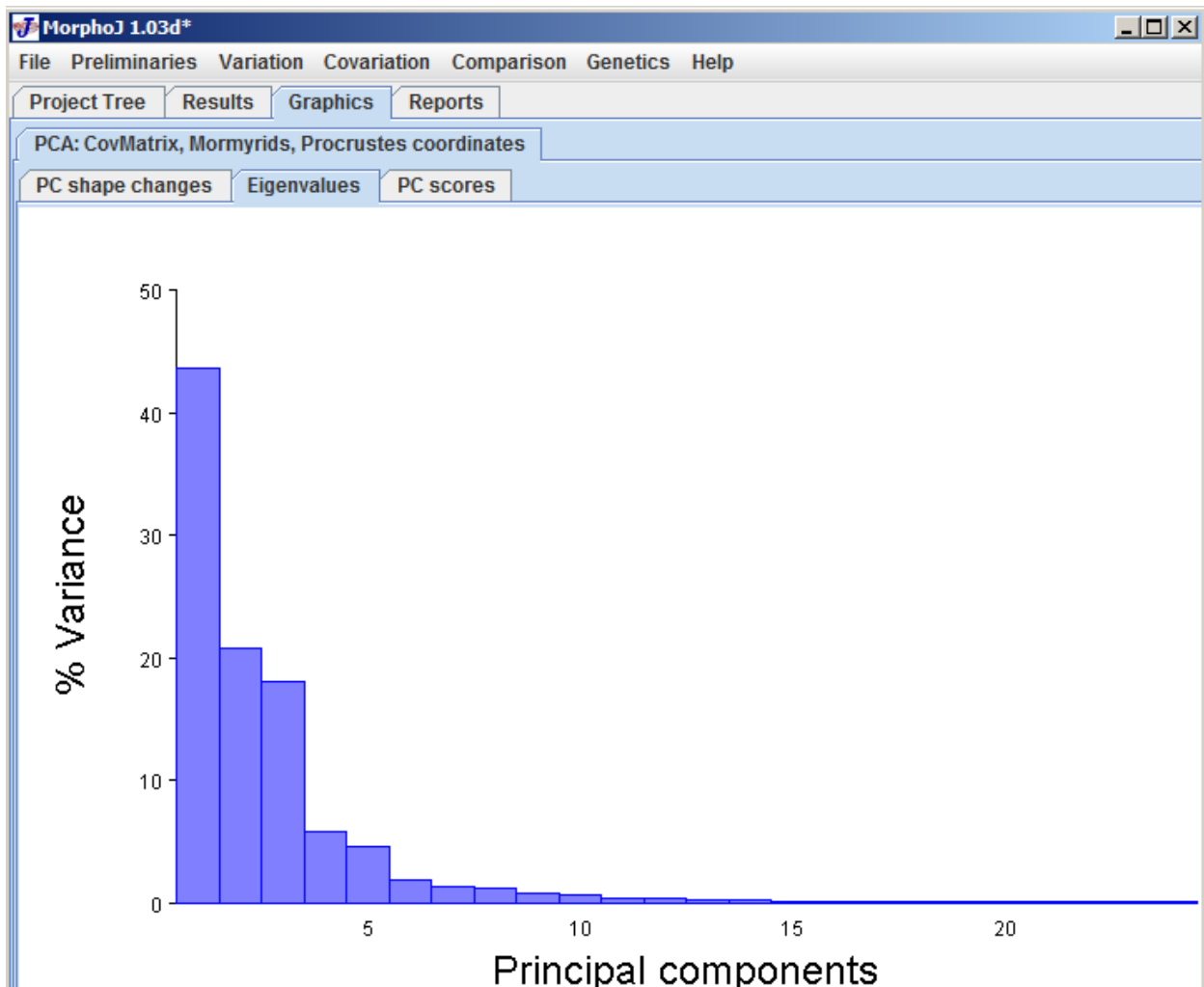
Selecting **PC scores** under **Graphics tab** gives the following user interface window:
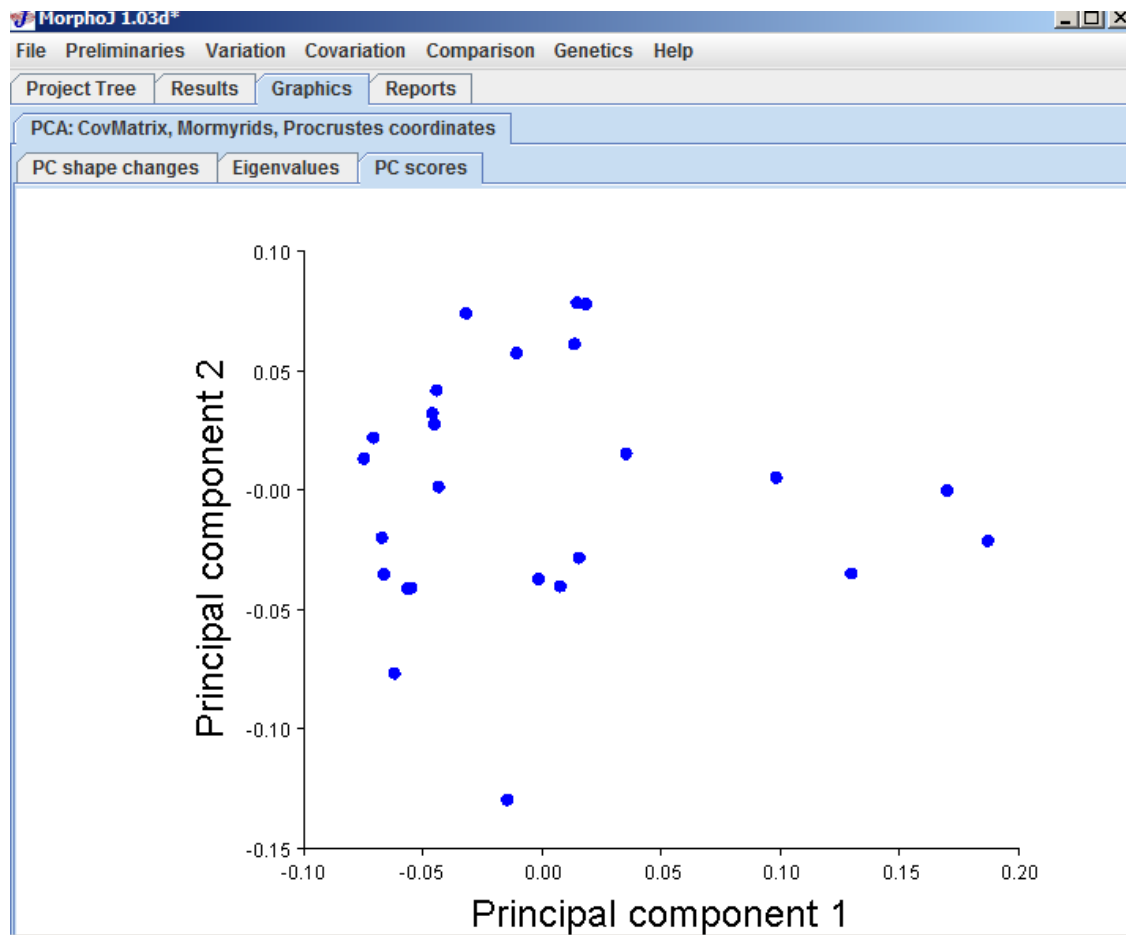
This is the scatter plot of PC scores. This graph is only produced if the covariance matrix on which the PCA is derived, directly comes from a dataset.

Right click on this scatter plot giver the following tab which gives lots of option to edit the scatter plot and also how to save and export the graph.
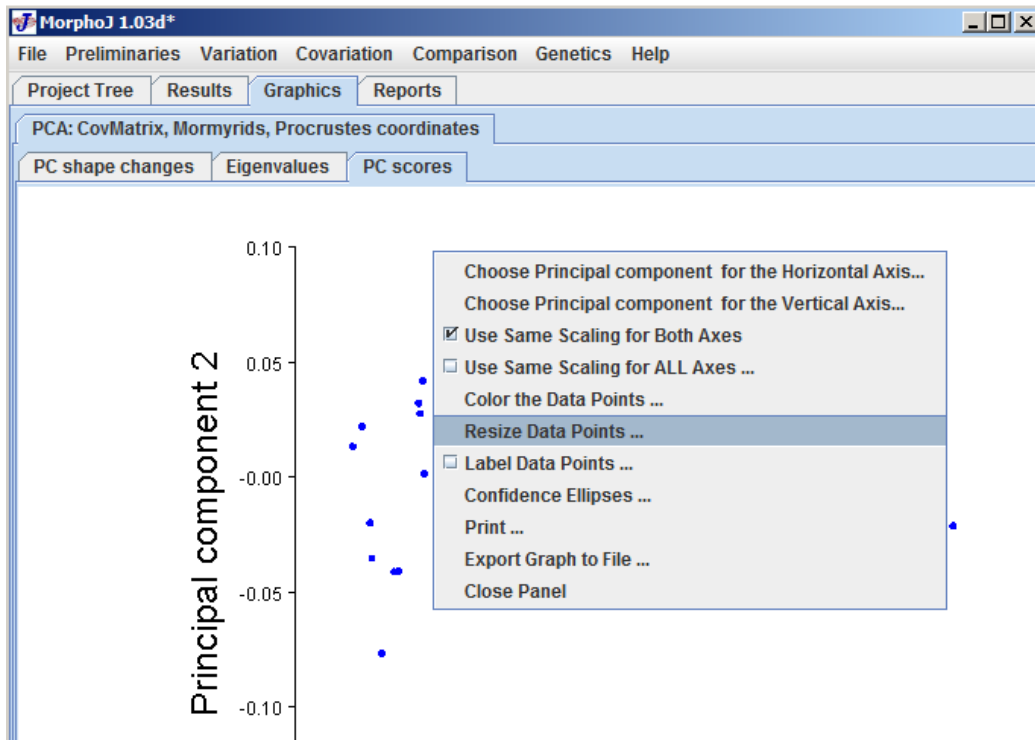
Showing Output Report as Text:

To see the text output for PCA choose Results and they are output as eigenvalues and PC coefficients.

The other way to see the report is selecting the Project tree window and right click on the PCScores and click result:
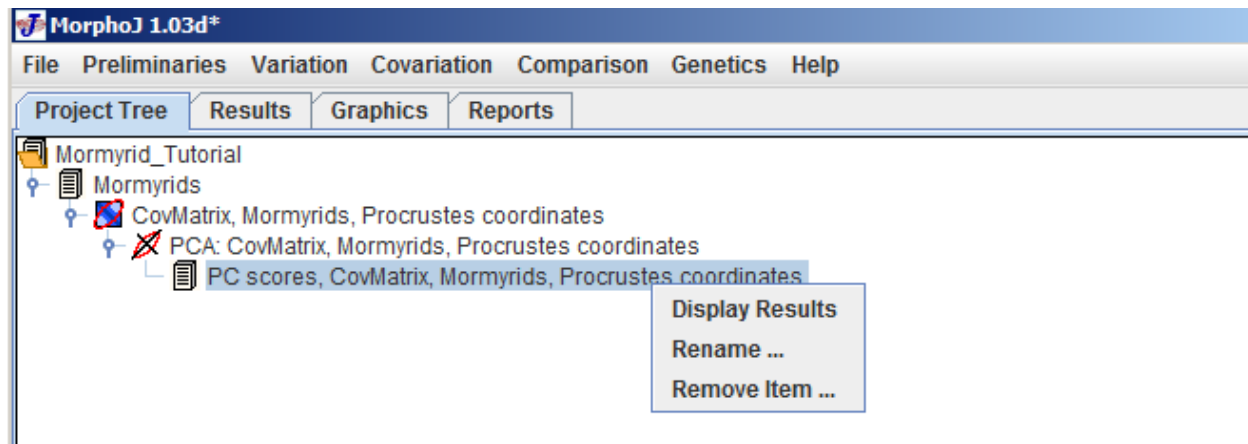
After clicking the following user interface window appears with the text output under the **Results** tab The text output in the Results window contains the following information:

MorphoJ 1.03d*

File Preliminaries Variation Covariation Comparison Genetics Help

Project Tree | Results | Graphics | Reports

Principal Component Analysis: PCA: CovMatrix, Mormyrids, Procrustes coordinates

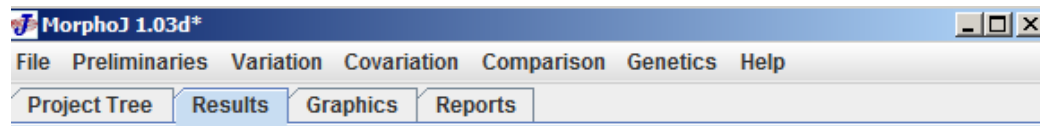|	Eigenvalues	% Variance	Cumulative %
1.	0.00549040	43.597	43.597
2.	0.00260612	20.694	64.291
3.	0.00226382	17.976	82.267
4.	0.00073164	5.810	88.076
5.	0.00057228	4.544	92.621
6.	0.00023287	1.849	94.470
7.	0.00016908	1.343	95.812
8.	0.00014762	1.172	96.985
9.	0.00009382	0.745	97.730
10.	0.00007936	0.630	98.360
11.	0.00004873	0.387	98.747
12.	0.00004422	0.351	99.098
13.	0.00003094	0.246	99.344
14.	0.00002316	0.184	99.527
15.	0.00001631	0.129	99.657
16.	0.00001215	0.097	99.753
17.	0.00000810	0.064	99.818
18.	0.00000699	0.056	99.873
19.	0.00000580	0.046	99.919
20.	0.00000432	0.034	99.954
21.	0.00000265	0.021	99.975
22.	0.00000132	0.010	99.985
23.	0.00000109	0.009	99.994
24.	0.00000078	0.006	100.000

Principal Component Coefficients

|	PC1	PC2	PC3	PC4	PC5
x1	0.299560	-0.028888	-0.079528	0.023754	-0.330969
y1	0.166393	-0.165195	0.208535	-0.091882	-0.029748
x2	-0.306434	0.011898	0.102343	0.014502	-0.113483
y2	-0.302048	0.222358	-0.232372	-0.116990	-0.140060
x3	-0.317040	-0.031981	0.129470	0.071695	-0.016001
y3	0.301442	0.183804	0.214078	0.067327	0.076648

The text output contains two tables, one is Eigenvalues and the other is PC coefficientns.

Eigenvalues: In this table there are 3 columns. The first one is eigenvalues for each corresponding PC axis and the second one is as a percentage of the total variance. The third column is the cumulative percentage of total variance.

PC coefficients: The PC coefficents or eignevectors are in a tabular form which includes all the PC coefficints.In this coefficient table columns contain coefficients for one principal componenet.

11 Canonical Variate Analysis

Canonical variate analysis is a method used to find shape features which maximize the separation between groups. For this analysis group is assumed to be known apriori. For exploration taxonomy, CVA is generally inappropriate. You need to have a justifiable reason for defining groups. For CVA analysis select Canonical variates under comparison and the following dialog box opens:
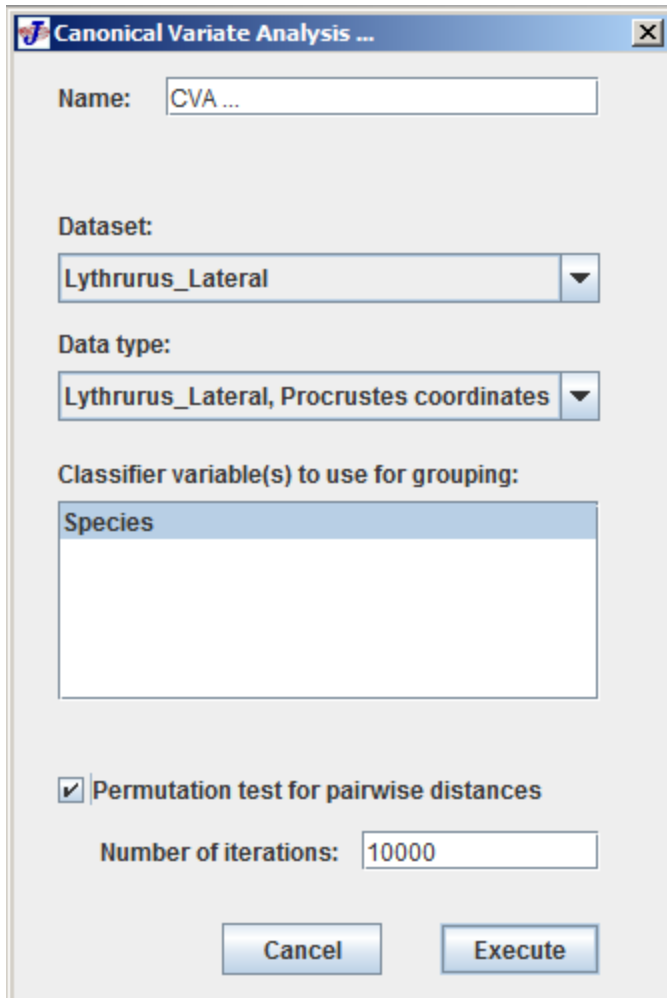
At the very top of this dialog box is a drop down menu to select what kind of analysis to run and by default it does a CVA.

The **Dataset** drop down menu gives an option to choose a dataset and here in this example it is the Lythrurus_Lateral. As the dataset is selected the Data type and Classifier variables are available to be selected for the next two elements of the dialog box.

At last, there is a check box that can be selected for a permutation test of pairwise distances between the groups. Click the **Execute** button to run the analysis. **Graphical Output**: As the CVA analysis is done there are two graphical outputs under the Graphics tab. One of them is **CV Shape Changes,** which shows the shape change associated with the CV axes. The number of axes is the number of species minus one. The other graphical output is **CV scores,** which is the scatter plot of the CV scores.

To view the results, choose the **Results Tab.** The following screen shot shows some of the elements of the user interface:

```
MorphoJ 1.03d*

File  Preliminaries  Variation  Covariation  Comparison  Genetics  Help

Project Tree    Results    Graphics    Reports

Canonical Variate Analysis: CVA ...


Dataset: Lythrurus_Lateral
Classification criterion: Species


Groups    Observations
1.        alegno  19
2.        ardens  30
3.        atrapi  25
4.        bellus  29
5.        fascio  20
6.        fumeus  35
7.        lirus_  21
8.        matuti  31
9.        roseip  30
10.       snelso  6
11.       umbrat  30


Variation among groups, scaled by the inverse of the within-group variation


          Eigenvalues      % Variance        Cumulative %
1.        8.51887316        36.719             36.719
2.        3.54695503        15.288             52.007
3.        3.23147984        13.929             65.936
4.        2.25753541         9.731             75.666
5.        1.45884809         6.288             81.954
6.        1.34017265         5.777             87.731
7.        1.07740052         4.644             92.375
8.        0.78608731         3.388             95.763
9.        0.56248547         2.424             98.187
10.       0.42053627         1.813            100.000
```

At the very biginning of the test output it gives information about the name of the analysis, dataset, number of groups etc.

The first entries are how many groups, the criteria to make the groups and how many individuals are in each group.

The second table shows the amounts of variation among groups. In this table there are 3 columns. First column gives the amounts of variation among groups, 2nd column is the percentage of the total for each CV and last column is the cumulative percentage.

The next two entries are matrices for Mahalanobish distance and Procrustes distance among all possible pairs of groups. If the option for Permutation test was selected during the analysis it gives two matrices with P values for all the pairwise tests for Mahalanobis and Procrustes distances.

Finally is a table with canonical coefficients which is related to the original shape coordinates to the CV scores.

12 Incorporating Phylogeny

MorphoJ can be used to map shape data onto phylogenies using squared-change parsimony (Maddison 1991). There is also a permutation test to determine if the morphometric data contains phylogenetic signal or not. Phylogeny files are imported into MorphoJ in the NEXUS file format. For this section we will be using Lythrurus.NTS data file.

Mapping onto phylogeny:

First you must have a phylogeny. Ideally, this phylogeny should have branch lengths, as branch lengths provide a better estimate for ancestral values in squared-change parsimony. This phylogeny can be created in most phylogenetic programs by saving the output tree with branch lengths. You cannot simply build this tree, but must have the original dataset input into the phylogenetic programs and can then build the tree. The taxa in the tree must match those in the shape analysis. Tree files can be edited in text reading programs.

To map a phylogeny onto shape data, the first step is to import a phylogeny file as NEXUS tree format (.tre). Before that everyone needs to make sure the project has PCA done so the main user interface looks like the following:
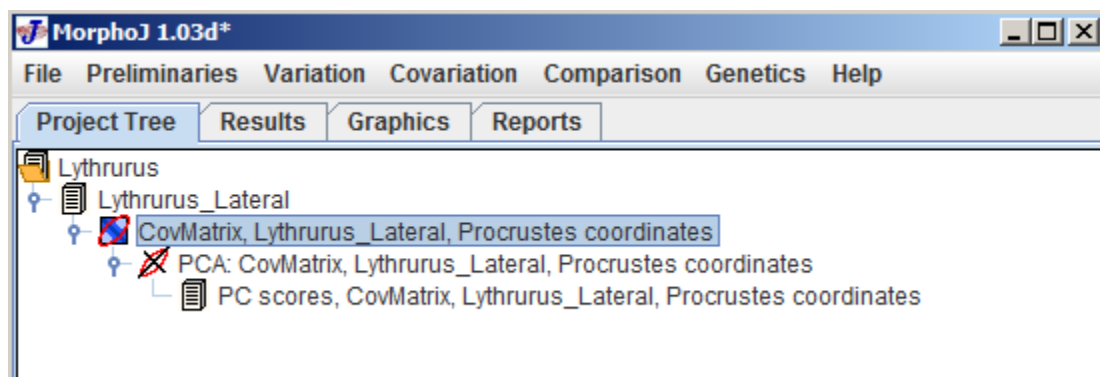

To import the phylogeny file click on File menu and click on **Import Phylogeny File**, which opens the following user interface:
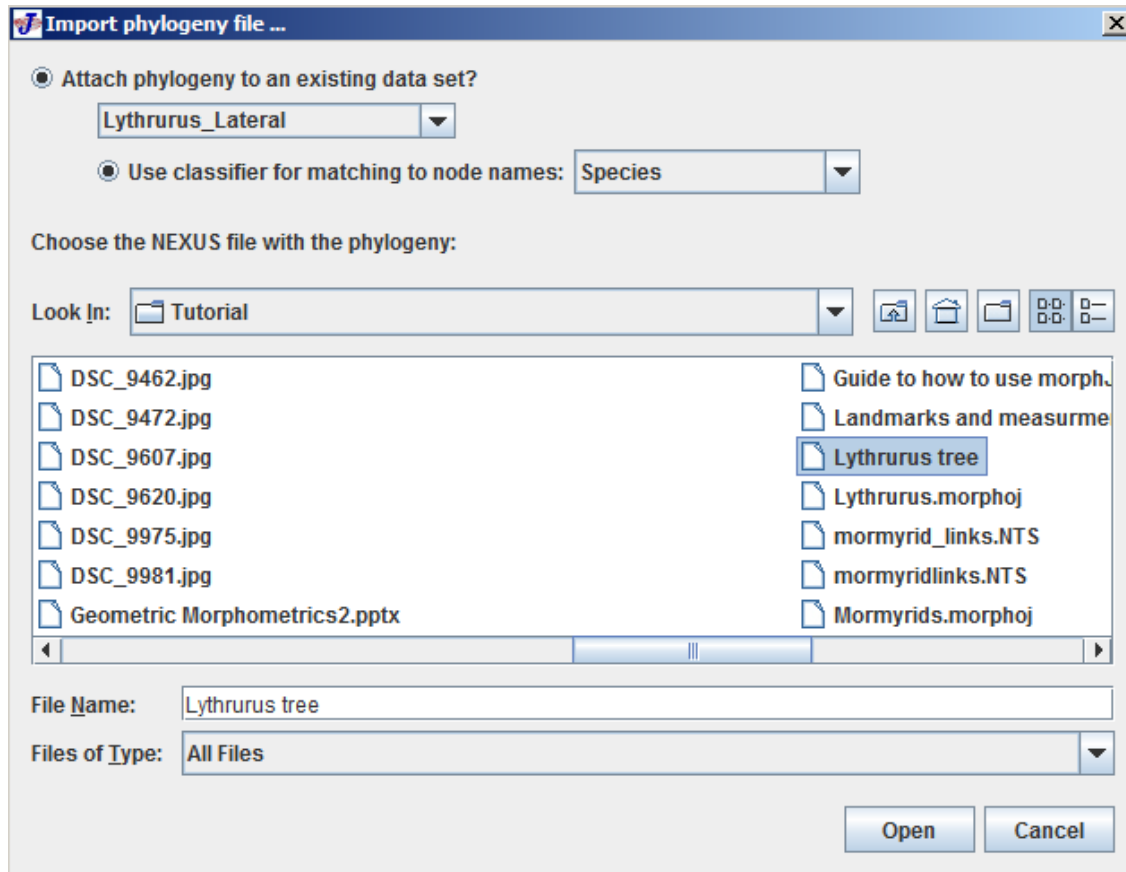
There are two buttons to choose the current dataset and choose the classifier. For this example the dataset is Lythrurus_Lateral and the classifier is species have been selected. The choose the tree file and select **Open.** When the tree opens the project tree window looks like the following:
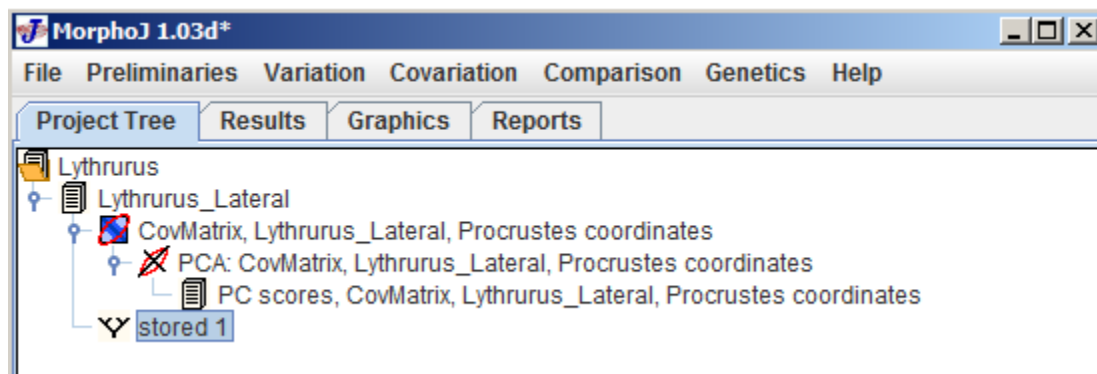

In this Project Tree the Phylogeny Tree is saved as Stored 1, to have an appropriate name for the tree, user can change it in the Nexus file. To check if the Phylogeny Tree has been imported proeprly a left mouse click on Stored1 and selecting **Display Graphs** shows the following dialog box:
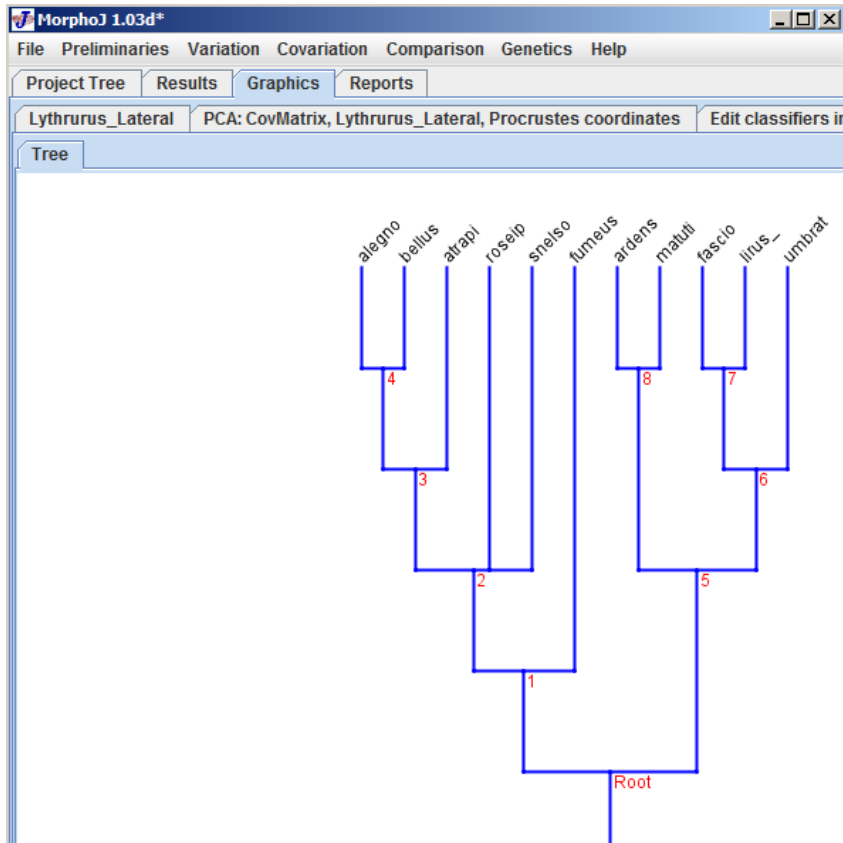
To map morphometric data onto phylogeny, select PC scores in the project Tree window and select **Map onto phylogeny** under **Comparision** menu. The user interface would look like the following:
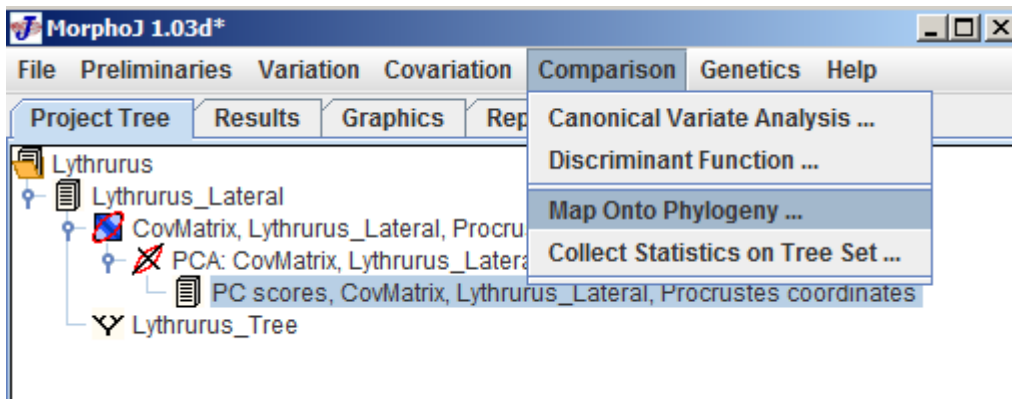

After selecting PC Scores, the following dialog box appears
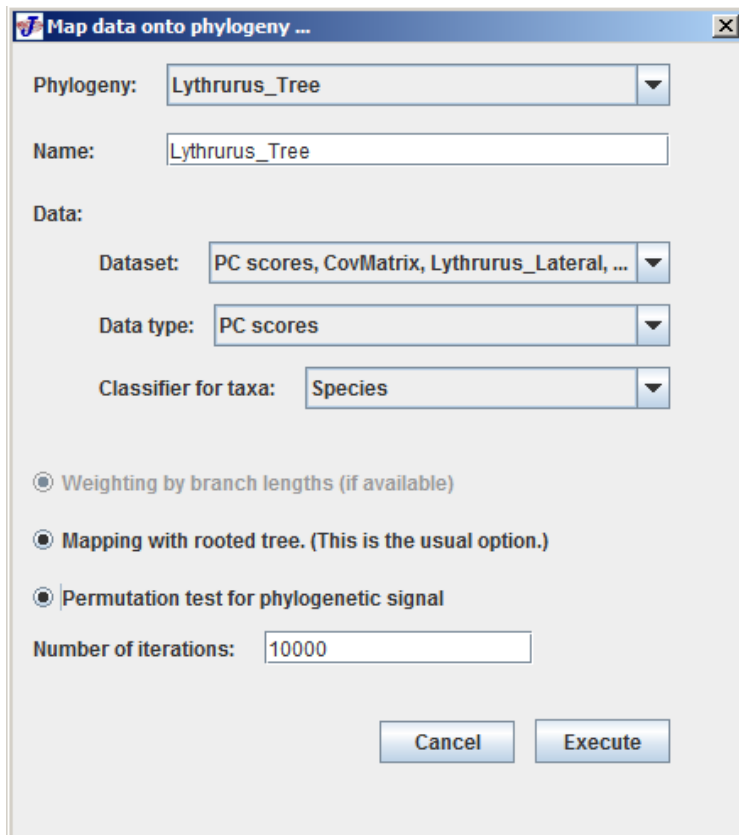
At the very top of the dialog box there are a few text boxes that give an option to choose the dataset and phylogeny tree.

The last button is **Permutation test for phylogenetic signal** which starts with the null hypothesis of not having any phylogenetic signal in the data set.

After all these selections click **Execute** and the following graphical user interface appears:
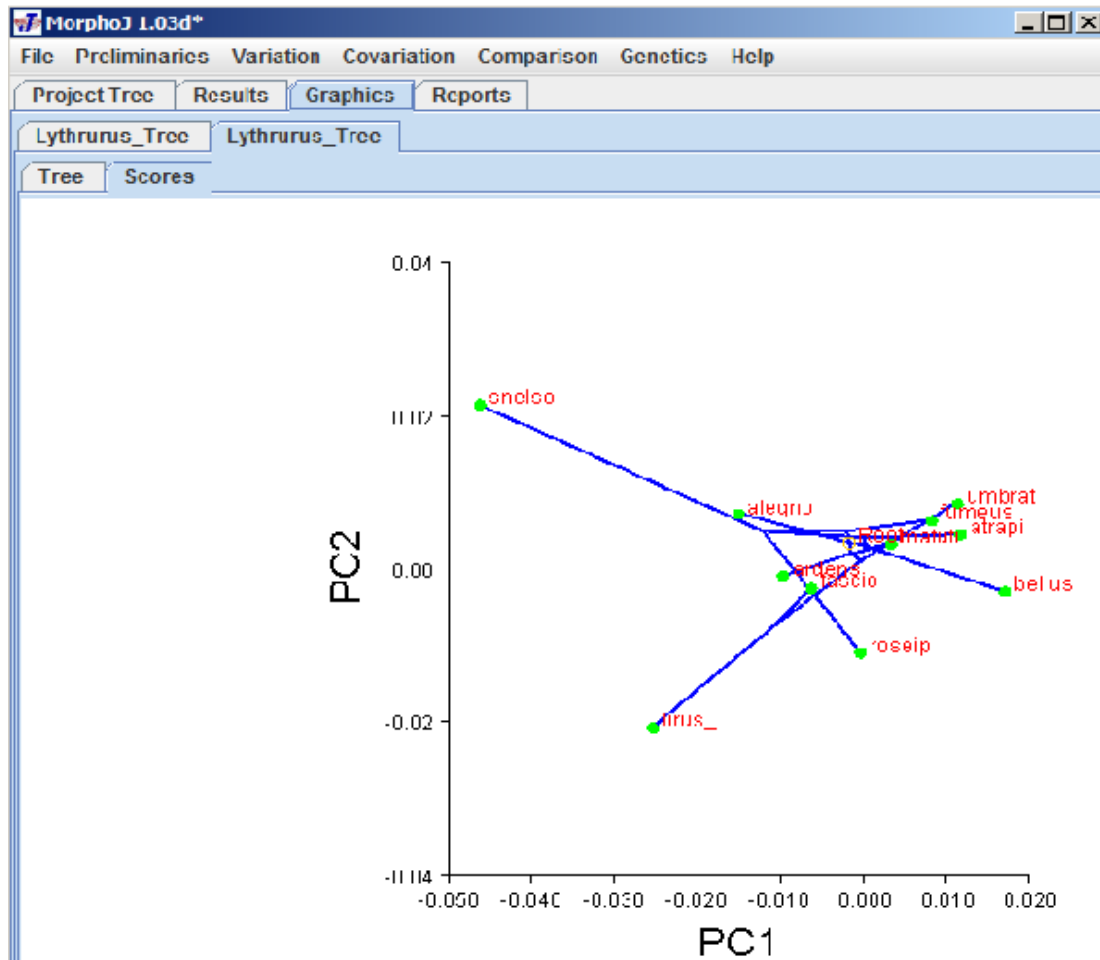
The graphical presentation depends on what kind of dataset has been used to overlay the phylogeny. A left click anywhere in this window gives a popup menu. This popup menu can be used to change the orientation of graph, scale factor, which PC axes to

show, and to export the graph file and save it.
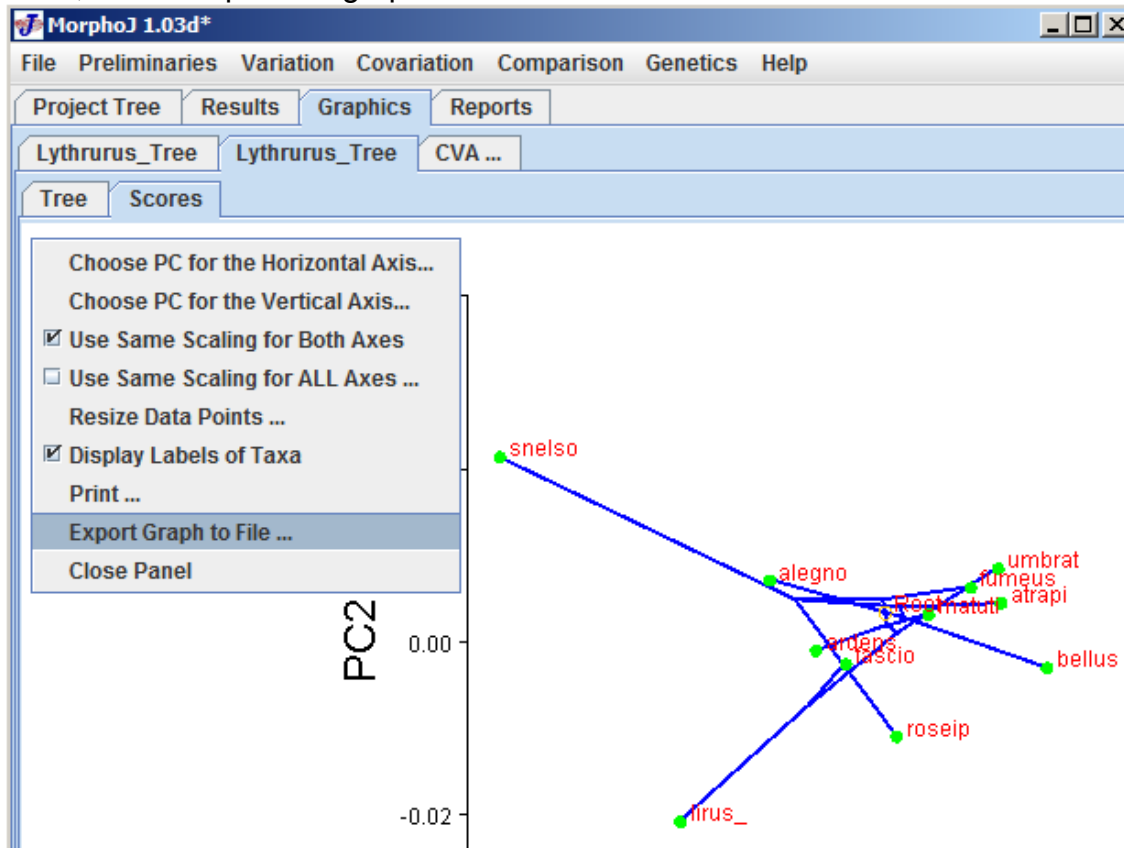
In this example we have mapped phylogeny tree on the PC data. Selecting the Result tab gives the result for this analysis and you can find the results of the permutation test. For this result the P-value = 0.7355, so no phylogenetic signal. Phylogeny can be mapped also onto Canonical Variates.

13 Conclusion

This user guide shows the most common features of MorphoJ to perform Geometric Morphometrics in taxonomic and phylogenetic research. It does not include every analysis that can be done with this very powerful program. We provided only basic information on the proper use of these anayses, and you should seek out more information on statistical analyses.

Reference:
MorphoJ user's guide at
http://www.flywings.org.uk/MorphoJ_guide/frameset.htm?index.htm

Klingenberg, C. P. 2011. MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources* 11**:** 353-357.