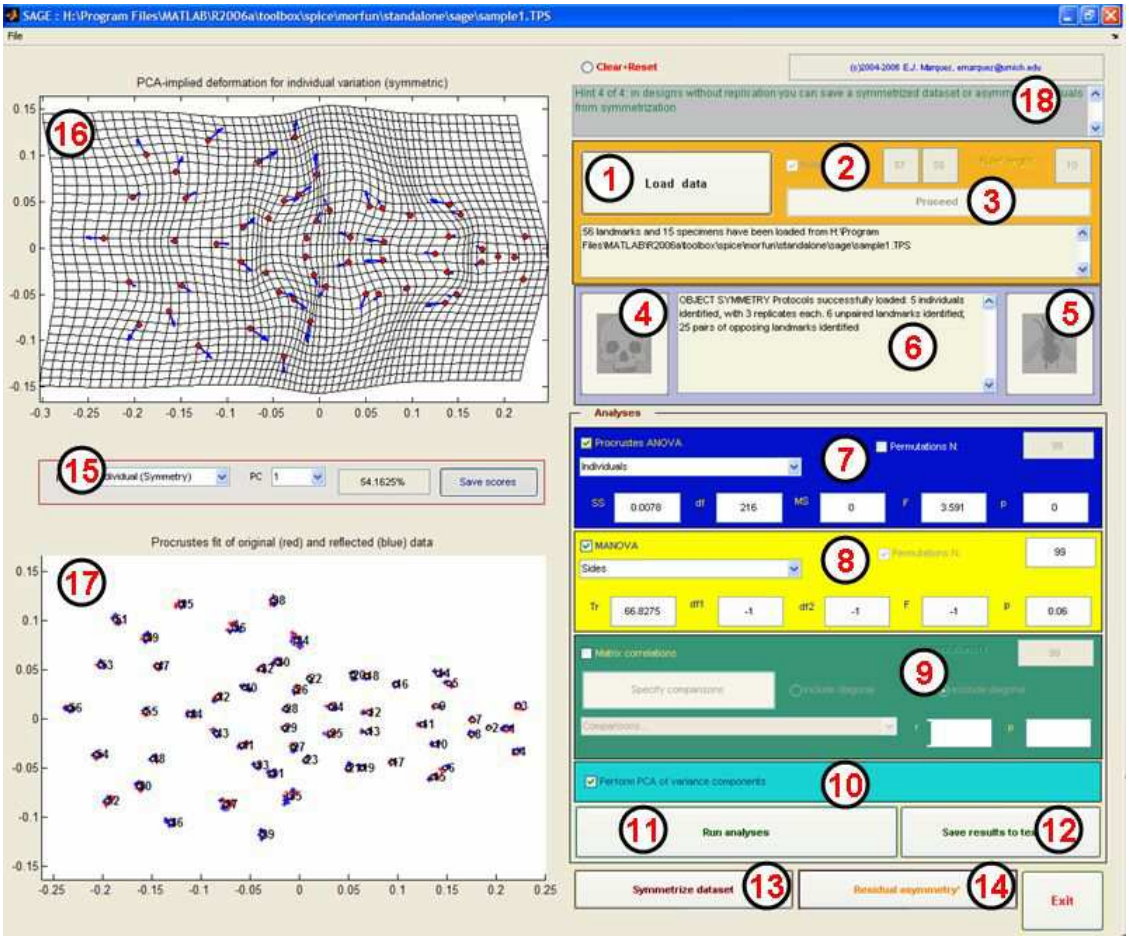
Sage: Symmetry and Asymmetry in Geometric Data
Version 1.21 (compiled 03/1114)

© Eladio Marquez 2012-2014
Mammals Division
University of Michigan Museum of Zoology
http://www-personal.umich.edu/~emarquez/morph/

Reference Guide

Most functions in Sage are accessible from the main panel. Additional functions, particularly file saving operations, may be found in the **File** menu. An overview of the main functions available through Sage is schematized in the following figure and explanation:



Description of functions:

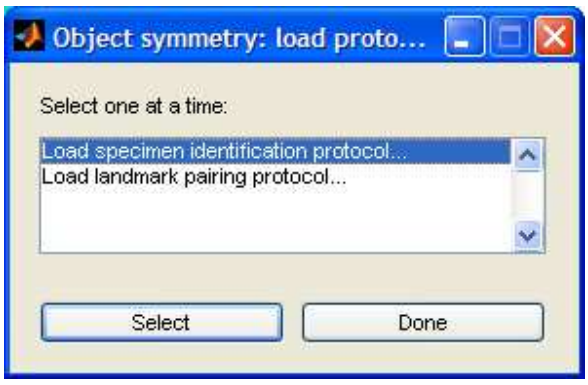
1. **Load data** button: this is the only activated button in Sage at startup. Data formats allowed by Sage are "XY" (X1,Y1,...,Xn,Yn, where each coordinate is in a column), David Sheet's IMP (X1,Y1,…,Xn,Yn,CS, where each coordinate is in a column and CS stands for centroid size of each individual configuration), and basic James Rohlf's TPS (currently supported tags are LM and SCALE, any other tags should be removed; file name extension must be .tps or .TPS in order for Sage to recognize this format). For object symmetry, specimens are not required to have any particular order. For matching symmetry, though, left-side landmarks must be stacked as a group below the right-side landmarks, using the **same order of specimens** in both cases. A sample dataset in TPS format, sample1.tps, is included in this release.

   Note: in the current version of Sage, all specimens must have the same number of replicates. Otherwise, the program will not function properly.

2. After choosing a data file to load, you can check **Use Ruler** if the two endpoints of a reference ruler have been included among the landmarks. Note that in TPS format, use of ruler overrides the use of the SCALE tag to scale data. After checking this box, enter the landmark numbers corresponding to the ruler. By default, the last two landmarks are set as the ruler endpoints.

3. Data is actually loaded when you press the **Proceed** button.

4. After loading your data, you need to inform the program about how they are organized. You do so by loading "protocols". Protocols are simple files which contain one or two columns of numbers that indicate how rows/columns in data files belong together. The button **Protocols for object symmetry** displays a dialog of options to load protocols unique to this type of symmetry:



Both protocols must be loaded for the Analysis panel to be enabled. The specimen identification protocol consists of a column of **numbers** identifying specimens for each row in the dataset. Testing for fluctuating asymmetry always requires at least two measurements (replicates) per individual. **All replicates of a single specimen must be labeled with the same ID number**. For example, a

protocol for 9 individuals with 3 replicates each, under object symmetry, may look like:

```
1 1 1 2 2 2 3 3 3 4 4 4 5 5 5 6 6 6 7 7 7 8 8 8 9 9 9
```

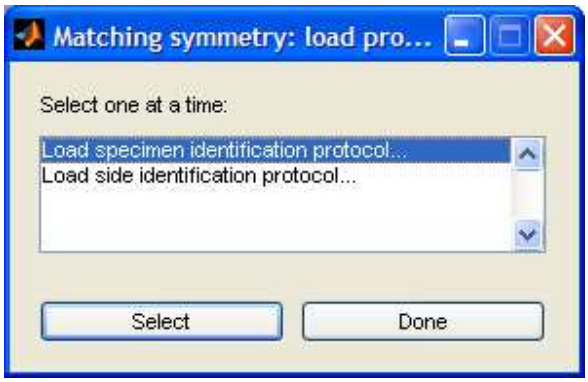
arranged in a column instead of a row. The corresponding dataset should have 27 rows.

The landmark pairing protocol consists of two columns with numbers representing landmarks in the dataset. Each row contains thus a pair of **different** numbers indicating which landmarks correspond to the same feature in different sides of the structure. Those landmarks that lie along the axis of symmetry (midline) are identified in the first column, with a zero in the corresponding position in the second one. For example, consider the following protocol for 17 landmarks:

```
 1     0
 2     0
 3     0
 4     5
 6     7
 8     9
10    11
12    13
14    15
16    17
```

In this dataset, landmarks 1 through 3 lie on the midline, while landmarks 4-17 lie outside the midline, and therefore are paired at both sides of a structure. Landmark 4 is the same as landmark 5, 6 the same as 7, and so on, each lying in a different side of the structure. Sample protocols for object symmetry, corresponding to the included sample dataset, are packaged with this release.

5. Pressing this button instructs the program to upload protocols for matching asymmetry, through the following dialog box:



The specimen identification protocol is similar to the one used for object symmetry, with the difference that the matching symmetry dataset consists of 2*nr*

rows, where *n* is number of specimens, and *r* is number of replicates per specimen, and therefore the specimen identification protocol must contain the same numbers repeated twice in the column. For a case with 9 specimens with 2 replicates each, this would be:

```
1 1 2 2 3 3 4 4 5 5 6 6 7 7 8 8 9 9 1 1 2 2 3 3 4 4 5 5 6 6 7 7 8 8 9 9
```

again, with these numbers arranged in a column.

A side identification protocol consists of the same number of entries as rows in the data file (and the individual identification protocol), but with entries equal to zero or one corresponding to different sides. The side identified with zero will be labeled as the Right side, and the one identified with one will be labeled as the Left side. For the same example with 9 specimens and 2 replicates, the side identification protocol would be (in a column):

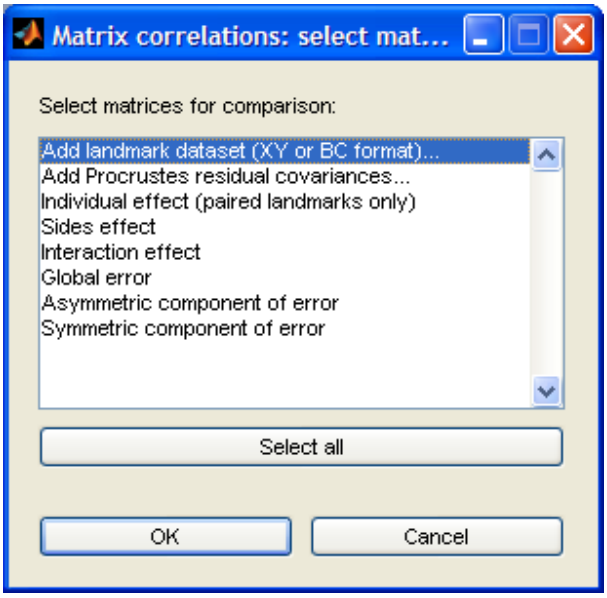
```
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

6. In these windows, the progress of file loading activity is reported. Check for inconsistencies or error messages.

7. **Procrustes ANOVA** panel: check if you want to perform this analysis. By default, p-values are computed using the F-distribution (parametric test). If you want to use a permutation procedure instead, check the **Permutations** box and enter the number of desired permutations. Once the results are computed, you can use the pull-down menu to access the ANOVA statistics for each effect. If matching symmetry is being used, an ANOVA for the size effect is automatically computed. Note that for this ANOVA to make sense either centroid size must be submitted as the last column in IMP format or data must not be scaled prior to loading it into the program.

8. **MANOVA** panel: same procedure as Procrustes ANOVA. However, computation of significance values requires that sample size is larger than the number of degrees of freedom for shape. If this requirement is not met, **Permutations** will be automatically selected and cannot be disabled, so significance values will be produced based on this procedure. The statistic used in MANOVA tests is the Lawley-Hotelling trace (labeled Tr in SAGE interface).

9. **Matrix correlations** panel: in this case, significance values for the correlations are always obtained by permutation of covariance matrices. Before running this analysis, though, comparisons to be performed must be selected by clicking on the **Specify comparisons** button. A dialog like the following will appear:



Here, you can select two or more of the covariance matrices already computed by Sage to perform comparisons. **All possible comparisons** between selected matrices are done. In addition, you can add a new landmark dataset (e.g. the original dataset) or a new covariance matrix for comparison. The number of landmarks in all matrices must be the same (the program will check for this before allowing you to use an uploaded matrix).

The option to include or exclude the diagonal of the covariance matrix is provided. Keep in mind that mixing covariances and variances in a matrix correlation analysis can produce spuriously high correlations simply driven by the difference in magnitude between them—variances are usually higher than covariances.

10. **PCA** button: performs a Principal Component Analysis of the covariance matrix associated to each component of variation (symmetry, directional asymmetry, fluctuating asymmetry, and their respective error matrices). Results are presented graphically in the upper-panel figure (see 16) and effects can be differentially displayed using the controls below this figure (see 15).

11. **Run analyses** button: this button will be active only if at least one analysis is checked. You must click this button for the analyses to proceed.

12. **Save results to text file** button: this will save only those results from the last run of Sage. In addition to checked analyses, the saved file includes a decomposition of the variance of each landmark, which must is provided to aid in general interpretation of results but should not be interpreted as individual landmarks' contribution to total variation (see Klingenberg and McIntyre, 1998). In the case of matching asymmetry, it also includes the components of variance of the Size variable (labeled as landmark zero).

13. **Symmetrize dataset**: no analyses are required to be carried out in order to produce a symmetric dataset. Clicking on this button allows saving the symmetric component of the variation in a dataset, generally defined as the average of two sides, after Procrustes superimposition (= [Right side + Left side]/2). During superimposition, one of the sides is reflected to match the opposite one, and the result is a set of symmetric configurations. If using object symmetry, the resulting dataset will be of the same size as the initial one. Under matching symmetry, both sides are identical, and thus only one is included in the output. *This function allows to symmetrize datasets irrespective of the number of replicates per specimen in it*, and replicates are not averaged in this case (use ANOVA/MANOVA and File>Save fitted data to get symmetric averages over replicates).

14. **Residual asymmetry***: this function allows saving the individual (or replicate) residuals from the symmetrization process described in 13. Thus, resulting configurations contain all information about asymmetry in the original dataset. Residual asymmetry is obtained as [Right side – Left side]/2, and it is the same, up to a reflection, for both sides in the original configurations. Therefore, only one side is produced when using matching symmetry. The * in the button label indicates that the Procrustes mean is added to the residuals from symmetrization to get full configurations. This mean is not part of the residuals, but its inclusion does not alter the results of any further analyses, while allowing plotting these residuals as regular geometric data.

15. PCA controls: these controls allow modifying what is plotted in the upper-panel figure, if a PCA was performed in the last analysis run. Use the pull-down menus to change effects of PCs to show. The box indicates the percentage of variance accounted for each PC. Press the **Save scores** button to produce a file with specimen/replicate scores. These scores are produced by projecting every entry in dataset, i.e. all replicates, onto corresponding PCs, such that all score files will have the same number of rows as the original data file. See below note about specimen ordering for more information.

16. Upper-panel figure: this figure displays Procrustes-fit coordinates of current dataset upon loading it, and PCA results as TPS deformations.

17. Lower-panel figure: this figure displays right and left (or reflected and unreflected) datasets upon loading of protocols, as well as plots of covariance matrix entries after running matrix correlation analyses.

> **Note**: right-clicking on a white space in any graph will pop up a menu offering the option to save the current file as an image and, in the case of TPS graph, to modify the properties of the deformation display.

18. This window presents some hints about general functioning of Sage, as well as other messages.

Note: in the current version of Sage, the ordering of specimens present in the original dataset may not be preserved through the analysis. This may affect the interpretation of adjusted data as output. **To ensure preservation of ordering**, follow these two rules when constructing datasets: (1) sort specimens so that individual identification numbers are already sorted in ascending order before loading the corresponding protocol, and (2) always place replicates of the same individuals together.

The **File** menu contains the following options:

- **Save fitted data**: allows saving landmark data transformed to represent different components of the original landmark variation.
  - **Individual component**: symmetric component.
  - **Interaction component**: fluctuating asymmetry component.
  - **Global measurement error**: measurement error component, including both symmetric and asymmetric components of the error.
  - **Asymmetric** and **Symmetric components of error**.
  - **Right** and **Left side deviations**: these correspond to the deviations of each side (or reflection, in the case of object symmetry) of the data from the right and left means, respectively, after superimposing both sides together. The output configurations are produced by adding the Procrustes mean to both sides' deviations.
- **Save covariance matrices**: these are the matrices employed in Principal Component Analyses, arising from the decomposition implied by the Fluctuating Asymmetry design employed throughout the program.
- **Save vectorized covariances**: same as above, with covariances stacked in a vector.
- **Save options**: two sets of attributes of saved data files are currently customizable. First, Sage allows saving entire configurations or only half of them, which is useful since most of the outcomes of interest from object symmetry analysis contain the same information in both sides (exceptions are the global error component and the asymmetric error component). If saving only half configurations, the option is provided to include or exclude midline landmarks. These options are disabled in matching symmetry analyses.

  The second set of saving options allows saving data in IMP format, with XY columns of scaled landmarks followed by a centroid size (CS) column, or as unscaled XY columns with no CS column at the end.

  Both of these sets of options apply to saves from the File menu as well as those described in 13 and 14.
- **Save design info**: saves a text file with basic information about the current design (sample sizes, type of symmetry, degrees of freedom per effect).

Note: an early beta version of Sage for 3D data (Sage3D) is available upon request only. As of August 2006, Sage3D allows symmetrization of datasets, plus ANOVA and MANOVA analyses as described above for Sage. It does not include a superimposition routine, though, so a third-party client has to be used for superimposition. It does not have any graphics capability or perform matrix correlation analyses either.