The ‘`Genesis`’ Manual

Version 0.1

1 Description

This manual describes `Genesis`, a program created for scientists to generate PCA (Principal Component Analysis) and structure/admixture graphs from data outputted by common tools such as eigenstrat and the SNPRelate package for PCAs and Admixture and CLUMPP for admixtures.

Genesis was developed with user-friendliness in mind as other tools can be complex to use and lack certain features. All elements of the graphs that would need to be edited can be done so using a Graphical User Interface where the graphs themselves are interactive and different elements can be viewed and changed at the click of the mouse. All this saves the time that scientists would rather be spending doing more important things.

Principal Component Analysis is a mathematical and statistical procedure that is used to analyse genetic data. It takes in a set of genetic data and outputs values (called principal components) that can then be plotted onto a set of axes for visual analysis.

Structure/Admixture mappings are used to analyse populations of mixed-ancestry and determine the ratios of proposed different ancestrys. These ratios can then displayed in stacked bar graphs as structure/admixture graphs.

1.1 Licence

`Genesis` was written by Robert W Buchmann, and is released under the Affero General Public Licence version 3. `Genesis` uses the iText Software Corporations iText library also released under the Affero General Public Licence.

2 Installing and Running Genesis

2.1 Installing Genesis

2.1.1 Java

`Genesis` does not require installation and the file can simply be executed using Java. It does, however, require the Java SE Runtime Environment 1.7 (or higher) to be installed on the user's system. The latest code can be found at `http://www.bioinf.wits.ac.za/software/Genesis`

The latest version of the Java SE Runtime Environment 1.7 can be found at: `http://www.oracle.com/technetwork/java/javase/downloads/jre7-downloads-1880261.html`

To check if you already have Java installed, you can open your Operating System's command line interface and enter

```
java -version
```

If Java is installed, one line of the response should read

```
Java(TM) SE Runtime Environment (build 1.x.0_*)
```

where x is the version number (and should be 7 or higher) and the `*` is not important.

Alternatively, Windows users can open Control Panel and then open `Add/Remove Programs` and check the list for Java SE Runtime Environment (in windows XP and earlier) or open `Programs` then `Programs and Features` and if `Java` x is in the list, then Java is installed where x is the version number (in Windows Vista, 7 or 8).

Mac OSX requires X11 to be installed. This can be downloaded from `http://xquartz.macosforge.org/landing/`.

2.1.2 Compatible Operating Systems

`Genesis` is compatible with the 32 and 64 bit varieties of Windows, Mac OSX and Linux.

2.2 Running Genesis

2.2.1 Windows and Linux

If the Java SE Runtime Environment 1.7 is installed, the user should be able to open `Genesis` by double-clicking the file. If this does not work, the user will have to launch the jar manually through a command line editor (cmd in windows or terminal in linux) using the following command.

```
java -jar Genesis.jar
```

2.2.2 OSX

Ensure that the Java SE Runtime Environment 1.7 is installed. In OSX, the user must launch the file through the command line with an extra argument as follows

```
java -XstartOnFirstThread -jar Genesis.jar
```

3 Structure Plotsjava -XstartOnFirstThread -jar Genesis.jar

3.1 Inputting Data

Genesis produces structure charts by taking input from the Structure, Admixture or CLUMPP tools, together with a PLINK-style fam and optionally a phenotype file which would contain population labels.

To input these files, click `File -> New Admixture` or the `New Admixture` button on the toolbar. On the screen that opens, click `Import Data File` and navigate to the admixture data file outputted by the Admixture/CLUMPP tools. Then click `Import Fam File` and navigate to the fam data file. Finally, click `Import Pheno File` and navigate to the phenotype data file.

You can import multiple data files into the same project by clicking `Import Data File` again. These files can be imported in any order relative to the importing of the fam and pheno file.

In the drop-down menu select the column of the phenotype file that will be used to group the data. To draw the graph, click `Finish` or click `Next` to access the Appearance Options menu (See below...).

3.1.1 Editing Phenotype Column/Importing Additional Data Files

If a graph has already been plotted, and you wish to change the column of the phenotype data used to group the data or import additional data files, you can access the initial menu by clicking the `Data Options` button on the toolbar, or clicking `Graph -> Data Options`.

3.2 Appearance Options

The Appearance Options menu can be accessed through the `New Admixture` button by clicking `Next` after importing the files or by clicking the `Appearance Options` button on the toolbar, or clicking `Graph -> Appearance Options`.

3.2.1 Heading

To set/change the heading, open the `Appearance Options` menu and type the heading into the text box that says “Set Heading.” To change the font of the heading, click `Select Heading Font` and select the font of choice.

3.2.2 Border

To show/hide the border open the `Appearance Options` menu and (un)check the `Show Border` checkbox.

3.2.3 Population Group Labels

To show/hide the population group labels, open the `Appearance Options` menu and (un)check the `Show Population Group Labels` checkbox. To change the font of the population group labels, click `Select Population Group Label Font` and select the font of choice.

3.2.4 Graph Size and Spacing

To change the size and spacing of a structure plot, open the `Appearance Options` menu and select values for the `Set Graph Height`, `Set Thickness of Each Subject` and `Set Distance Between Graphs` boxes.

3.2.5 Horizontal/Vertical Orientation

To set the orientation of the structure plot, and select `Horizontal` or `Vertical` from the bottom drop down menu.

3.3 Interacting with the Graph

Other settings and options can be changed by interacting with the graph. Certain elements can be clicked to view or modify their options.

3.3.1 Selecting Subjects

To select an individual subject on the structure plot, click on the subject on the plot. This will bring up a subject menu where you can view the data about the subject (from the phenotype data file) and gives the option to hide the subject from the plot.

3.3.2 Population Group Name and Order

To select a population group, click on the population group label for the group. This will bring up a menu where you can change the Name of the group and also shift the group left/right (or up/down) in the graph.

3.3.3 Changing the Colours and Order of Colours

To change the colours and order of the colours in the graph, right click anywhere on the graph and select `Population Group Options`. From the next menu, click the button next to the colour you wish to change and from the next dialog, click the `Select Colour` button to change the colour or the `Shift Up` and `Shift Down` buttons (or the `Shift Left` and `Shift Right` buttons) to change the order of the colours.

3.4 Annotating the Graph

3.4.1 Creating a Label

To create a label and annotate the graph, right click on the graph where the label is to be placed, and click `Create Label at Mouse Pointer`. Then enter the label's text and click `OK` to place the label.

3.4.2 Editing or Deleting a Label

To edit a label that has been created, click on the label to bring up the Label dialog. From this dialog you can edit the label's text, reposition the label or delete the label.

3.5 Hiding Subjects from the Graph

3.5.1 Hiding Subjects

To hide a subject from the graph, click on the subject on the graph to bring up the subject dialog. From this menu you can choose to hide the subject from the graph.

3.5.2 Hiding Groups

To hide a population group from the graph, click on the group in the key to bring up the population group dialog and check the box labelled `Hide this group from the graph`.

3.5.3 Showing Hidden Subjects and groups

To reshow a subject that has been hidden, click on the `Select Hidden Individuals or Groups` button and from the drop down menu that pops up, select the name of the subject or group that you wish to show. From the subject dialog you can now uncheck the `Hide this Individual from the Graph` or `Hide this group from the Graph` checkbox.

3.6 Searching for a Subject

To search for a subject in the graph by name, click the `Search for individual` button in the toolbar. In the dialog, enter the Name (first, last or both) of the individual you wish to find and click `Ok`. If the individual was found in the data, it will be selected and the subject dialog for that individual will open. If the individual was not found, a message will display.

4 PCA Plots

4.1 Inputting Data

To input Eigenstrat files, click `File -> New PCA` or the `New PCA` button on the toolbar. On the screen that opens, click `Import Data File` and navigate to the PCA data file outputted by the Eigenstrat software. Then click `Import Phenotype File` and navigate to the phenotype data file.

To input Relate Data, click `File -> New PCA` or the `New PCA` button on the toolbar. On the screen that opens, click `Import Data File` and navigate to the PCA data file outputted by the Relate package. The relate package file includes the phenotype information in the data file.

In the drop-down menus, select the 2 or 3 PCAs to plot as the axes and select the column of the phenotype file that will be used to group the data. To draw the graph, click `Finish` or click `Next` to access the Appearance Options menu (See below...).

4.1.1 Editing PCAs/Phenotype Column

If a graph has already been plotted, and you wish to change the PCAs to be the axes or the column of the phenotype data to represent the data, you can access the initial menu by clicking the `Data Options` button on the toolbar, or clicking `Graph -> Data Options`.

4.2 Appearance Options

The Appearance Options menu can be accessed through the `New PCA` button by clicking `Next` after importing the files or by clicking the `Appearance Options` button on the toolbar, or clicking `Graph -> Appearance Options`.

4.2.1 Heading

To set/change the heading, open the `Appearance Options` menu and type the heading into the text box that says “Set Heading.” To change the font of the heading, click `Select Heading Font` and select the font of choice.

4.2.2 Border

To show/hide the border open the `Appearance Options` menu and (un)check the `Show Border` checkbox.

4.2.3 Axes

To show/hide the axes, axis labels, grid and scale, open the `Appearance Options` menu and (un)check the relevant checkboxes.

4.2.4 Key Options

To set the position of the key, open the `Appearance Options` menu and select the key position from the drop down menu. To hide the key, select `No Key` from the drop down menu. To change the font of the key, click `Select Key Font` and select the font of choice.

Other settings and options can be changed by interacting with the graph. Certain elements can be clicked to view or modify their options.

4.2.5 Selecting Subjects

To select an individual subject on the structure plot, click on the subject on the plot. This will bring up a subject menu where you can view the data about the subject (from the phenotype data file) and gives options to change the subject's icon as well as to delete the subject and hide or place the subject on top.

4.2.6 Changing Subject's Icon

To change an individual subject's icon, click the subject on the graph and from the dialog that opens, select the icon shape and colour. Selecting the shape `Default` will set the shape of the icon to the group's shape and checking `Clear all icon data specific to this individual` will set the icon's shape and colour to the group's.

4.2.7 Placing a Subject on Top

To place an individual subject on top, click the subject on the graph and from the dialog that opens, check the `Place this individual on top` checkbox and click `Done`.

4.2.8 Changing the Population Group Names

To change a population group's name, click on the group in the key. This will bring up the Population Group dialog. From here you can set the group name and click `Done`.

4.2.9 Changing the Population Group Icons

To change a population group's icon, click on the group in the key. This will bring up the Population Group dialog. From here you can set the shape and colour of the icon and click `Done`.

4.2.10 Changing the Order of Population Groups in the Key

To change a the order of the population groups in the key, click on the group in the key. This will bring up the Population Group dialog. From here you can click `Shift Up` or `Shift Down` shift the group's order in the key.

4.3 Annotating the Graph

4.3.1 Creating a Label

To create a label and annotate the graph, right click on the graph where the label is to be placed, and click `Create Label at Mouse Pointer`. Then enter the label's text and click `OK` to place the label.

4.3.2 Editing or Deleting a Label

To edit a label that has been created, click on the label to bring up the Label dialog. From this dialog you can edit the label's text, reposition the label or delete the label.

4.4 Hiding Subjects and Population Groups from the Graph

4.4.1 Hiding Subjects

To hide a subject from the graph, click on the subject on the graph to bring up the subject dialog. From this menu you can choose to hide the subject from the graph.

4.4.2 Hiding Groups

To hide a population group from the graph, click on the group in the key to bring up the population group dialog and check the box labelled `Hide this group from the graph`.

4.4.3 Showing Hidden Subjects and Groups

To reshow a subject that has been hidden, click on the `Select Hidden Individuals or Groups` button and from the drop down menu that pops up, select the name of the subject or group that you wish to show. From the subject dialog you can now uncheck the `Hide this Individual from the Graph` or `Hide this group from the Graph` checkbox.

4.5 Searching for a Subject

To search for a subject in the graph by name, click the `Search for individual` button in the toolbar. In the dialog, enter the Name (first, last or both) of the individual you wish to find and click `Ok`. If the individual was found in the data, it will be selected and the subject dialog for that individual will open. If the individual was not found, a message will display.

4.6 Rotating the Graph

To rotate a 3D PCA plot, click the `Show/Hide 3D PCA Rotate Panel` button in the toolbar. This will bring up the rotate panel which contains a slider which can be dragged to rotate the graph about the z-axis.

5 Saving, Loading and Exporting Graphs

5.1 Saving a Project

To save a project currently in use, click the `Save Project` button on the toolbar or click `File -> Save Project`. Enter a filename and navigate to the location you wish to save the file and click `Save`.

5.2 Loading a Project

To load a project previously saved, click the `Load Project` button on the toolbar or click `File -> Load Project`. Navigate to and select the file you wish to load and click `Load`.

5.3 Exporting Graphs

Once you are satisfied with the graph you are currently working on, you can export the graph as an image file (.png) or in a PDF. To do so, click the `Export` button on the toolbar or click `File -> Export`.

From the dialog that opens, select whether to export as a PNG image or as a PDF, then navigate to the location you wish to save the file and click `OK`.

5.3.1 Notes on exporting as a PDF

Genesis will export the graph as it is on the screen in front of you, and will scale this image to fit on the width of an A4 page. This may cause larger admixture project's images to be of too high detail to view all the detail on the monitor of your computer all at once. This is normal and expected. The detail will be stored there - it may be necessary to zoom in to view it all - and if the image is printed, the detail will be there (assuming the printer is of high enough quality).

5.3.2 Notes on exporting as a PNG image

PNG (Portable Network Graphics) is a free open-source cross-platform image file that can be handled by almost all good image viewers and editors. When exporting as a PNG file, note that the image will be the same size as it is on the screen as you export it. This means that resizing the Genesis window between PCA exports will create different sized images (Admixture sizes are determined by the options in the `Appearance Options`). This is important as PNG images do not scale up, i.e. a smaller image will become blurry if it is stretched too big. Keep this in mind when exporting as a PNG image.

6 Conclusion

Any technical questions on the use of `Genesis` should be sent to `genesis@bioinf.wits.ac.za`.