GeneLink User Manual Draft

Greg O'Brien, Robert Shields[2], Natalie Maricic[2], Robert A. Burne[2]

July 7, 2018

Contents

1   Getting Started

1.1   What is GeneLink

GeneLink, is a software tool for analyzing growth curves and clustering similar experiments together. Made available to the end-user, a large assortment of clustering algorithms are automatically combined into a robust, agglomerative, hierarchal result. This approach takes the guess-work out of clustering. Maybe include a sentence here about what Genelink can and cant do.

1.2   Installing

Be sure that the most recent version of GeneLink is installed. Users may download the latest version at github.com/mygregstuff/GeneLink.

1.3   Importing Data

Upon launching GeneLink, users are able to select data files containing optical density, time and strain names for each set of optical density measurements over time. Data files may be .xls, .xlsx, .cvs or plain text formats. Multiple data files may be selected at once for comparison of larger data sets. Figures are saved to the same folder as the data file and named accordingly.

Table 1: Example Data Table : Sheet 1

...	...	...	...	...	...	...
...	time	blank	strain name 1	...	strain name 2	...
...	10	0	.1	...	.1	...
...	20	0	.1	...	.2	...
...	30	0	.2	...	.2	...

1.4 Combing Data

When multiple data files are imported for each run, all data should be formatted similarly, such as having similar measurement intervals for the same length of time. In general, the shorter the optical density measurement intervals, the better.

2 Settings and Optional Features

2.1 The Settings File

The settings file is a text file that lets users customize GeneLink to their preferences. The settings file, commented by '%', has included descriptions and optional values for each setting.

2.2 Location Dependencies and Default Settings

The settings file should be in the same folder as the imported data files, or default settings will be implemented.

2.3 Import Options

3 optional methods exist for importing data. The user may select strain names, observation times, and optical density data for each xls or xlsx data file (as default). For importing multiple files, users may declare locations for names, time and optical densities in the Settings.txt file, to avoid being prompted for every single file. The third method (auto) allows GeneLink to automatically chose the correct data by recognizing the words "time" and "blank" within the file. Using the third method requires data files to formatted as such:

2.4 Inspecting Fits and Data

Its a VERY good ideal to inspect a few data points GeneLink uses, to see if they are in agreement with expected values. Figures of individual fits may be produce by the boolean settings option ¨fit_figures¨in the settings file, which is off by default.

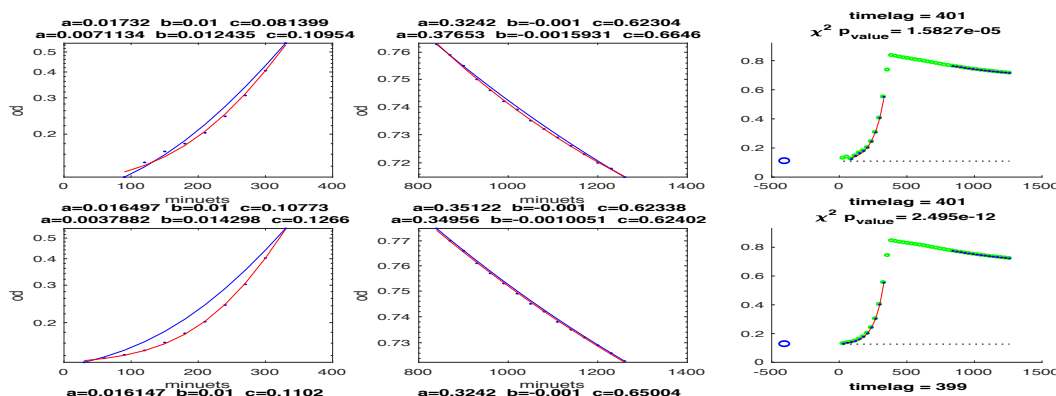**Biological Replicate of Cultured Bacteria**

Figure 1: Fits of 2 biological replicates. An example output of GeneLink. Experimental variability causes no two estimates to be the same, demonstrating the need of a robust analysis method for biological replicates.

2.5 Figure Saving Options

Figures may be saved as .pdf (by default), .jpeg or .png. PDFs produce the best resolution for publishing and scale with near zero resolution loss. Users my also set the resolution option in the settings file by declaring different DPI values.

3 Trouble Shooting

3.1 Installing

A general rule of thumb is to have everything up-to-date on your computer. Currently GeneLink is only available for windows.

4

3.2 Opening

GeneLink may be slow opening on computers with disk drives and appeared to have crashed. If you receive an error message, try the standard procedure of reinstalling GeneLink and then restarting your computer.

3.3 Importing

GeneLink sometimes crashes. The most often cause is when GeneLink encounters a text character it is unsure what to do with. Make sure all optical density and time data are numerical and contains no strings. This is best done by using the manual import method in section 2.3

3.4 Crashing

GeneLink sometimes crashes. The most often cause is when GeneLink encounters a text character it is unsure what to do with. Make sure all optical density and time data are numerical and contains no strings. This is best done by using the manual import method detailed in the 'Importing' section

3.5 Slowness

GeneLink can be slow to open, sometimes taking up to 10 minuets to initialize depending on you computers hard drive speed. GeneLink may be additionally slowed down by producing figures for every fit or a large number of bootstrap iterations.

3.6 Non informative Figures

Poor Figures with most samples clustered into a single group gives little information to the user. Poor quality figures can arise for a number of reason, many which can be addressed. Try importing the data manually and make sure all data is being used. Be sure enough time points exist to produce accurate fits. This may be determined by examining fits of figures. Additionally not enough boot strap iterations will produce poor chi-squared statistics and

lead to bad estimations. Some bacterial cultures may not also have grown, and can be discarded by placing lower and upper bounds on the carrying capacity. GeneLink, as well, examines p-values to decide which variables are useful. Adjusting the p-value cutoff will reduce the number of variables and will provide better contrast in clustering, but at a lower resolution. Finally users may adjust the sensitivity of how items are clustered across different algorithms. A higher sensitivity will weight better cluster results higher, and lower f-scores cluster results lower. This produces a similar effect of better contrast but lower resolution in pictures. A final word of caution to the user about over fitting the data to produce a desired result. You may often find supporting evidence for a hypothesis with a unique set of settings. However, if a result only shows up under a specific set of options, and not under others, the results may be non robust and statistically non-significant. Secondly, the default values in the settings are known to work and should only be substantially adjusted far form the default value only if there is a specific reason.

3.7 Saving Figures

Poor resolution and figures that will not open may be caused by incapable figure options. PDF (the option used by default) usually works best.

4 Additional Resources

For additional information users may refer to several sources. The paper, "Genelink" provides a brief over view of GeneLink. The supplemental information file, available on the GeneLink Github repository, contains background information about growth dynamics, fitting, clustering and graphics. Mathworks clustering page /citematlabcus, an excellent source provides detailed knowledge on how clustering works.

4.1 Personal Contact Information

Further questions can by answered by contacting Greg O'Brien at greg999@ufl.edu or Rob Shields at rsheilds@ufl.edu.

5   Acknowledgments

References