

script manual: **bar_plotter.rb**

Summary

bar_plotter.rb is a tool to create ready to publish vector graphic barplots from output of the software STRUCTURE (Pritchard, Stephens, and Donelli 2000).

I wrote it for two reasons:

1. The Bayesian clustering algorithm implemented in STRUCTURE is one of the most widely used methods for population assignment based on microsatellite data. Its popularity may be due to its high quality analysis but also to the ease of use of STRUCTURE's graphical user interface (GUI). STRUCTURE output is commonly presented as barplot graphs and triangle plots, which often find their way into high-ranking journals. Unfortunately however, both graph types are exportable from STRUCTURE's GUI version in jpeg format only. Authors regularly find themselves redrawing the graphs in image editing software such as Adobe Photoshop or Illustrator, to change the graph's colors according to color codes used in other graphs, to change the graph's height/width ratio, to improve the clearness of the illustration, or simply for esthetic reasons.
2. Apparently the GUI version of STRUCTURE 2.2 has conflicts with the new MacOS X operating system, Leopard. Or the problem is something else instead, in any case I found STRUCTURE's GUI crashing quite often recently. Either runs don't really start at all or stop right after the burn-in period. The command line version of STRUCTURE does not seem to have this problem. But of course the command line version does not produce barplots...

Method

bar_plotter.rb reads STRUCTURE output files. It extracts the table containing the inferred ancestry of individuals and stores all these values in a matrix. It prepares output in SVG format, where bar sizes and positions are calculated from these matrix values. Color coding of the individual clusters can be chosen using a separate color list file. If no color list is specified, greyscale colors are used. A few other parameters can be chosen (see parameter list).

Input file

The output files of STRUCTURE's command line version can be directly read by bar_plotter.rb. If you're using STRUCTURE's GUI version, you'll find the right files in a subfolder of STRUCTURE's output folder (something like 'standard > Results'), and they are named 'standard_run_1_f'. LF, CR, or CRLF line endings can be read, so don't worry about Windows vs. Mac vs. Mac Classic problems.

Output

The output is written to a separate file with the same name as the input file, but with the extension removed (if there was any) and with "_barplot.svg" added instead. This file will be located in the same folder as the input file. If a file with an identical name already exists in this folder, it will be overwritten. The bar_plotter.rb output file has LF line endings, which should be readable by Adobe Illustrator on any operating system.

How to run the script

bar_plotter.rb is written in Ruby and thus needs Ruby installed on your system. By default, this is the case on Mac OS X, but probably not on any Windows computer.

On Macs, do the following: Copy bar_plotter.rb into the folder, where your input file is. Then open the Terminal and change into this directory by typing 'cd ', following by the path to the directory. Hit enter and see whether you really are in the right directory by typing 'pwd' (print working directory) or 'ls' to see all the files in that folder. Then execute bar_plotter.rb by typing '**ruby bar_plotter.rb -i inputfilename**', where "inputfilename", naturally, is the name of your input file. Hit enter, and your output file should be written within milliseconds.

Parameter list

The default values should be fine if you need greyscale images for publication anyway. If not, use the -c option and prepare a separate color list file. Hex codes for colors can be determined with the Adobe Photoshop colorpicker. Alternatively google for "Colorpicker" and you'll find lots of tools.

-c specifies that a color list file is provided. This parameter must be followed by the name of the file, otherwise the script won't execute. The color list file should be in the same directory as the script and the input file, and just as the input file, it may have LF, CR, or CRLF line endings. bar_plotter.rb expects the color list file to contain nothing but the colors, given either by name (e.g. "red") or by hex code (e.g. "ff0000"), and listed below each other. If more colors are specified than clusters in the STRUCTURE analysis, left-over colors will be ignored. If less colors are specified, the list will be completed automatically with greyscale values. An example file is provided at www.evolution.unibas.ch/salzburger/software.

-g causes a semi-transparent gradient to be applied on top of the barplot, which has none but esthetic value. The default gradient color is gray, but if -g is followed by a color name or a hex code, this color will be used (e.g. -g red or -g ff0000)

-t produces vertical ticks at the top and bottom of the barplot between every individual. The default tick length is 3 pixels, but this can be changed by adding an integer after -t (e.g. -t 5). Note that this may increase the output file size remarkably.

-v specifies that vertical lines should be added between every individual. In order not to confuse this with the lines between populations (which are black), the default line color is gray. This can be changed by adding a color name or hex code after -v (e.g. -v red). Personally, I don't think this option is useful, and it may increase output file size if you have many individuals.

References

Pritchard, J. K., M. Stephens, and P. Donnelly (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155 (2):945-959

