This file describes the usage of GenoMap.

Copyright of this document and the GenoMap script:
Naoki Sato, October 29, 2002.
Revised, December 4, 2002.

Essentially identical description is provided with the GenoMap script.

A. General notes

The file 'Genomap.tk' is a Tcl/Tk script that allows easy visualization of gene content, either the arrangement of genes, expression pattern of genes, GC content etc. GenoMap needs only a single file describing the location of CDS, RNA genes, and various sequence tags related to the expression of genes obtained from microarray experiments, or sequence feature information such as GC content etc. Upto four different kinds of TAG data are displayed. These data are displayed in a circular map with various cirlular bands. The size of the entire circle, the width of the circular bands, and the color of data representation are customizable by selecting various values from the menus. In addition, the data are displayed in one of the three ways, namely, linear, log, and negative log. An offset value can be used for each TAG to show clearer differences of values. By default, the plot is scaled to fit the highest value within the width of the circular band. However, the scale of the plot may be given.

Tcl/Tk can handle text files with any types of end of line codes. However, we sometimes need to convert end-of-line code for text files imported from different platforms. To facilitate handling of these imported or exported files, a command 'txtr' is implemented to convert end-of-line codes in any manner.

This script is expected to run with Tcl/Tk version 8.1 or higher, although it was mostly tested with version 8.3 and 8.4. I recommend using Linux or MacOS X to run this script. On MacOS X, the Aqua version of Wish has some problems in showing names of buttons, but the Tcl/Tk compiled on X window system (XFree86 4.2) is perfect.

Usage of various settings is explained briefly in the command list window. This window is shown if the 'Command list' button is pressed.

By default, the application is intended to be run on Unix. However, in most cases, the correct OS name is set automatically. Alternatively, OS name can be defined in the first part of the script.

I found that the data are not correctly displayed in the circular map with the Windows version of Tcl/Tk. This is probably due to a bug (incompatibility of Windows
system library and Tcl library?) in the drawing of very narrow arc in the canvas widget. I have fixed tentatively this problem, and the GeneMap can now be used on Windows.

I took care of various trivial errors occurring with improper use of this software, but in some cases, error is displayed on the Tcl/Tk window. In such cases, you just need to click on the OK button of the error message window. If you still have problems, just put the 'exit' button to close the GenoMap window.

This file is available from the window of GenoMap.
Press 'show description' button.

Any comments are welcome.

Contact address:

Naoki Sato
Department of Molecular Biology
Faculty of Science
Saitama University
Shimo-Ohkubo 255
Saitama 338-8570 Japan
naokisat@molbiol.saitama-u.ac.jp

B. Notes for the usage

1. The extended GRS file format

The GRS file is a tab-delimited text file, either generated by computer, or converted from EXCEL file. An example file is shown below. The lines 'Organism' and 'Size' are necessary. Other lines in the header are not used in the program.

The GRS format was used in the software called GRS by T. Sicheritz. In this application, we extend the file format to include expression data or other quantitative values. The data type of these data are described as 'TAG', and three different kinds of TAGs are allowed in the circular map. This can be extended, but there is a limit of space for such circular map, and we suppose three TAGs are a good compromise.

In this file, empty lines are allowed. Each line represents a single data entry, and include items that are arranged with a tab as a delimiter. The data entries can be placed in any order, i.e., the lines with CDS, TAG and RNA can be mixed. But this might not be very convenient for most purposes.

New feature: In a new version, color can be specified in each TAG entries. This is useful when the data with non-significant difference (judged by t-test or other tests) are displayed without highlight. The color is optional. If the color is not specified, the color settings of the GenoMap is used.
See the file 'FileFormat.pdf' for explanation.

#GTF
Organism: Nostoc sp. BA000019
Type: type_unknown
Size: 6413771
Contigs: 0

definition of format:

name	type	orient	start	stop	length	description	color
all0002	CDS	R	1718	981	737	some gene 1	
asl0003	CDS	R	2805	2617	188	some gene ABC	
atpC	CDS	R	4365	3418	947	ATP synthase subunit	
chlP	CDS	F	131490	132710	1220	geranylgeranyl hydrogenase	
rpaA	CDS	R	133777	133034	743	two-component response regulator	
trnL	RNA	R	179948	179865	83	tRNA-Leu	
trnL	RNA	F	185099	185180	81	tRNA-Leu	
rrn5Sa	RNA	F	238209	238221	118	5S ribosomal RNA	
1	TAG	1	1	2565	2565	1.288	
2	TAG	1	1613	5067	3455	0.723	
3	TAG	1	4475	7618	3144	0.9825	
230	TAG	2	603266	606407	3142	1.013	gray
231	TAG	2	606213	609389	3177	0.923	gray
232	TAG	2	607164	610508	3345	1.199	

----

2. How to generate GRS file?

The genome database file in the GenBank or EMBL format can be converted to the GRS format by the SISEQ package. This software has been described in the following article.

Sato, N. (2000) SISEQ: Manipulation of multiple sequence and large database files for common platforms. Bioinformatics 16, 180-181.

Use the following command to convert GenBank or EMBL file.

siseq genlist <infile> <outfile> t

The 't' option is to generate GRS table. <infile> is the name of the GenBank or EMBL file. These file formats are recognized automatically. <outfile> is the name of the GRS file to be generated. Note that SISEQ can be used on UNIX, Windows and (conventional) MacOS. It is also useful on the UNIX environment of MacOS X.

The 'TAG' data are normally edited as an EXCEL file, and finally converted to a tab-delimited file. Finally, the CDS and RNA data and TAG data are combined to form a single large file in the extended GRS format.

3. How to save and process images created by GenoMap?

Press the 'save PostScript' button located under the map window. Before pressing this button, you should specify the name of the output file in the main panel. By default, the output file is set to <infile>.ps, but you can change the name as you like. After some seconds or tens of seconds, a message saying 'Postscript image saved to file: ********.ps' will appear in the main panel. If you have not specified the output file, a message saying 'Specify te name of output file first.' is displayed in the main panel. Then, you can set the output file, and press 'Save PostScript' button, withou closing the map window.
  The generated PostScript file can be viewed by any PostScript viewer, such as ghostscript or GIMP. However, a large image file (by setting the diameter of the circle to as large as 6000 pixels!) is not rapidly displayed by these viewers. I recomment using Photoshop version 6 to rasterize the Post- Script image. The GIMP is also rapid on Silicon Graphics machines such as O2. Then, the image is converted to JPEG or other types of files to edit or decorate manually.

4. Configuration file

The settings of various variables such as colors and dimension of various objects can be saved in a file genomap.cf by selecting the submenu 'Save settings' from the 'SETTING' pull down menu. The name of this file can be changed by editing the genomap.tk script. If genomap.cf file is present in the current directory, the settings are automatically loaded during the initialization of the program. To avoid this and to use the initial settings indicated in the genomap.tk script, just remove or rename the genomap.cf file.

5. Japanese menu

The menu can be shown in Japanese if the language variable is set 'Japanese'. This should not be done with OS that is not compatible with Japanese language, including X window in MacOS X. The buttons are not Japanized in any case.

C. List of commands and settings

This list may be displayed by pressing the 'Command list' button.

```
-------------------------------------------------------------------------------
```
Setting environmental variables

1. Window size   Setting the size of actual square window.
2. Molecular form   Form of DNA, only circular is enabled now.
3. Canvas size   Size of the drawing canvas in square shape.
4. Diameter of outermost circle
5. Width of circular band
6. Language   Language in menubar.
7. Show data type   Select data to draw: CDS, RNA and/or TAG.

Setting colors

21. Background color   Background of the drawing canvas.
22. Line color   Color of the lines.
23. Forward ORF color Color of the ORFs in the forward direction.
24. Reverse ORF color Color of the ORFs in the reverse direction.
25. Forward ORF filling color   Color of the interior of the ORF boxes
                                in the forward direction.
26. Reverse ORF filling color   Color of the interior of the ORF boxes
                                in the reverse direction.

Setting TAG variables

41. TAG Background color   Background color for the TAG field
42. Color of TAG1
43. Color of TAG2
44. Coor of TAG3
45. Coor of TAG4
46. Width of TAG band Width (or the maximum hight of the data bars)
                               of the TAG field
47. Data conversion of TAG1   Linear or conversion to log or negative log
48. Data conversion of TAG2   Similar to above
49. Data conversion of TAG3   Similar to above
50. Data conversion of TAG4   Similar to above
51. Offset for TAG1   Offset value for TAG1. This is subtracted from the data
                The scale of the plot can be given after the offset value with a comma
                as a delimiter. No space is necessary before and after the comma.
52. Offset for TAG2   Similar to above
53. Offset for TAG3   Similar to above
54. Offset for TAG4   Similar to above
55. Font
56. Size of the text font
57. Interval of text lines   Maximum value of the TAG values is shown on the
                               upper left corner. These defines the format of
                               this description

100. Txtr   A command to change the end-of-line code of text files.
        This is normally not necessary for reading the data files into this
        software. But this is useful in importing from or exporting to other
        platforms text files. The MacOSX version of EXCEL generates
        a tab-delimited text file with Windows style end-of-line code.

```
-------------------------------------------------------------------------------
```