

# R Graph Automation Documentation



Status

In progress



This Guide is written with the assumption that you already have R studio installed.

## 1. R settings

- a. Click the files button on the right hand side of the screen
- b. Find the Gears button just below the files bottom on the right side of the screen
- c. Click "Set As Working Directory"
- d. Click on the R script
- e. Change the name of the file to your desired file
- f. Use Command+Shift+Enter (or Control+Shift+Enter on Windows) to run the entire file
- g. Click Export to export your plot.

The screenshot shows the RStudio interface. On the left, a data table with 10 rows and 5 columns is displayed. The columns are Gene.Type, Gene\_name, Statistical\_Significance, Background, and Foreground. The data is as follows:

	Gene.Type	Gene_name	Statistical_Significance	Background	Foreground
1	Both	CNGA11	No	1.328684	1.92
2	Both	CNGA20	Yes	1.370099	1.85
3	Cone	CNGA11	No	1.124724	1.91
4	Cone	CNGA12	Yes	1.740157	1.97
5	Cone	CNGA19	No	1.456275	1.01
6	Cone	CNGA7	No	1.264260	1.15
7	Cone	CNGA8	Yes	1.433880	1.49
8	Rod	CNGA1	Yes	1.400218	1.37
9	Rod	CNGA2	No	1.369840	1.60
10	Rod	CNGA4	Yes	1.235931	1.55

On the right, the 'Files' pane shows the project structure. A red arrow points to the 'Files' tab, and another red arrow points to the 'Gear Button' (the settings icon) in the top right corner of the Files pane.

The screenshot shows the RStudio interface with R code on the left and a plot on the right. The R code is as follows:

```

1 ## Plots clade model dN/dS estimates by gene
2 ## Last updated 20240607 by Chong
3
4 library(tidyverse)
5
6 dnds <- read.csv("Sample_Gene_Bank - Sheet1.csv")
7
8 # Transforms to wide format for plotting lines
9 dnds_line <- spread(dnds, key = Background_Foreground, value = dNdS)
10
11
12 ## Plots the dN/dS values for each clade by gene and grouped by photoreceptor
13 graph <- ggplot() +
14
15 # Overlay a segment graph as a background to represent significance
16 geom_segment(
17   data = dnds_line,
18   aes(x = rep(-Inf, length(Gene_name)),
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```

A red arrow points to the file name "Sample\_Gene\_Bank - Sheet1.csv" in the code, with the text "change to your file's name" written next to it. On the right, a plot is shown. The plot displays dN/dS values for various genes (CNGA8, CNGA7, CNGA4, CNGA20, CNGA2, CNGA19, CNGA12, CNGA11, CNGA1) grouped by photoreceptor type (Background and Foreground). A red arrow points to the 'Export' button in the top right corner of the plot, with the text "export your plot here" written next to it.

## 2. CSV/Excel Formatting

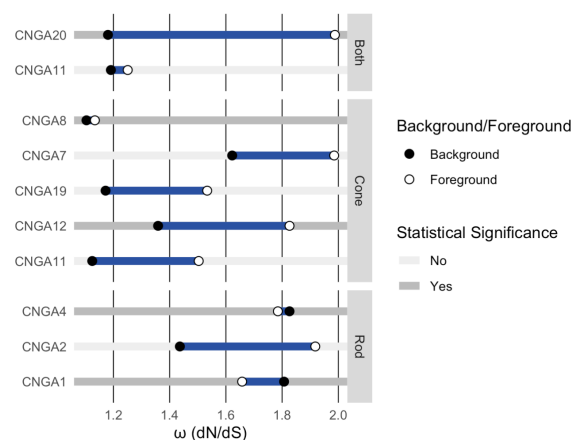
- Columns (Ordered doesn't matter, R is just picky about names):  
Gene\_type, Gene\_name, Background\_Foreground, dNdS,

## Statistical\_Significance

- Gene Type: Please enter the name of the string, or enter "na"
- Gene\_name: Please enter the name of the gene
- Background\_Foreground: Please enter either "Background" or "Foreground"
- Statistical\_Significance: Please enter either "Yes" or "No"
- dNdS: Numeric Value

### 3. Sample data with Gene Types

Gene Type	Gene_name	dNdS	Statistical_Significance	Background_Foreground
Rod	CNGA1	1.420898271	Yes	Background
Rod	CNGA1	1.399120233	Yes	Foreground
Rod	CNGA2	1.798042993	No	Background
Rod	CNGA2	1.287029638	No	Foreground
Rod	CNGA4	1.538076909	Yes	Background
Rod	CNGA4	1.736751058	Yes	Foreground
Cone	CNGA7	1.872901679	No	Background
Cone	CNGA7	1.862169926	No	Foreground
Cone	CNGA8	1.311531231	Yes	Background
Cone	CNGA8	1.0391834	Yes	Foreground
Cone	CNGA11	1.869341625	No	Background
Cone	CNGA11	1.010740569	No	Foreground
Cone	CNGA12	1.201309048	Yes	Background
Cone	CNGA12	1.089178849	Yes	Foreground
Cone	CNGA19	1.748352827	No	Background
Cone	CNGA19	1.752109061	No	Foreground
Both	CNGA20	1.90966686	Yes	Background
Both	CNGA20	1.466176017	Yes	Foreground
Both	CNGA11	1.222447703	No	Background
Both	CNGA11	1.691968101	No	Foreground



### 4. Sample data without Gene Types

Gene Type	Gene_name	dNdS	Statistical_Significance	Background_Foreground
na	CNGA1	1.672654992	Yes	Background
na	CNGA1	1.712828314	Yes	Foreground
na	CNGA2	1.063048757	No	Background
na	CNGA2	1.302023844	No	Foreground
na	CNGA4	1.401174259	Yes	Background
na	CNGA4	1.068695445	Yes	Foreground
na	CNGA7	1.740858486	No	Background
na	CNGA7	1.160488964	No	Foreground
na	CNGA8	1.492000778	Yes	Background
na	CNGA8	1.291456166	Yes	Foreground
na	CNGA11	1.258751012	No	Background
na	CNGA11	1.406918501	No	Foreground
na	CNGA12	1.293412095	Yes	Background
na	CNGA12	1.472288487	Yes	Foreground
na	CNGA19	1.265750147	No	Background
na	CNGA19	1.542183529	No	Foreground
na	CNGA20	1.48079509	Yes	Background
na	CNGA20	1.275036697	Yes	Foreground
na	CNGA11	1.314555225	No	Background
na	CNGA11	1.907426764	No	Foreground

