

Tutorial: SeqAPASS Boxplot Generator

1. Access SeqAPASS by opening <https://seqapass.epa.gov/seqapass/> using Mozilla Firefox web browser
2. Open the 'About' link on the login page or upon logging in to SeqAPASS, click on the 'Home' tab and then the 'About' link. At the bottom of the 'About' page, under the 'SeqAPASS Help' section, there are links to the Data Visualization components, including a **RData** link. By clicking on the 'RData' link an R workspace can be opened and saved to an easily accessible location on your personal computer (e.g., a folder on your Desktop).
3. Use the SeqAPASS application to generate a .csv for your query sequence. You will see the SeqAPASS data output on the 'View SeqAPASS Reports' Page. This R-based visualization tool handles both the Level 1 (primary amino acid sequence) and Level 2 (functional domain) datasets in either the Primary or Full Report formats. To download the report, click on the radio button associated with the query protein of interest and then click 'Request Selected Report.'

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) [Log out](#)

Home Request SeqAPASS Run SeqAPASS Run Status View SeqAPASS Reports Settings

SeqAPASS Reports Version 1.0 Logged in as: lalone.carlie@epa.gov

Main Level 1

Partial Protein Sequence Request Selected Report Refresh Available Reports

Available Reports

Search: Enter keyword

SeqAPASS Run Id	NCBI Accession	Query Protein Name	NCBI Taxonomy ID	Query Species Name	Query Common Name	Taxonomy	SeqAPASS
6	CAD01096.1	5-enolpyruvylshikimate-3-phosphate synthase	29674	Eleusine indica	goosegrass	Liliopsida	2016 03
5	NP_001171681.1	thyroid peroxidase	9823	Sus scrofa	pig	Mammalia	2016 03
4	AJE70269.1	nicotinic acetylcholine receptor beta2 subunit	7460	Apis mellifera	honey bee	Insecta	2016 03
3	ABB54707.1	cytochrome b	27349	Puccinia sorghi	basidiomycetes	Pucciniomycetes	2016 03
2	CAA70212.1	ecdysteroid receptor	7102	Heliothis virescens	tobacco budworm	Insecta	2016 03
1	NP_001091690.1	nicotinic acetylcholine receptor alpha1 subunit	7460	Apis mellifera	honey bee	Insecta	2016 03
1	NP_001011625.1	nicotinic acetylcholine receptor alpha2 subunit precursor	7460	Apis mellifera	honey bee	Insecta	2016 03
1	NP_001073029.1	nicotinic acetylcholine receptor alpha3 subunit precursor	7460	Apis mellifera	honey bee	Insecta	2016 03
1	NP_001091691.1	nicotinic acetylcholine receptor alpha4 subunit precursor	7460	Apis mellifera	honey bee	Insecta	2016 03
1	AAJ87893.1	nicotinic acetylcholine receptor alpha4 subunit	7460	Apis mellifera	honey bee	Insecta	2016 03

(1 of 3) 1 2 3 10 Download Table:

Top of Page

The Level 1 data will then open in a different page and can be downloaded by clicking on the csv icon on the lower right-hand side of the data table (for detailed instructions, see the SeqAPASS User Guide). You can download the .csv file to a location of your choice on your personal computer and name it as you choose.

Primary Report Partial Hit Protein Sequence Percent Similarity > 100% Responsible = N. Ontology Column = 4

Level 1 Data - Primary

Search: Enter keyword

NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Scientific Name	Common Name	Protein Name	BLAST BitScore
NP_001091690.1	26879	7460	Insecta	Apis mellifera	honey bee	nicotinic acetylcholine receptor alpha1 subunit	1233.
XP_012348317.1	10105	7463	Insecta	Apis florea	little honeybee	PREDICTED_ acetylcholine receptor subunit alpha-like 1	1192.
XP_003397561.2	10799	30196	Insecta	Bombus terrestris	buff-tailed bumblebee	PREDICTED_ acetylcholine receptor subunit alpha-like 1	1066.
XP_012151345.1	9522	143996	Insecta	Megachile rotundata	alfalfa leafcutting bee	PREDICTED_ acetylcholine receptor subunit alpha-like 1	1055.
XP_012249741.1	10602	132113	Insecta	Bombus impatiens	common eastern bumble bee	PREDICTED_ acetylcholine receptor subunit alpha-like 1	1050.
XP_011893111.1	238	64793	Insecta	Wasmannia auropunctata	little fire ant	PREDICTED_ acetylcholine receptor subunit alpha-like 1	1043.
XP_012232337.1	117	83486	Insecta	Linepithema humile	Argentine ant	PREDICTED_ acetylcholine receptor subunit alpha-like 1	1031.
XP_011050487.1	34299	103372	Insecta	Argemoxenus schinotator	Panamanian leafcutter ant	PREDICTED_ acetylcholine receptor subunit alpha-like 1	1030.
XP_012054749.1	417	12867	Insecta	Atta cephalotes	ants	PREDICTED_ acetylcholine receptor subunit alpha-like 1	1028.
XP_011261493.1	34389	104421	Insecta	Crematogaster floridanus	Florida carpenter ant	PREDICTED_ acetylcholine receptor subunit alpha-like 1	1024.

(1 of 39) 1 2 3 4 5 6 7 8 9 10 10 Download Table:

Top of Page

When you run the boxplot generator, you will be prompted to select this .csv file as an input.

If you are an experienced R user, skip to the end of this tutorial for traditional [R documentation](#) for the functions in this R workspace. If you are a beginning R user, or have never used R before, see detailed step-by-step descriptions for each function below. For [troubleshooting help](#), go to the end of this document.

4. If you do not have R on your computer, **download R** at <http://www.r-project.org/>¹. It is a free software used for statistical computing and data visualization. On the R download site, you will be asked to choose a CRAN (Comprehensive R Archive Network) mirror. Choose the mirror that is closest to your current location and the R-version that matches your operating system.
5. Open R then go to the drop-down menu **File → Load Workspace → seqapassboxplot.RData**. If the workspace is loaded correctly you will only see one line of code appear. Your screen might look something like this:

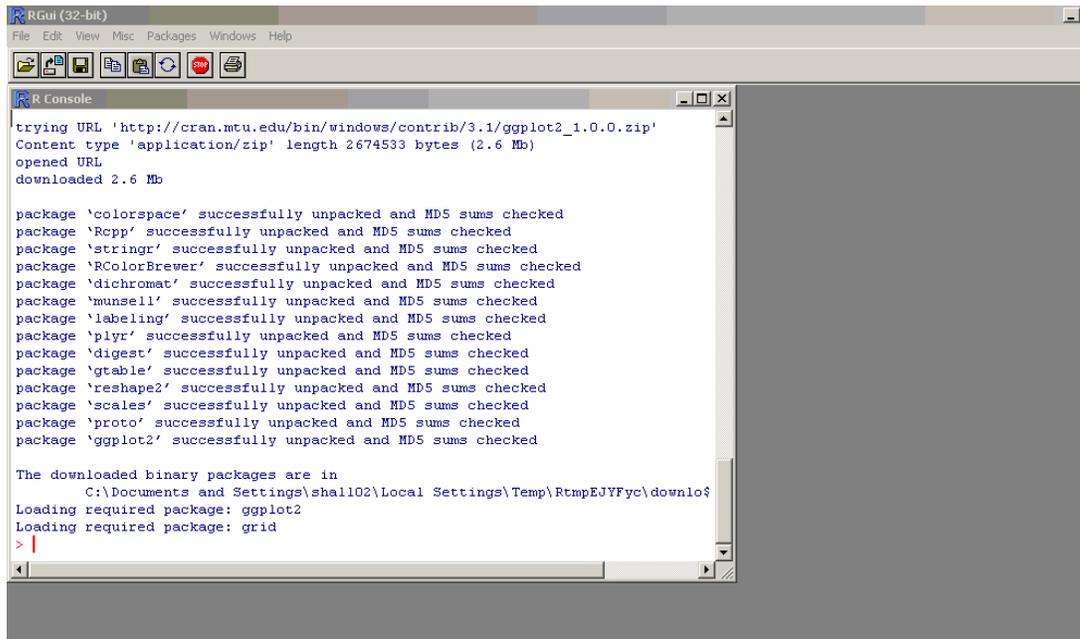
```
> load("C:\\Users\\srobin08\\Documents\\seqapassbox.RData")
```

You will now have access to five different boxplot functions on your R workspace.

1. [seqapassbox\(\)](#) - Creates box-plot of a downloaded .csv file from the SeqAPASS tool.
2. [orthologbox\(\)](#) - Creates box-plot of a downloaded .csv file from the SeqAPASS tool with a red dotplot overlay of ortholog candidates.
3. [taxaselectbox\(\)](#) - Creates box-plot of a downloaded .csv file from the SeqAPASS tool and filters for only the taxa you specify in addition to the taxon of the query species.
4. [subunitbox\(\)](#) - Creates box-plot of downloaded .csv files from the SeqAPASS tool comparing similarity of protein subunits within a specific taxon.
5. [orthosubunitbox\(\)](#) - Creates box-plot of downloaded .csv files from the SeqAPASS tool comparing a specified organism class with specified protein subunits and red dotplot overlay identifying ortholog proteins.

¹ R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Note: These functions utilize the R package ‘ggplot2’.² The functions will test if this package is already downloaded, and if it isn’t, the package will download automatically using the Michigan Technological University CRAN mirror (<http://cran.mtu.edu/>). The user will not have to provide additional inputs, but should be aware that this package is being installed as part of the function call. The first time you call a function per session, you should see the following packages installed. If the user does not have write access to the default R library then downloading the ‘ggplot2’ package will fail. See the [troubleshooting section](#).



```

RGui (32-bit)
File Edit View Misc Packages Windows Help

R Console
trying URL 'http://cran.mtu.edu/bin/windows/contrib/3.1/ggplot2_1.0.0.zip'
Content type 'application/zip' length 2674533 bytes (2.6 Mb)
opened URL
downloaded 2.6 Mb

package 'colorspace' successfully unpacked and MD5 sums checked
package 'Rcpp' successfully unpacked and MD5 sums checked
package 'stringr' successfully unpacked and MD5 sums checked
package 'RColorBrewer' successfully unpacked and MD5 sums checked
package 'dichromat' successfully unpacked and MD5 sums checked
package 'munsell' successfully unpacked and MD5 sums checked
package 'labeling' successfully unpacked and MD5 sums checked
package 'plyr' successfully unpacked and MD5 sums checked
package 'digest' successfully unpacked and MD5 sums checked
package 'gtable' successfully unpacked and MD5 sums checked
package 'reshape2' successfully unpacked and MD5 sums checked
package 'scales' successfully unpacked and MD5 sums checked
package 'proto' successfully unpacked and MD5 sums checked
package 'ggplot2' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Documents and Settings\shall02\Local Settings\Temp\RtmpEJYFyc\downlo$
Loading required package: ggplot2
Loading required package: grid
> |

```

SeqAPASS Boxplot Generator Across Taxa

seqapassbox()

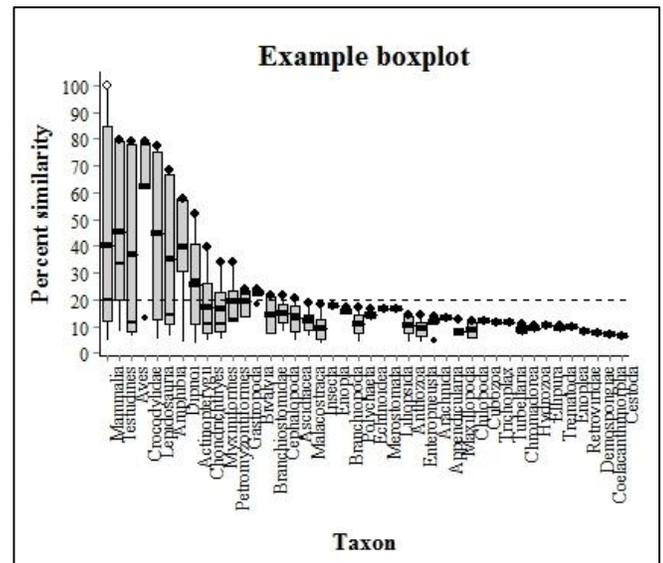
Description of the function: Creates box-plot of a downloaded .csv file from the SeqAPASS tool

- Type **seqapassbox()** into the active R command line, paying careful attention to the underscore, spelling, spacing, and case sensitivity. Close the parentheses but leave them empty.
- When prompted by the pop-up window to select a file, **select the .csv** file containing your SeqAPASS output data from wherever it is stored on your personal computer (Stored during step 2. of this tutorial). **Note: The input file MUST BE IN .csv FORMAT**
- Enter the susceptibility cut-off** as specified in the “Cut-off” column of a SeqAPASS output .csv. Note: Although the R Console may show the spinning “busy icon”, you should proceed to enter the susceptibility cut-off either as a decimal or integer and press enter. Note: The decimal mark should be a “decimal dot” not a comma. The function will take this value and plot it as a dashed line on your graph.

² H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

- d) **Type in an appropriate title** for your boxplot. Do not include quotation marks or any other characters. If you would prefer no title, please press 'Enter'. If you type the title incorrectly, simply call the **seqapassbox()** function again to generate a new boxplot.
- e) Save the resulting boxplot in the desired format by ensuring that the R graphics window is the active window (designated by the word ACTIVE in parentheses), then **File** → **Save as** → **jpeg, PDF etc.** The graph can also be saved or copied from the "R Graphics" window by clicking on the graph, right clicking, then scrolling to the desired save or copy format. To plot fewer taxa see [taxaselectbox\(\)](#).

Function Output: A boxplot to visualize percent similarity among all classes relative to the query species/protein (indicated by an open circle). The *mean* percent similarity appears as a thicker crossbar compared to the thin *median* crossbar. Dashed line represents the susceptibility cut-off. The upper whisker extends from the top of the box (75th percentile) to the highest value that is within 1.5 * IQR of the 75th percentile, where IQR is the inter-quartile range, or distance between the first and third quartiles. The lower whisker extends from the bottom of the box (25th percentile) to the lowest value within 1.5 * IQR of the 25th percentile. Data beyond the end of the whiskers are outliers and plotted as points (as specified by Tukey).



orthologbox()

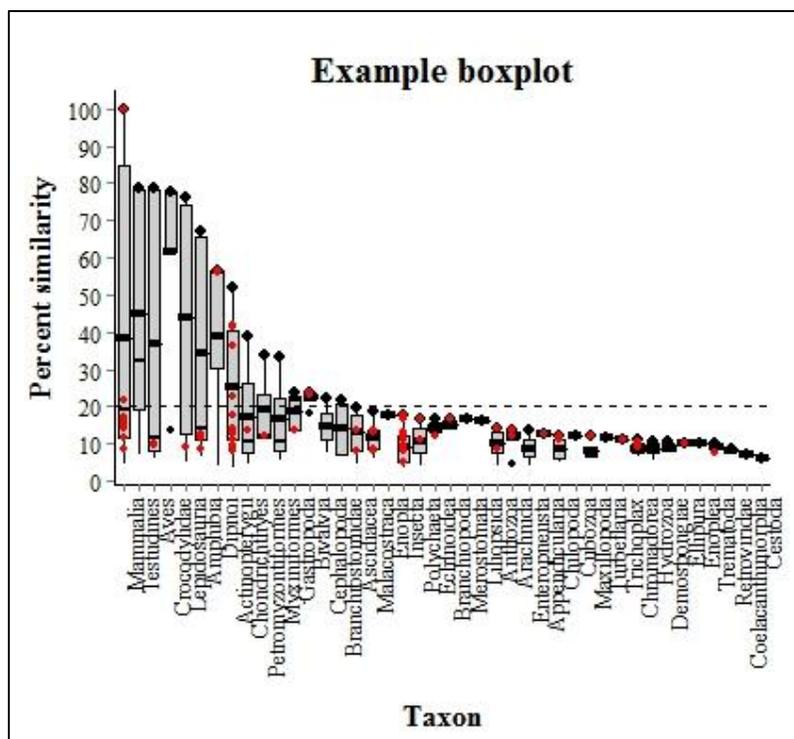
Description of the function: Creates box-plot of a downloaded .csv file from the SeqAPASS tool with a red dotplot overlay of ortholog candidates.

- a) To call this function, **type orthologbox()** into the active R command line, paying careful attention to the underscore, spelling, spacing, and case sensitivity. Close the parentheses but leave them empty.
- b) When prompted by the pop-up window to select a file, **select the .csv** file containing your SeqAPASS output data from wherever it is stored on your personal computer (Stored during step 2. of this tutorial).
Note: The input file MUST BE IN .csv FORMAT
- c) **Enter the susceptibility cut-off** as specified in the "Cut-off" column of a SeqAPASS output .csv. Note: Although the R Console may show the spinning "busy icon", you should proceed to enter the susceptibility cut-off either as a decimal or integer and press enter. Note: The decimal mark should be a "decimal dot" not a comma. The function will take this value and plot it as a dashed line on your graph.
- d) **Type in an appropriate title** for your boxplot. Do not include quotation marks or any other characters. If you would prefer no title, please press 'Enter'. If you type the title incorrectly, simply call the **seqapassbox()** function again to generate a new boxplot.

- e) Save the resulting boxplot in the desired format by ensuring that the R graphics window is the active window (designated by the word ACTIVE in parentheses), then **File** → **Save as** → **jpeg, PDF etc**. The graph can also be saved or copied from the “R Graphics” window by clicking on the graph, right clicking, then scrolling to the desired save or copy format.
- f) The scientific names, common names, taxa and percent similarity for all ortholog candidates have been saved to a .csv file on your computer. They are saved in the current working directory under the name ‘Ortholog_Candidates.csv’. To find out what this working directory is (where this file has been saved) type **getwd()** in the R command line. Alternatively, you can find it on your computer using the search or Finder tool and typing in *Ortholog_Candidates.csv*.

Continued on next page

Function output: A scatterplot to visualize ortholog candidates (red dots) overlaid on boxplots of taxonomic percent similarity relative to the query species/protein (indicated by an open circle). The *mean* percent similarity appears as a thicker crossbar compared to the thin *median* crossbar. Dashed line represents the susceptibility cut-off. The upper whisker extends from the top of the box (75th percentile) to the highest value that is within 1.5 * IQR of the 75th percentile, where IQR is the inter-quartile range, or distance between the first and third quartiles. The lower whisker extends from the bottom of the box (25th percentile) to the lowest value within 1.5 * IQR of the 25th percentile. Data beyond the end of the whiskers are outliers and plotted as points (as specified by Tukey).



If you have any questions or error messages, please see the [troubleshooting section](#) or contact the corresponding author for additional support.

taxaselectbox()

Description of the function: Creates box-plot of a downloaded .csv file from the SeqAPASS tool and filters for only the taxa you specify in addition to the taxon of the query species.

- a) To call this function, **type taxaselectbox()** into the active R command line, paying careful attention to the underscore, spelling, spacing, and case sensitivity. Close the parentheses but leave them empty.
- b) When prompted by the pop-up window to select a file, **select the .csv** file containing your SeqAPASS output data from wherever it is stored on your personal computer (Stored during step 2. of this tutorial). **Note: The input file MUST BE IN .csv FORMAT**
- c) **Enter the susceptibility cut-off** as specified in the “Cut-off” column of a SeqAPASS output .csv. Note: Although the R Console may show the spinning “busy icon”, you should proceed to enter the susceptibility cut-off either as a decimal or integer and press enter. Note: The decimal mark should be a “decimal dot” not a comma. The function will take this value and plot it as a dashed line on your graph.
- d) **Type in an appropriate title** for your boxplot. Do not include quotation marks or any other characters. If you would prefer no title, please press ‘Enter’. If you type the title incorrectly, simply call the **seqapassbox()** function again to generate a new boxplot.
- e) You will be prompted for the organism classes you would like to visualize. You will be given a list of numbers. **Type the numbers for the classes you would like to visualize, either as a range (e.g. 1-20) or individual numbers, separated by commas, but without spaces or any other punctuation. The taxon of the query species will be included automatically.**

```

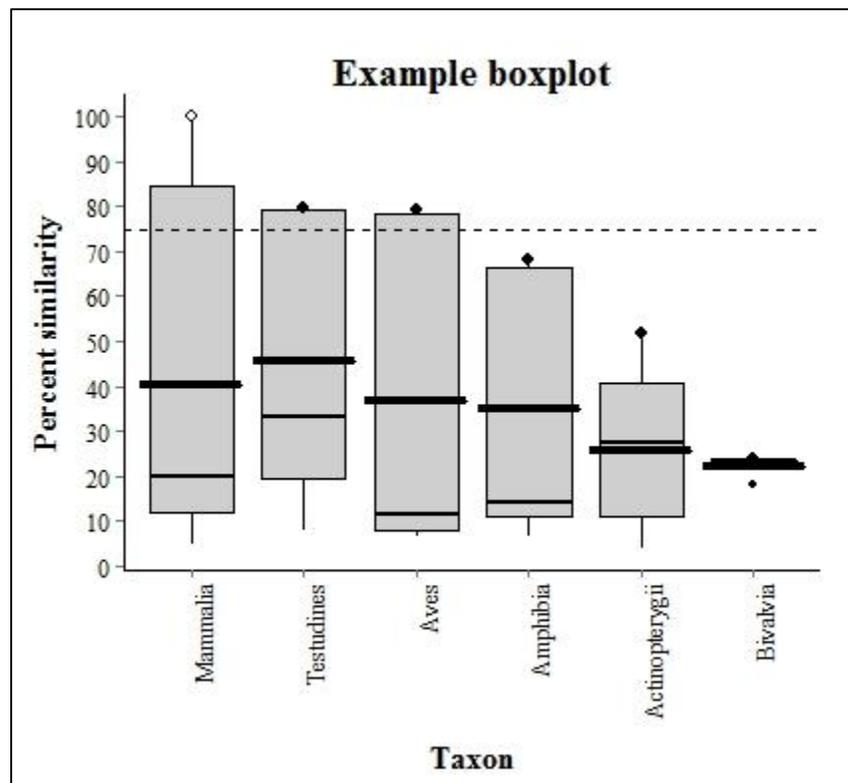
      txns
1      Actinopterygii
2      Amphibia
3      Anthozoa
4      Appendicularia
5      Arachnida
6      Ascidiacea
7      Aves
8      Bivalvia
9      Branchiopoda
10     Branchiostomidae
11     Cephalopoda
12     Chlorophyceae
13     Choanoflagellida
14     Chondrichthyes
15     Chromadorea
16     Chroococcales
17     Deltaproteobacteria
18     Echinoidea
19     Enoplea
20     Enteropneusta
21     Gammaproteobacteria
22     Gastropoda
23     Gloeobacteria
24     Hirudinida
25     Hydrozoa
26     Insecta
27     Lepidosauria
28     Malacostraca
29     Mammalia
30     Maxillopoda
31     Merostomata
32     Oscillatoriales
33     Pelagophyceae
34     Planctomycetia
35     Polychaeta
36     Testudines
37     Trematoda
Using the above list, enter the numbers corresponding to the classes
you would like to include in the boxplot e.g. '1,2,7,10,14':|

```

Continued on next page

- f) Save the resulting boxplot in the desired format by ensuring that the R graphics window is the active window (designated by the word ACTIVE in parentheses), then **File** → **Save as** → **jpeg, PDF etc.** The graph can also be saved or copied from the “R Graphics” window by clicking on the graph, right clicking, then scrolling to the desired save or copy format.

Function output: A boxplot to visualize percent similarity among specified classes to the query species/protein (indicated by an open circle). The *mean* percent similarity appears as a thicker crossbar compared to the thin *median* crossbar. Dashed line represents the susceptibility cut-off. The upper whisker extends from the top of the box (75th percentile) to the highest value that is within 1.5 * IQR of the 75th percentile, where IQR is the inter-quartile range, or distance between the first and third quartiles. The lower whisker extends from the bottom of the box (25th percentile) to the lowest value within 1.5 * IQR of the 25th percentile. Data beyond the end of the whiskers are outliers and plotted as points (as specified by Tukey).



If you have any questions or error messages, please see the [troubleshooting section](#) or contact the corresponding author for additional support.

SeqAPASS Boxplot Generator for Multi-Subunit Proteins

The following pages pertain to the functions `subunitbox()` and `orthosubunitbox()`

Prior to calling either `subunitbox()` or `orthosubunitbox()`:

1. Use the SeqAPASS database to generate .csv files for each protein subunit of interest. You will see these on the 'View SeqAPASS Reports Page'. To download the .csv files, click on the SeqAPASS Query ID link, then Actions → Download → CSV. You will have to **download the file for each subunit individually** and rename them. Download the same type of report (either Full or Primary) for all subunits. Rename each .csv file **EXACTLY AS YOU WOULD LIKE THE LABEL TO APPEAR ON THE X-AXIS OF YOUR BOXPLOT**.
2. Store **ALL** of the subunit .csv files you wish to visualize in a **single folder in your working directory**. **Do not include any additional .csv files which don't pertain to your protein of interest**. To identify your current working directory **type `getwd()` in the R command line**. Set your working directory to the folder where you have all of your protein subunit .csv files stored using the command `setwd()`.

On a Windows PC the command will look like this:

```
> getwd()
[1] "C:/Users/srobin08/Documents"
> setwd("C:/Users/srobin08/Documents/Subunit_Data")
> |
```

On a Mac the command will look like this:

```
> getwd()
[1] "/Users/SerinaRobinson/Desktop/Data"
> setwd("/Users/SerinaRobinson/Desktop/Data")
>
```

See [troubleshooting section](#) at the end of this document if errors occur. Errors are likely due to spelling, punctuation, case, spacing, missing quotation marks, or misplaced backslashes.

Continued on next page

subunitbox()

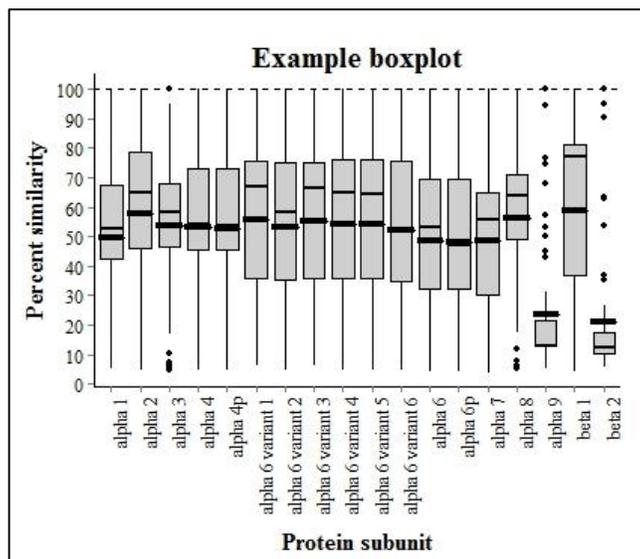
Description of the function: Creates box-plot of a downloaded .csv file from the SeqAPASS tool comparing similarity of protein subunits within a specific taxon.

- To call this function, **type subunitbox()** into the active R command line, paying careful attention to the underscore, spelling, spacing, and case sensitivity. Close the parentheses but leave them empty.
- When prompted, enter the taxonomic class you are interested in comparing to the query species. Relevant taxonomic classes are identified in the “Taxonomic Group” column of the SeqAPASS .csv file. **Type the taxonomic class** in the command line, without any punctuation as follows:

```
>
> Subunit_box()
Please enter the class of organism you are interested in e.g. Insecta. To abort, type exit:Insecta
Please enter the label you would like on the x-axis:|
```

- If the taxonomic group is not recognized (e.g. not existing in the current data set or spelling error), you will be informed of this and no boxplot will be generated.
- Next, enter the desired x-axis label and boxplot title. If you choose not to enter one or both of these, hit the return key twice and no x-axis label or boxplot tile will be included on the plot.
- Save the resulting boxplot in the desired format by ensuring that the R graphics window is the active window (designated by the word ACTIVE in parentheses), then **File → Save as → jpeg, PDF etc.** The graph can be saved or copied from the “R Graphics” window by clicking on the graph, right clicking, then scrolling to the desired save or copy format.

Function Output: A boxplot showing percent similarity for different protein subunits for all species in the specified organism class. The *mean* percent similarity appears as a thicker crossbar compared to the thin *median* crossbar. Dashed line represents the query species/protein at 100% similarity. The upper whisker extends from the top of the box (75th percentile) to the highest value that is within 1.5 * IQR of the 75th percentile, where IQR is the inter-quartile range, or distance between the first and third quartiles. The lower whisker extends from the bottom of the box (25th percentile) to the lowest value within 1.5 * IQR of the 25th percentile. Data beyond the end of the whiskers are outliers and plotted as points (as specified by Tukey).



If you have any questions or error messages, please see the [troubleshooting section](#) or contact the corresponding author for additional support.

orthosubunitbox()

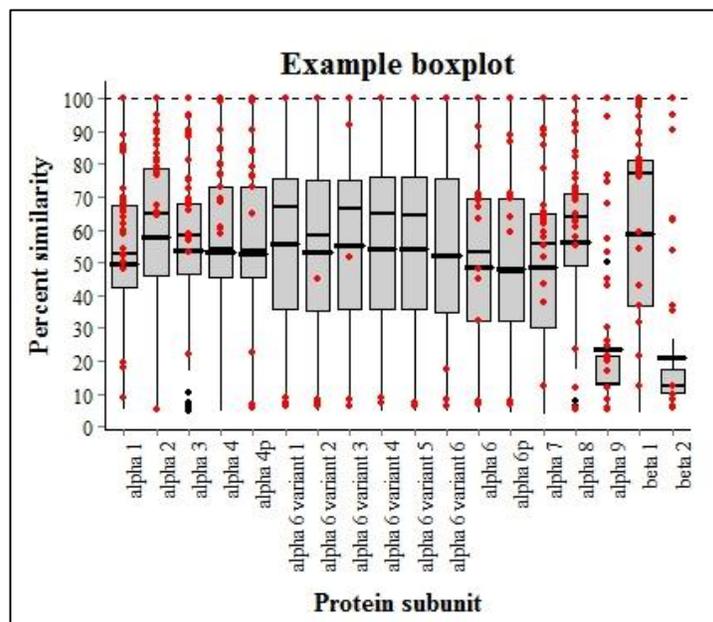
Description of the function: Creates box-plot of a downloaded .csv file from the SeqAPASS tool comparing a specified organism class with specified protein subunits and red dotplot overlay identifying ortholog proteins.

- To call this function, **type orthosubunitbox()** into the active R command line, paying careful attention to the underscore, spelling, spacing, and case sensitivity. Close the parentheses but leave them empty.
- When prompted, you will be asked to enter the taxonomic class you are interested in comparing to the query species. Relevant taxonomic classes are identified in the “ClassName” column of the SeqAPASS .csv file. **Type the taxonomic class** in the command line, without any punctuation as follows:

```
> Orthosubunit_box()
Please enter the class of organism you are interested in e.g. Insecta. To abort, type 'exit':Mammalia
Please enter the label you would like on the x-axis:|
```

- If the class is not recognized (e.g. not existing in the current data set or spelling error), you will be informed of this and no boxplot will be generated. If this occurs, call the function again.
- You will next be asked to **enter the desired x-axis label and boxplot title**. If you would not like to enter one or both of these, simply hit the return key and no x-axis label or boxplot tile will be included on the plot.
- Save the resulting boxplot in the desired format by ensuring that the R graphics window is the active window (designated by the word ACTIVE in parentheses), then **File → Save as → jpeg, PDF etc**. The graph can be saved or copied from the “R Graphics” window by clicking on the graph, right clicking, then scrolling to the desired save or copy format.

Function Output: A scatterplot to visualize ortholog candidates (red dots) overlaid on a boxplot showing percent similarity for different protein subunits for all species in the specified organism class. The *mean* percent similarity appears as a thicker crossbar compared to the thin *median* crossbar. Dashed line represents the query species/protein at 100% similarity. The upper whisker extends from the top of the box (75th percentile) to the highest value that is within 1.5 * IQR of the 75th percentile, where IQR is the inter-quartile range, or distance between the first and third quartiles. The lower whisker extends from the bottom of the box (25th percentile) to the lowest value within 1.5 * IQR of the 25th percentile. Data beyond the end of the whiskers are outliers and plotted as points (as specified by Tukey).



If you have any questions or error messages, please see the [troubleshooting section](#) or contact the corresponding author for additional support.

Troubleshooting

Installing the 'ggplot2' Package

If you receive the following error message, you do not have write access to the default R library

```
> install.packages("ggplot2")
Installing package(s) into 'C:/Program Files/R/R-2.15.2/library'
(as 'lib' is unspecified)
Warning in install.packages :
  'lib = "C:/Program Files/R/R-2.15.2/library"' is not writable
```

In this case, at the start of each session you will need to change the library directory using the `.libPaths()` function.

For example:

```
.libPaths("C:/Rpackages")
```

Type `?libPaths` in the R command line for additional documentation

If you are unsure of the `.libPaths()` function, copy and paste the following line into the R command line

```
install.packages("ggplot2", lib="")
```

And the user will be prompted with a pop-up box to use a personal library instead. Select 'Yes', and a personal library will be automatically set up. Once the library is set up, type `.libPaths()` into the command line, then copy and paste the following line, replacing the "your library name here" with first path specified from the `.libPaths()` command

```
library(ggplot2, lib.loc="your library name here")
```

Now call `seqapassbox()` or other function again and proceed. You will need to do this every time you start a new R session.

Setting the working directory

You will need to set the working directory for the `subunitbox()` and `orthosubunitbox()` functions. If you copy and paste your path, make sure that the backslashes are forward slashes.

For example, if you receive any of the following error messages you will need to go in (using the arrow keys) and change the backslashes to forward slashes.

```
> setwd('C:\Users\srobin08\Desktop\')
Error: '\U' used without hex digits in character string starting "'C:\U"
> |
| > setwd("C:\Documents and Settings\shall02\Desktop\serina")
| Error: '\D' is an unrecognized escape in character string starting "'C:\D"
| > setwd("C:/Documents and Settings\shall02\Desktop\serina")
| Error: '\s' is an unrecognized escape in character string starting "'C:/Documen$
```

If you receive one of the following error messages, it most likely means there was a spelling or punctuation error. In the cases below, the problem is missing quotation marks and misplaced backslashes.

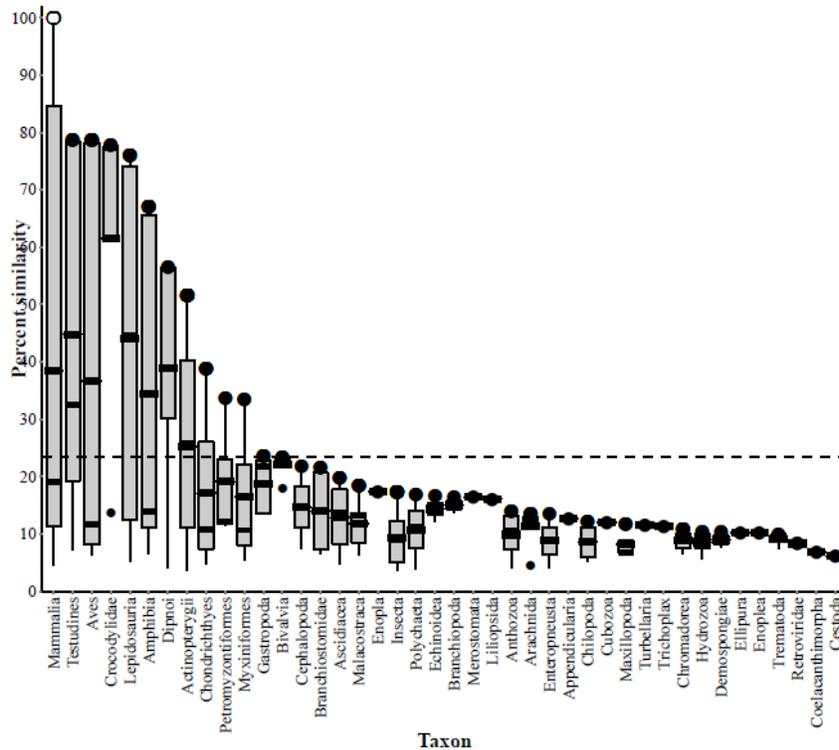
```
> setwd(C:/Users/srobin08/Documents/Subunit_Data/)
Error: unexpected '/' in "setwd(C:/"
> setwd(C:/Users/srobin08/Documents/Subunit_Data)
Error: unexpected '/' in "setwd(C:/"
> setwd(C:Users/srobin08/Documents/Subunit_Data)
Error in setwd(C:Users/srobin08/Documents/Subunit_Data) :
  object 'Users' not found
> setwd(/C:/Users/srobin08/Documents/Subunit_Data)
Error: unexpected '/' in "setwd(/"
```

If you receive the error message below, be sure to check that the folder actually exists and is spelled correctly in the location you are specifying.

```
> setwd("C:/Users/srobin08/Documents/Nonexistent folder")
Error in setwd("C:/Users/srobin08/Documents/Nonexistent folder") :
  cannot change working directory
|
```

Overlapping axis labels

If the graphical output has overlapping axes labels, it is likely a result of your specific computer. If possible, try a different computer, or contact the corresponding author for modified code with rotated axis text.



Spinning 'busy' icon

If you call a function and see the spinning "busy icon", DO NOT be alarmed. This is normal as the function waits for user input. It may be waiting for you to select a file, or to type in the susceptibility cut-off or boxplot title. If you cannot see your blinking cursor and your screen looks like the picture below, expand the R console window so you can see the blinking cursor, then type your response, even if the wheel is spinning.

```
> SeqAPASS_box()
When prompted by the pop-up window, select your SeqAPASS report$
Enter the susceptibility cut-off from your SeqAPASS output, for$
```



Functions not found

If you receive the following message, first make sure that both the spelling and case of the function typed into the command line is correct.

```
> seqapassbox()
Error: could not find function "seqapassbox"
```

If the spelling is correct, then it is possible the R workspace did not load properly. Type `rm(list=ls())` into the command line to clear your existing workspace, then go to **File** → **Load workspace** → **seqapassboxplot.RData** and test your function again.

You can also use the `ls()` command to see the functions loaded in your workspace and check the spelling if you are entering the function wrong.

```
> ls()
 [1] "basebox"          "cleanup"          "maxmean"          "orthologbox"
 [5] "orthosubunitbox" "promptfiles"      "selectclass"      "seqapassbox"
 [9] "subbasebox"       "subunitbox"       "taxaselectbox"    "userprompts"
. |
```

Note that additional functions such as “basebox” and “cleanup” are called within the main functions like `seqapassbox()`, and do not need to be called by the user.

Escaping a function

Press ‘Esc’

If you ever run into the situation shown below, where you have a ‘+’ instead of a ‘>’ in the command line, this means there has been some sort of bracketing error and the function hasn’t ended. If you attempt to type things in when there is a ‘+’, the command won’t execute. To escape this situation, simply press ‘Esc’

```
> {
+
+
+
+
+
+
+ |
```

R documentation

seqapassbox

Description

Generates a boxplot of SeqAPASS output using Hadley Wickham's ggplot2 package.

Usage

```
seqapassbox(filename, suscutoff, boxtitle)
```

Arguments

`filename` (*optional*) The name of the file which the data are to be read from. Each row of the table appears as one line of the file. If it does not contain an *absolute* path, the file name is *relative* to the current working directory, `getwd()`

`suscutoff` (*optional*) A numeric argument for the susceptibility cut-off as specified in the "Cut-off" column of a SeqAPASS output .csv

`boxtitle` (*optional*) A character argument for the desired boxplot title. If no title is desired, enter ""

Value

A ggplot2 boxplot with each box sorted by organism class and a horizontal dashed line indicating the susceptibility cut-off. Thicker crossbars on the boxplot represent mean percent similarity, while the thinner crossbars represent median percent similarity. The query species is specified by an open circle.

Note

If arguments are not provided, then the user will be prompted with `file.choose()` and `readline()` commands to select the file, enter the boxplot title and susceptibility cut-off.

Examples

```
seqapassbox('SeqAPASS_output.csv', 21.3, 'Estrogen receptor similarity')
```

References

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

orthologbox

Description

Generates a boxplot of SeqAPASS output with ortholog candidate dotplot overlay using Hadley Wickham's ggplot2 package.

Usage

```
orthologbox(filename, suscut, boxtitle)
```

Arguments

`filename` (*optional*) The name of the file which the data are to be read from. Each row of the table appears as one line of the file. If it does not contain an *absolute* path, the file name is *relative* to the current working directory, `getwd()`

`suscut` (*optional*) A numeric argument for the susceptibility cut-off as specified in the "Cut-off" column of a SeqAPASS output .csv

`boxtitle` (*optional*) A character argument for the desired boxplot title. If no title is desired, enter ""

Value

A ggplot2 boxplot with each box sorted by organism class with a red ortholog candidate dotplot overlay. A horizontal dashed line indicates the susceptibility cut-off. Thicker crossbars on the boxplot represent mean percent similarity, while the thinner crossbars represent median percent similarity. The query species is specified by an open circle.

A CSV file titled 'Ortholog_candidates.csv' written to the working directory storing the taxa, scientific names, common names, and percent similarity of all ortholog candidates.

Note

If arguments are not provided, then the user will be prompted with `file.choose()` and `readline()` commands to select the file, enter the boxplot title and susceptibility cut-off.

Examples

```
orthologbox('SeqAPASS_output.csv', 21.3, 'Estrogen receptor similarity')
```

References

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

taxaselectbox

Description

Generates a boxplot for a user-specified number of taxa from SeqAPASS output using Hadley Wickham's `ggplot2` package.

Usage

```
taxaselectbox(filename, suscut, boxtitle)
```

Arguments

`filename` (*optional*) The name of the file which the data are to be read from. Each row of the table appears as one line of the file. If it does not contain an *absolute* path, the file name is *relative* to the current working directory, `getwd()`

`suscut` (*optional*) A numeric argument for the susceptibility cut-off as specified in the "Cut-off" column of a SeqAPASS output `.csv`

`boxtitle` (*optional*) A character argument for the desired boxplot title. If no title is desired, enter ""

`numselect` (*optional*) An integer argument specifying the number of taxa you would like to visualize (for example, 20)

Value

A `ggplot2` boxplot with each box sorted by organism class, with the taxa specified shown in the x-axis. A horizontal dashed line indicates the susceptibility cut-off. Thicker crossbars on the boxplot represent mean percent similarity, while the thinner crossbars represent median percent similarity. The query species is specified by an open circle.

Note

If arguments are not provided, then the user will be prompted with `file.choose()` and `readline()` commands to select the file, enter the boxplot title and susceptibility cut-off.

Examples

```
taxaselectbox('SeqAPASS_output.csv', 21.3, 'Estrogen receptor similarity', 40)
```

References

H. Wickham. `ggplot2`: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

subunitbox

Description

Generates a boxplot of multiple protein subunits using multiple SeqAPASS output CSV files stored in the working directory. Uses Hadley Wickham's ggplot2 package.

Usage

```
subunitbox(ClassName, Xlab, boxtitle)
```

Arguments

`ClassName` (*optional*) The taxon name across which protein subunits will be compared. Often is the taxon of the query species, but does not have to be. Argument must be spelled correctly, but is not case-sensitive.

`Xlab` (*optional*) A character argument for the desired x-axis label. If no title is desired, enter ""

`boxtitle` (*optional*) A character argument for the desired boxplot title. If no title is desired, enter ""

Value

A ggplot2 boxplot with each box sorted by organism class, with the protein subunits specified shown in the x-axis. Boxplot x-axis text will be labeled with the file name of each CSV file. A horizontal dashed line indicates the susceptibility cut-off. Thicker crossbars on the boxplot represent mean percent similarity, while the thinner crossbars represent median percent similarity.

Note

If arguments are not provided, then the user will be prompted with readline() commands to select the classname, x-axis label, and boxplot title.

Examples

```
subunitbox('Insecta', 'Protein subunit', 'Estrogen receptor similarity')
```

References

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

orthosubunitbox

Description

Generates a boxplot of multiple protein subunits with a red ortholog dotplot overlay using multiple SeqAPASS output CSV files stored in the working directory. Uses Hadley Wickham's ggplot2 package.

Usage

```
orthosubunitbox(ClassName, Xlab, boxtitle)
```

Arguments

`ClassName` (*optional*) The taxon name across which protein subunits will be compared. Often is the taxon of the query species, but does not have to be. Argument must be spelled correctly, but is not case-sensitive.

`Xlab` (*optional*) A character argument for the desired x-axis label. If no title is desired, enter ""

`boxtitle` (*optional*) A character argument for the desired boxplot title. If no title is desired, enter ""

Value

A ggplot2 boxplot with each box sorted by organism class, with the protein subunits specified shown in the x-axis. Ortholog candidates are overlaid as a red dotplot. Boxplot x-axis text will be labeled with the file name of each CSV file. A horizontal dashed line indicates the susceptibility cut-off. Thicker crossbars on the boxplot represent mean percent similarity, while the thinner crossbars represent median percent similarity.

Note

If arguments are not provided, then the user will be prompted with readline() commands to select the classname, x-axis label, and boxplot title.

Examples

```
orthosubunitbox('Insecta', 'Protein subunit', 'Estrogen receptor similarity')
```

References

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.