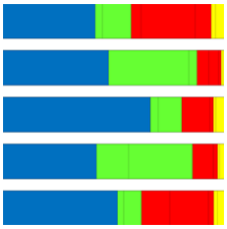


Explicet Basic Tutorial

“Explicet” is from the Latin: explain, unfold, extend, set forth, exhibit, disentangle

A brief introduction to program capabilities and functions for new users of the Explicet software

v2.8.6_20131010

Table of Contents

Note: Clicking on any level of the table of contents will take you directly to that section.

I. Begin a New Project 3
A. Create a New Project 3
II. Import OTU Data 5
A. Import the OTU Data 5
III. Import Metadata 7
A. Import the Metadata 7
IV. Save the Project 9
A. Save the Project 9
V. Adjust the Display 10
A. Hierarchy, OTU, or Both 10
B. Counts, % of Library, % of Total 10
C. OTU displays 10
D. Hierarchy Level 10
E. Sort Libraries Based on a Metadata Tag (Anatomical Position) 11
VI. Make an OTU Stacked Bar Chart 15
A. Create an OTU Stacked Bar Chart of the Top 10 Most Prevalent Taxa 15
B. Change the Title and Label the Axes 18
C. Save the OTU Stacked Bar Chart as a Figure 19
VII. Make a Pie Chart 23
A. Create a Pie Chart of the Project Components 23
B. Make a Pie Chart with Sub-Wedges 26
C. Change Wedge Colors in the Pie Chart 29
VIII. Create a Workspace 32
A. Create a New Workspace 32
IX. Apply a Filter 33
A. Create a Filter 33
B. Set Up the Filter Parameters 34
X. Beta Diversity (Morisita-Horn) 42
A. Create a Morisita-Horn Heatmap 42
XI. Alpha Diversity 44
A. Run a Good's Coverage Test 44
XII. Two-Part Test 48
A. Run a Two-Part Test 48
B. Set Up Filter Parameters 50
XIII. Close the Project 62

© 2012-2013 Charles E. Robertson and Incubix, Inc.

The data used in this tutorial come from an analysis of 16S ribosomal RNA gene sequences obtained from many distinct skin sites of healthy humans (Grice EA, et al. (2009) Topographical and Temporal Diversity of the Human Skin Microbiome. Science 324(5931): 1190–1192). To produce a concise tutorial, the data have been reduced from the original dataset and may not represent the findings of the original study.

I. Begin a New Project

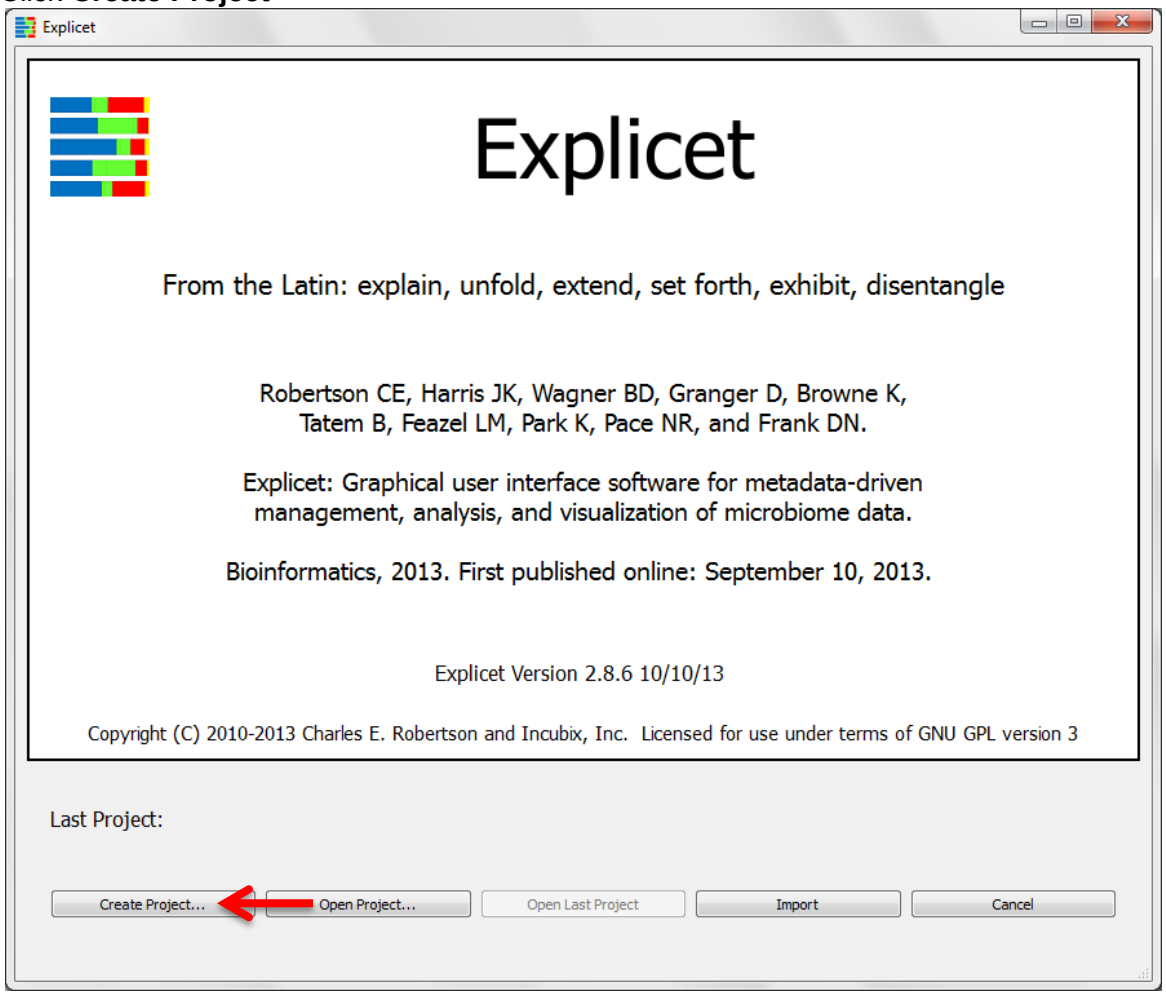
An Explicet project is a single file that contains all of the OTU data, sample names (a.k.a. library names) and metadata that are to be analyzed as a unit. In other words, all data analyzed for one publication are drawn together into a single Explicet project, independent of how many 454/Miseq runs are involved.

We will begin by creating a project and importing an OTU table. The tutorial example we have selected is based on the Human Skin Microbiome paper published by Grice, et al. This example was picked because it is relatively small and has a nice set of intuitive metadata available.

Please do not hesitate to ask questions or make suggestions via our online Explicet forum. The Explicet forum link can be found on our web site: www.explicet.org.

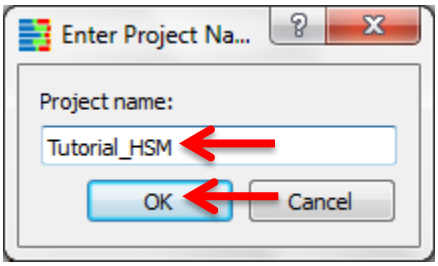
A. Create a New Project

Open Explicet
A pop-up window will open with several different options
Click **Create Project**

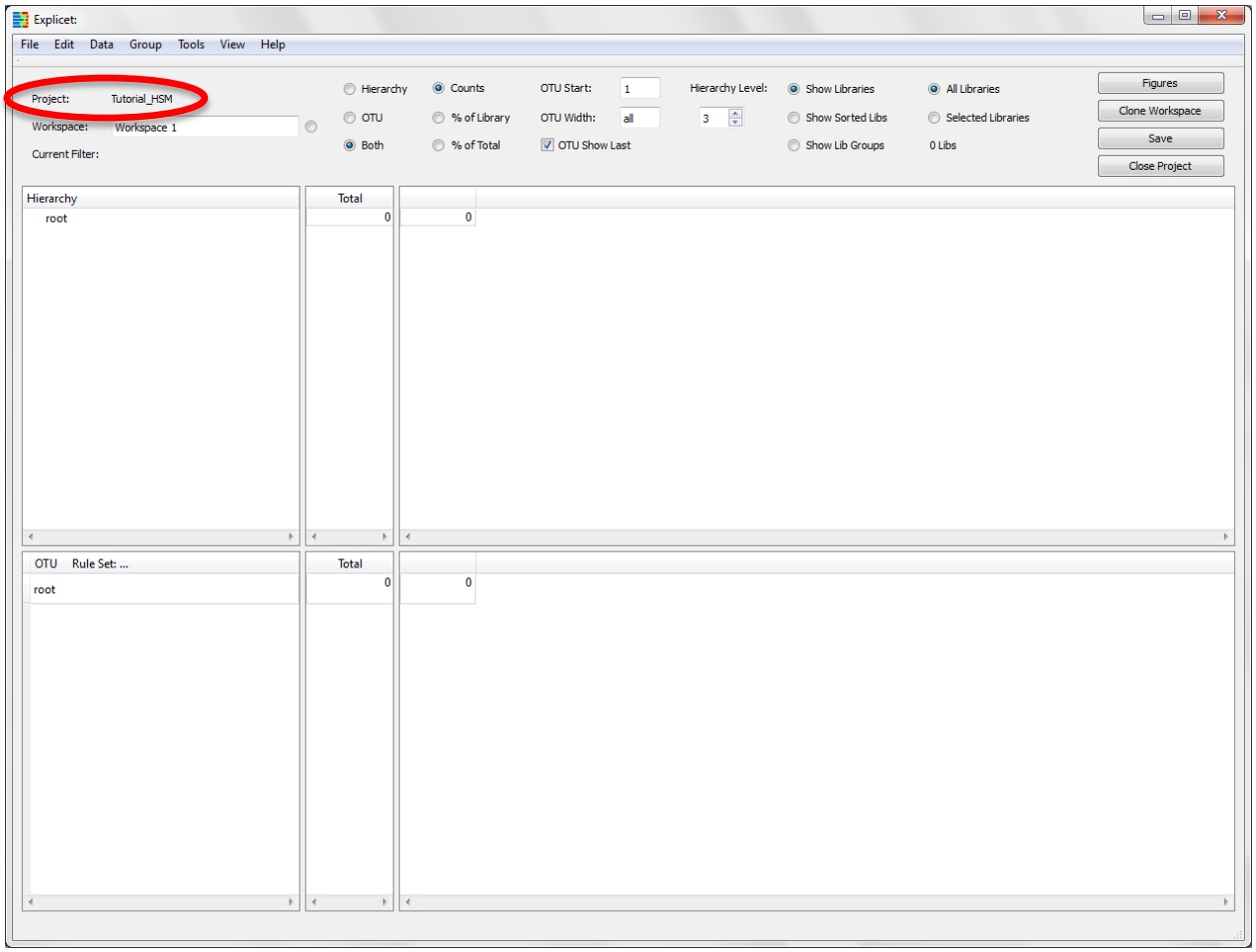


© 2012-2013 Charles E. Robertson and Incubix, Inc.

Enter **Project Name**, “Tutorial_HSM”, when prompted
Click **OK**



We now have a blank project in Explicet, and the name of the current project is displayed in the upper left corner of the current workspace window.



© 2012-2013 Charles E. Robertson and Incubix, Inc.

II. Import OTU Data

The first step in a new project is to import the data that comes out of the 16S pipeline runs into Explicet. In general, OTU tables are the most convenient form of data commonly generated by pipelines. For detailed information on how OTU tables are formatted, please see the Explicet Handbook. In short, OTU tables are a delimited file (tab-separated or comma-separated file) in which the rows are the OTUs, and the columns represent the number of each OTU seen in a given sample.

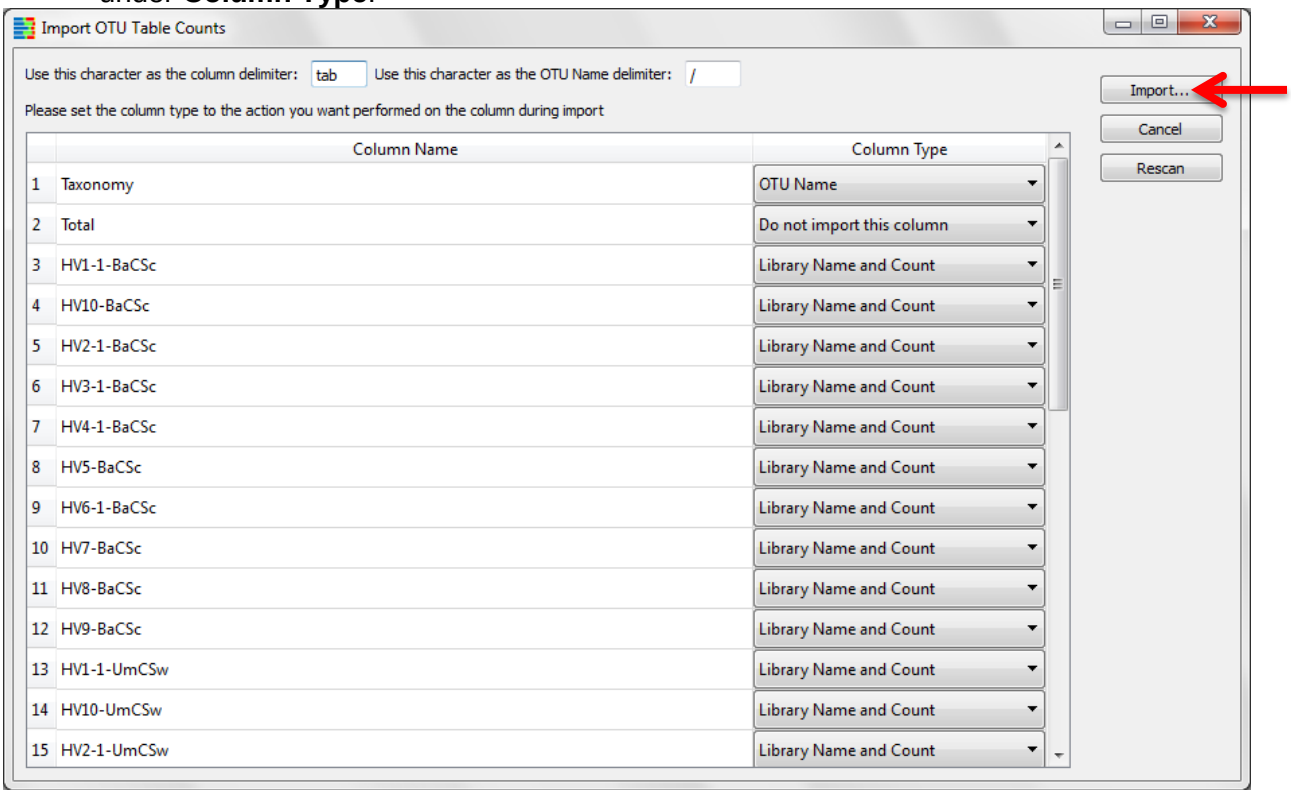
Explicet supports many other formats for importing the OTU data. For more details on the other OTU import formats, please see the Explicet Handbook. Later, we will discuss more data management tools that allow you to explore and modify subsets of the dataset without disrupting the larger project.

Now we will import the data that will belong to the new project. Once data are imported to a project, they are permanently associated with the project. Additional data can be incrementally imported to the same project. Thus, the Explicet project file can grow as a project evolves.

A. Import the OTU Data

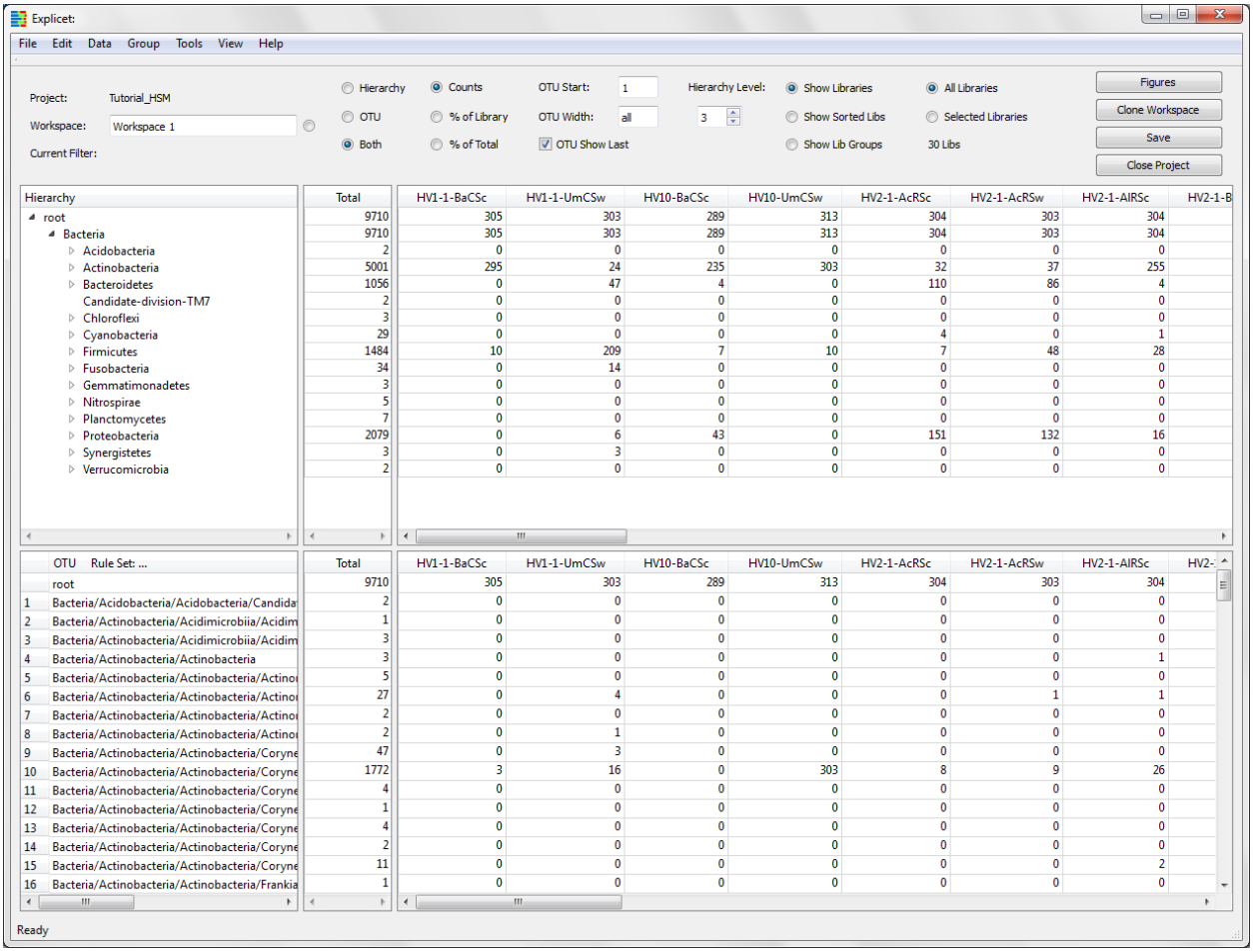
File → Import → File → OTU Table Counts
Select "Tutorial_HSM_OTU_2_Explicet"
Click **Open**
A pop-up window will open
Click **Import**
A dialog box below will open

- On this dialog, Explicet tells the user how it is interpreting the rows and columns in the OTU table. The user needs to verify that Explicet has interpreted the table correctly. Note that in this case, Explicet is telling the user that it is not going to import column 2, "Total", as it will generate that sort of information itself. If Explicet gets it wrong, the user can adjust the interpretation using the provided pull down lists under **Column Type**.



The OTU data now appear in the current workspace window

© 2012-2013 Charles E. Robertson and Incubix, Inc.



© 2012-2013 Charles E. Robertson and Incubix, Inc.

III. Import Metadata

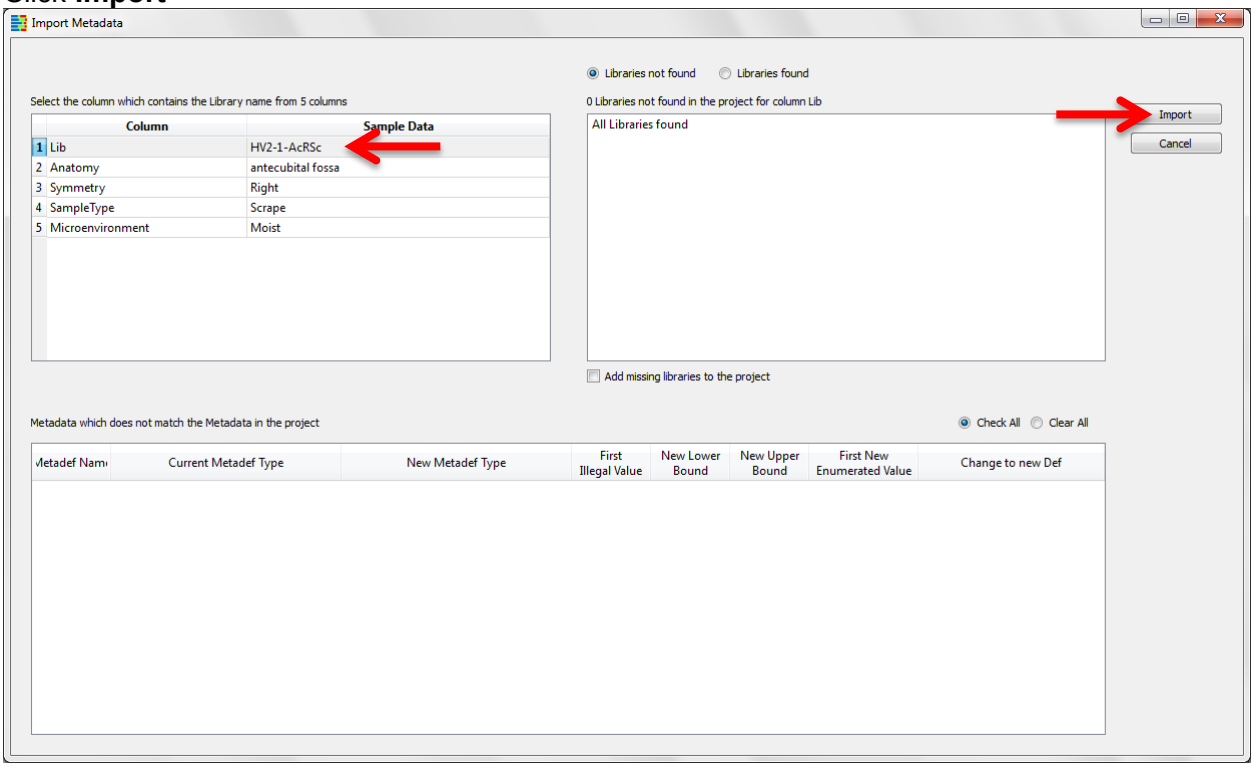
Now we will import the metadata associated with the OTU data. Metadata refers to information about the sequence data - in this case, a description of the samples and subjects from which the sequence data were generated. In our nomenclature, a "library" represents all of the sequences generated from a single sample (multiple libraries may be generated from a given sample, for example through multiple PCR reactions, but for this tutorial we will assume a one-to-one relationship between libraries and samples). In this study, the metadata for each library includes the anatomical position, microenvironment description, sample acquisition method, and side of the body associated with each skin sample. Just like the OTU data, metadata need be imported only once (unless you choose to add more metadata) - imported metadata are also incorporated into the Explicet project file. For detailed information on how to format metadata files, please see the Explicet Handbook. In short, the metadata file is a tab-separated or comma-separated file organized by columns, generally prepared with a spreadsheet package like Microsoft Excel. The first column contains the names of the libraries in the dataset; all subsequent columns are metadata items and their values associated with each library.

A. Import the Metadata

File → Import → Metadata
Select "Tutorial_HSM_Metadata"
Click **Open**
A pop-up window will open
Make sure that the column containing the library name is selected

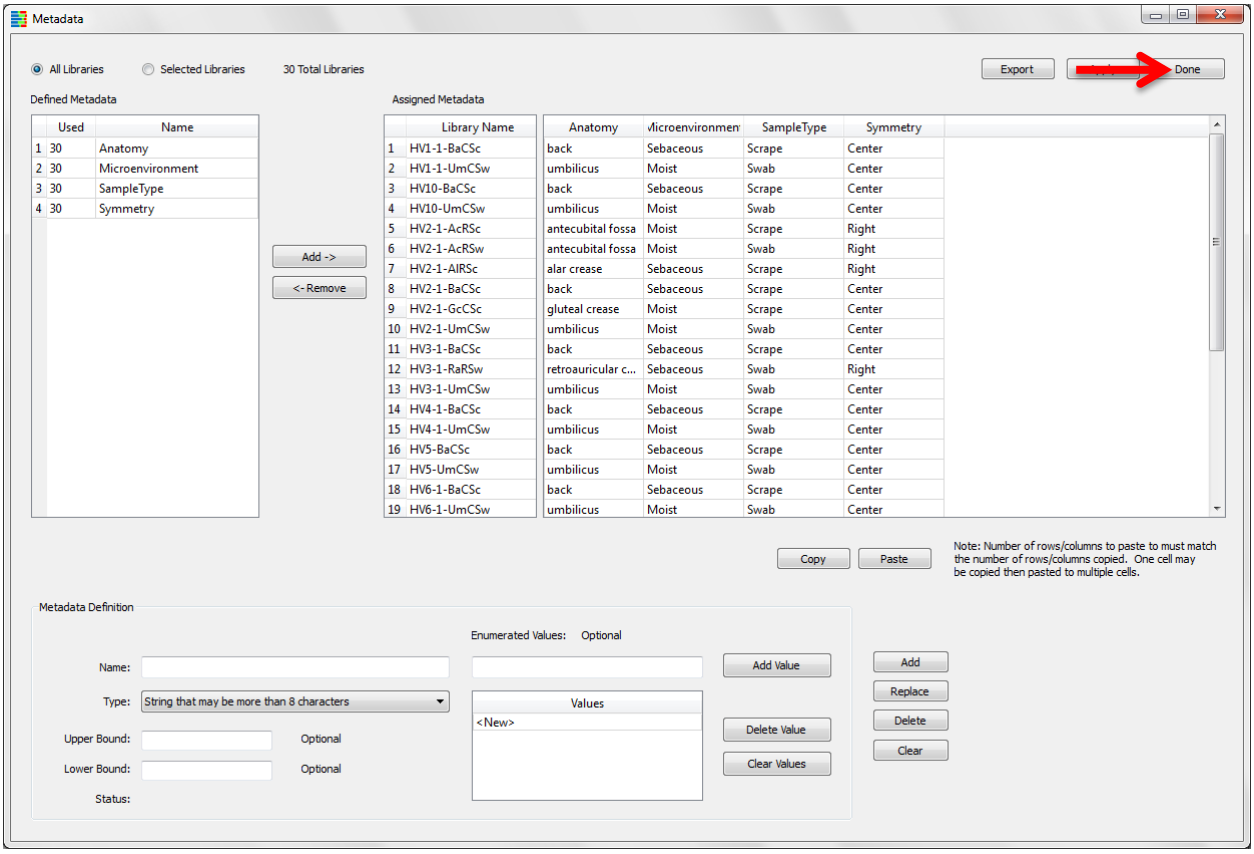
- Explicet searches all of the columns in the metadata file looking for the library names that were found when the taxonomy data were imported. In all but rare cases (e.g., when only a small portion of the sample names are present in the imported taxonomy data), Explicet will find the library column automatically.

Click **Import**



A new pop-up window will open which displays the imported metadata
Click **Done**

© 2012-2013 Charles E. Robertson and Incubix, Inc.



For our example dataset, all of the library names were found in the metadata file, as indicated in the left-hand pane: i.e., the number under **Used** (**30**) matches the total number of libraries shown above the two panes (**30 Total Libraries**).

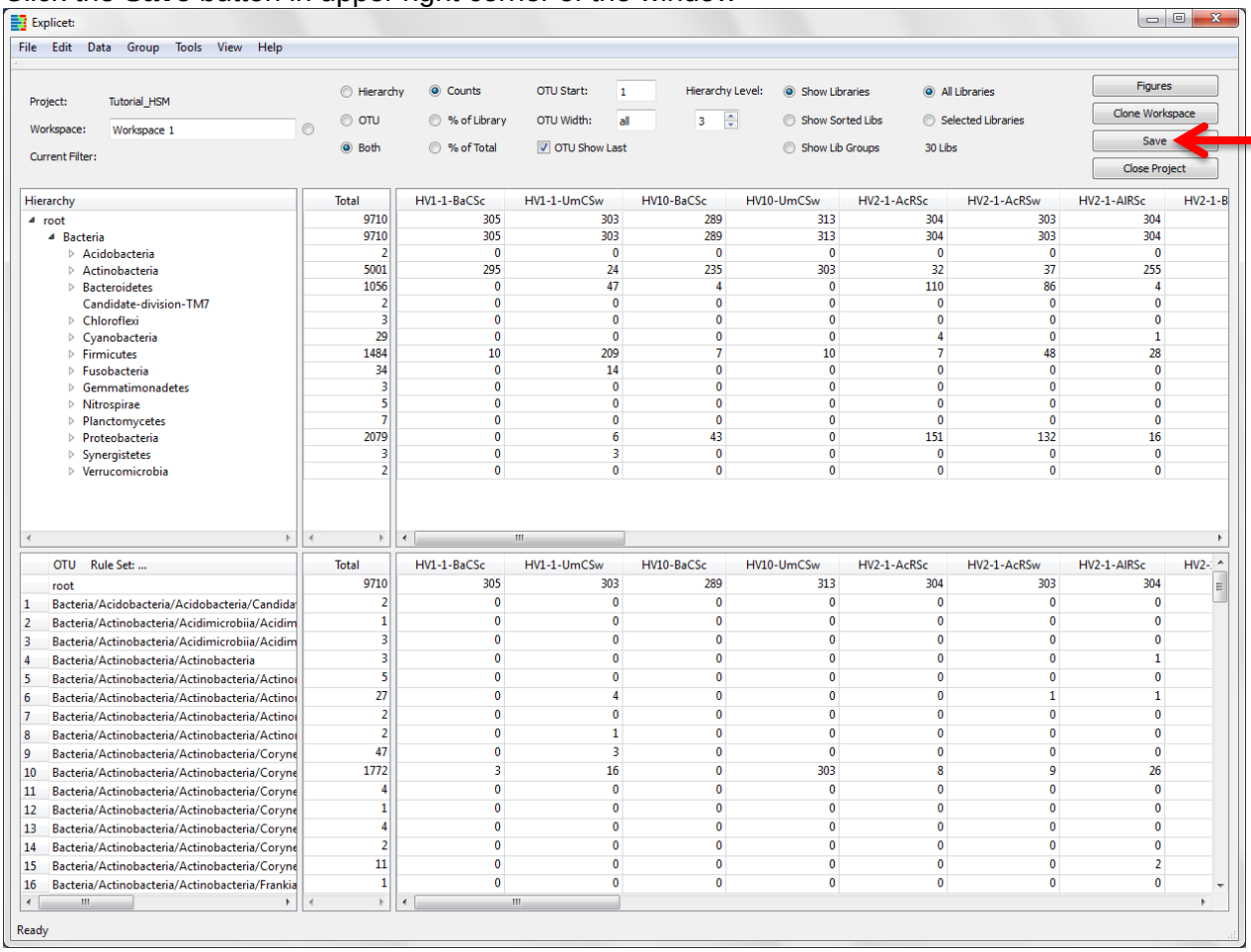
© 2012-2013 Charles E. Robertson and Incubix, Inc.

IV. Save the Project

Now that all of the data associated with the project are imported, the file should be saved. Explicet does not auto-save, so remember to save your project frequently!

A. Save the Project

Click the **Save** button in upper right corner of the window



Enter desired project name and location when prompted

- The default file name is the project name with an "_Explicet_Project" extension.

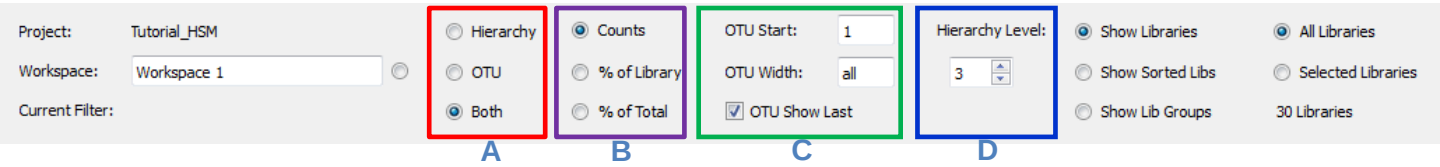
Click **Save**

All of the imported information is now saved within the project file.

© 2012-2013 Charles E. Robertson and Incubix, Inc.

V. Adjust the Display

Now we will adjust the current workspace window display for ease of use (detailed demonstration on next page).



A. Hierarchy, OTU, or Both

Both is the default

This option creates two panes on workspace screen; the upper pane shows the Hierarchy, and the lower pane shows the OTUs. The Hierarchy pane allows exploration of the dataset in a "big tree" hierarchical context, whereas the OTU pane shows a more literal view of the data from the 16S pipeline. The information in the OTU pane is used for input into the statistics and most of the plots (except for pie charts, which are graphical depictions of the Hierarchy pane).

B. Counts, % of Library, % of Total

Select **% of Library** (**Counts** is the default)

While Counts is the default (raw sequence data counts in integers), % of Library tends to be more useful. % of Library is relative abundance, which is important since the total number of Counts received from any library is beyond our control. Using the relative abundance, or % of Library, allows us to fairly compare libraries. Otherwise, the libraries that have a very large number of counts will skew conclusions.

C. OTU displays

These options control the manner in which the taxonomy lines are displayed on the OTU pane.

OTU Start: 1 is the default
This is the position (counting from one) of the first taxonomic category that the user desires to be displayed. In our tutorial example, the taxonomy lines in the OTU pane display will start with Bacteria (Bacteria is the "1"st lineage level).

Set **OTU Width** to 2 ("all" is the default)
This is the number of positions on the line to be displayed. To save space on the screen, now only 2 taxonomic levels will be displayed in the OTU taxonomy line. Taxonomies with more than 2 levels will be shown with an embedded ellipsis; for example, "Bacteria/Actinobacteria/Acidimicrobiia/Acidimicrobiales" becomes "Bacteria/Actinobacteria/…/Acidimicrobiales".

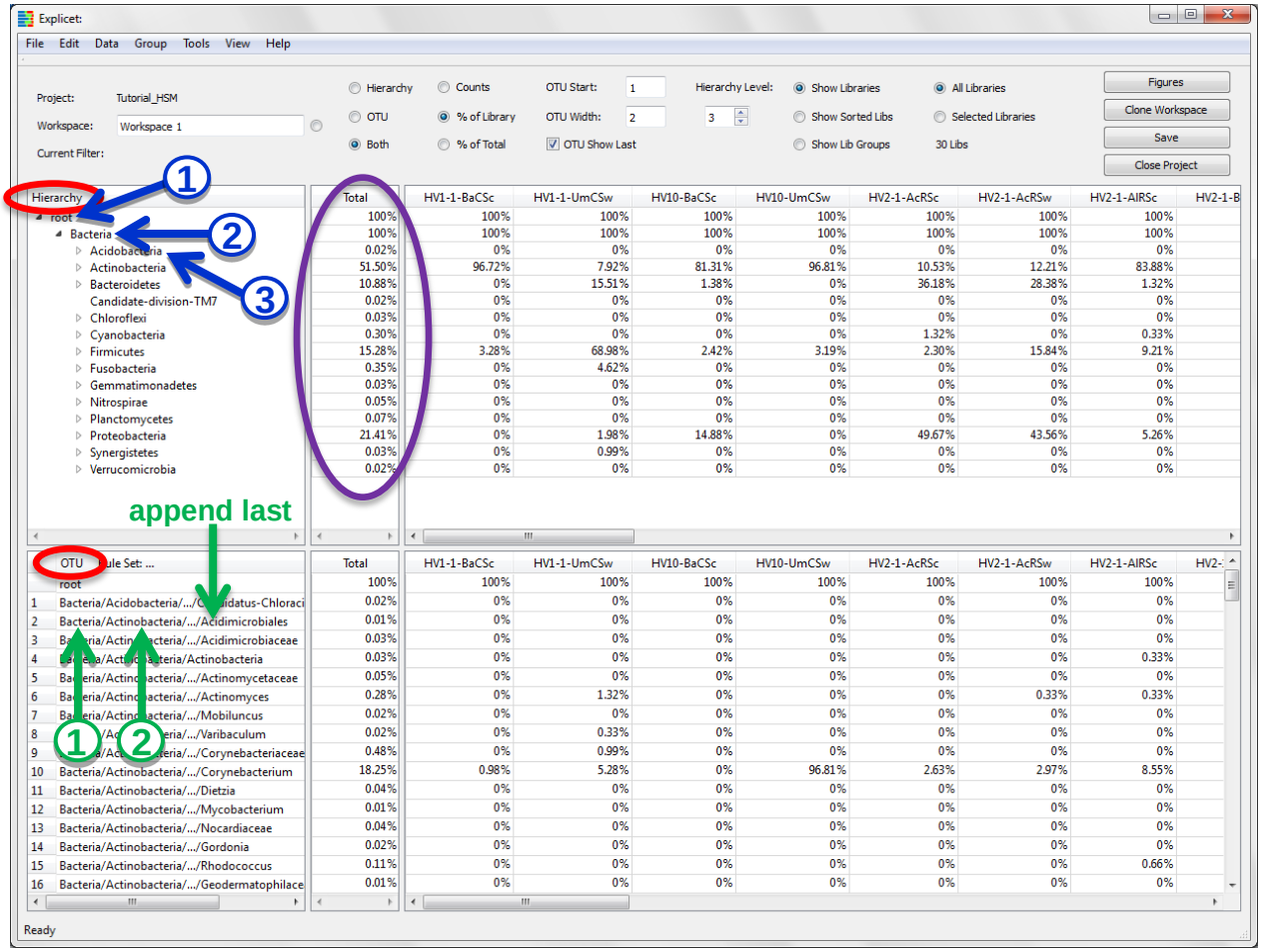
OTU Show Last *on* is the default
This option appends the last item in the taxonomic line onto a truncated OTU lineage.

D. Hierarchy Level

Hierarchy Level: 3 is the default

This controls the number of taxonomic categories that will be opened on the hierarchy pane.

© 2012-2013 Charles E. Robertson and Incubix, Inc.

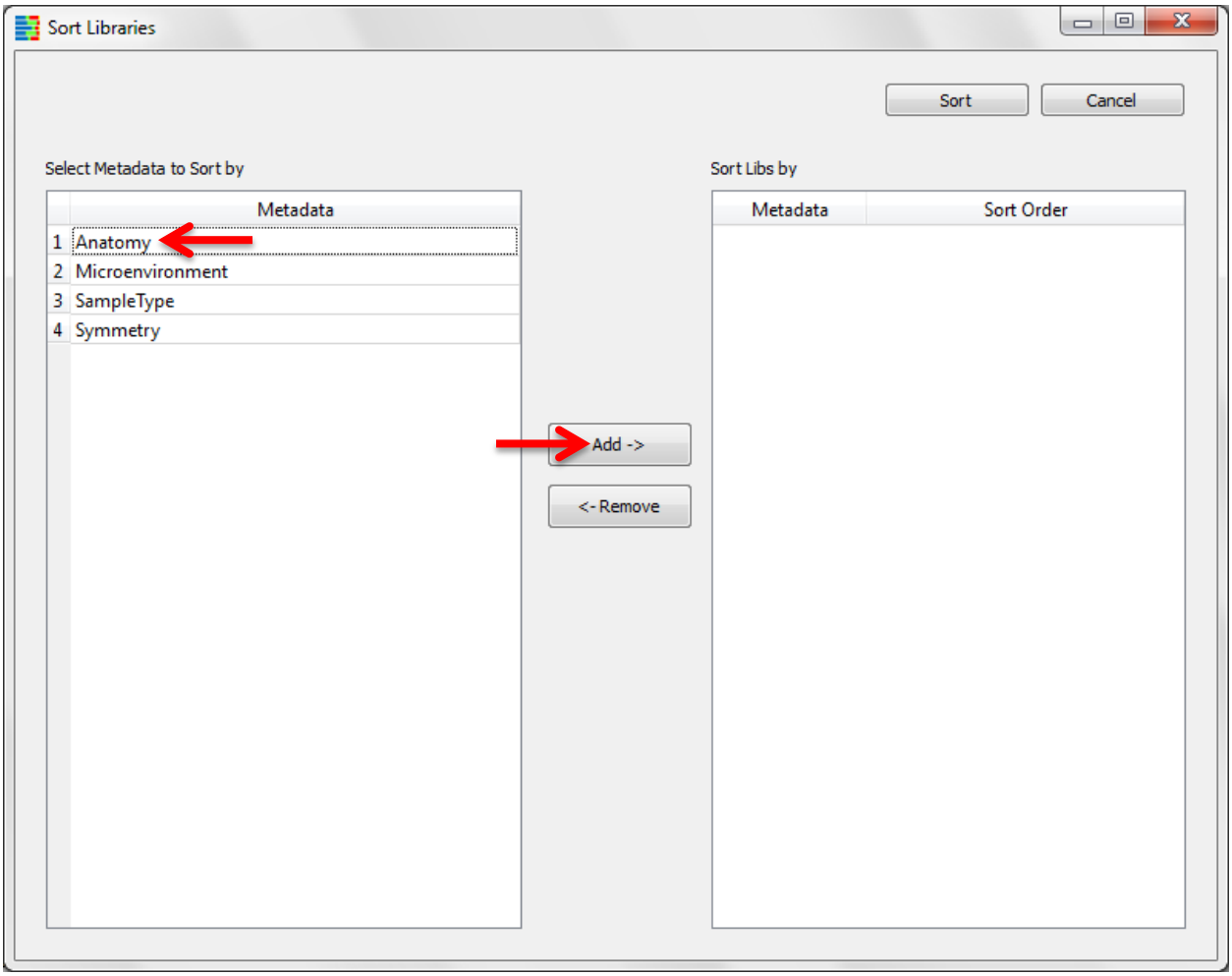


Since libraries are often cryptically named, it's nice to add a readable metadata tag in the view so that we have some context for the libraries we are viewing. To do this, we will sort the libraries in the view based on anatomical position.

E. Sort Libraries Based on a Metadata Tag (Anatomical Position)

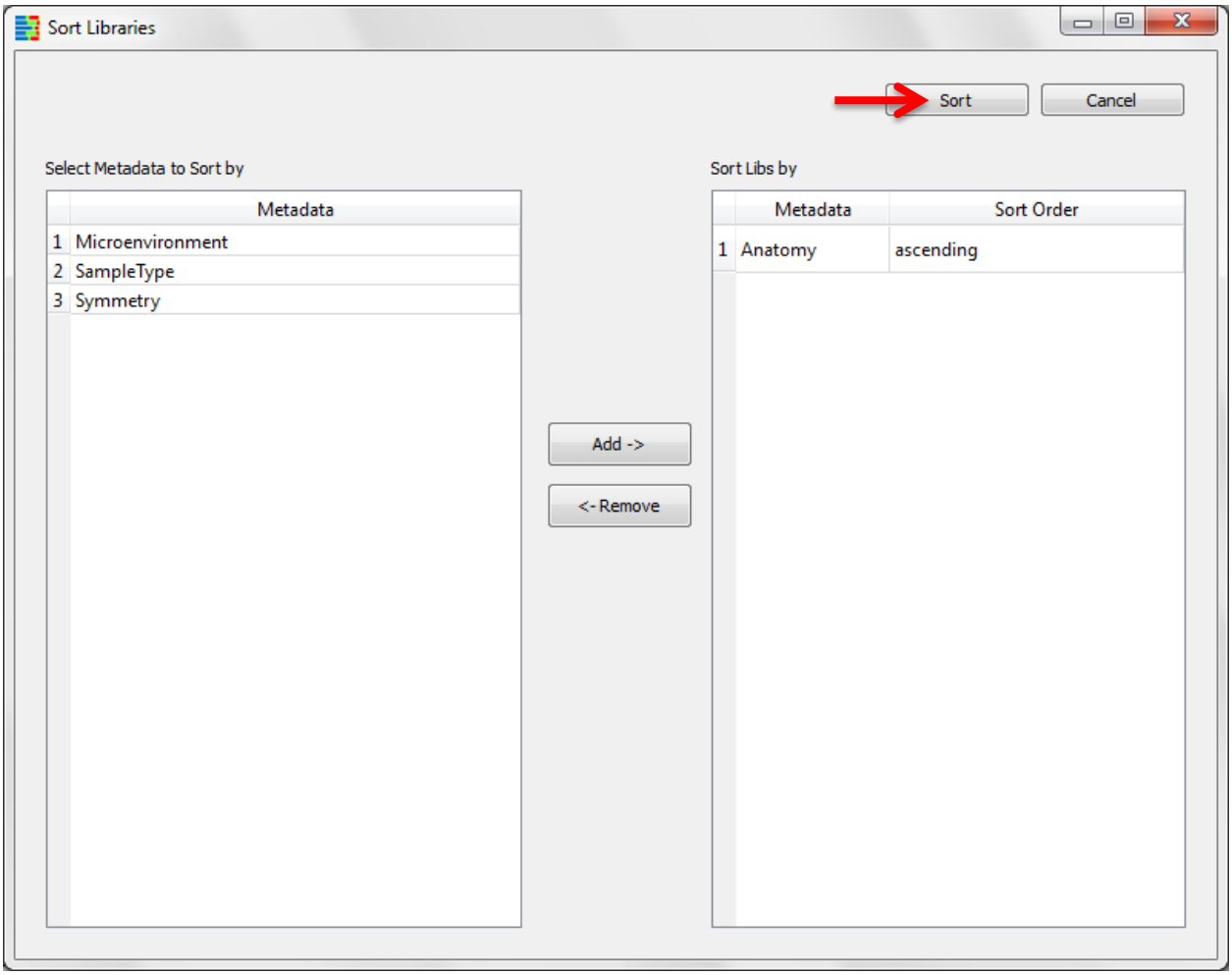
View → Sort Libraries
A pop-up window will open
In left panel, select **Anatomy**
Click **Add** button between panels

© 2012-2013 Charles E. Robertson and Incubix, Inc.



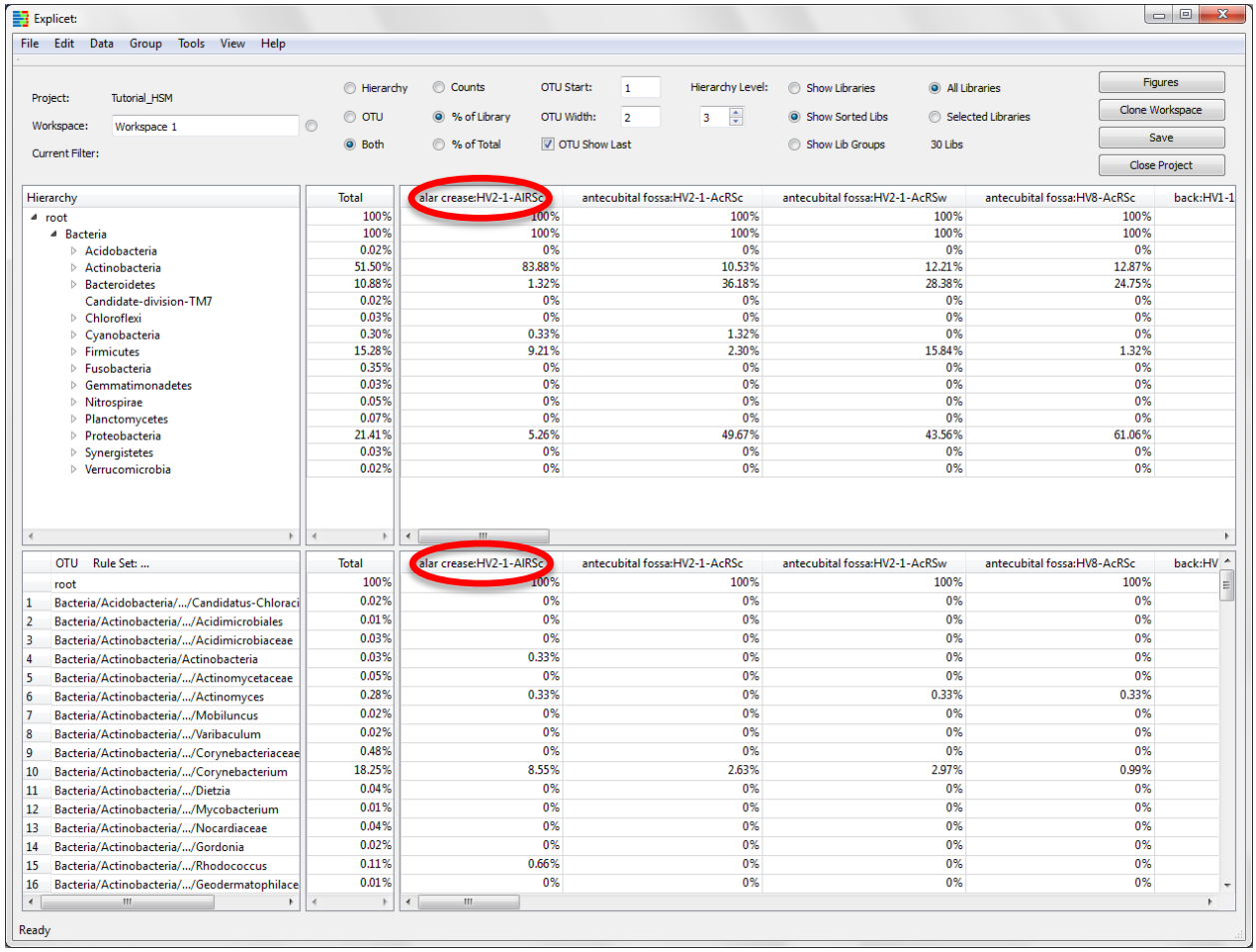
Name of metadata descriptor will appear in the right panel
Click **Sort**

© 2012-2013 Charles E. Robertson and Incubix, Inc.



Pop-up window will disappear
Both the hierarchy and OTU tables are now sorted by anatomical position

© 2012-2013 Charles E. Robertson and Incubix, Inc.



© 2012-2013 Charles E. Robertson and Incubix, Inc.

VI. Make an OTU Stacked Bar Chart

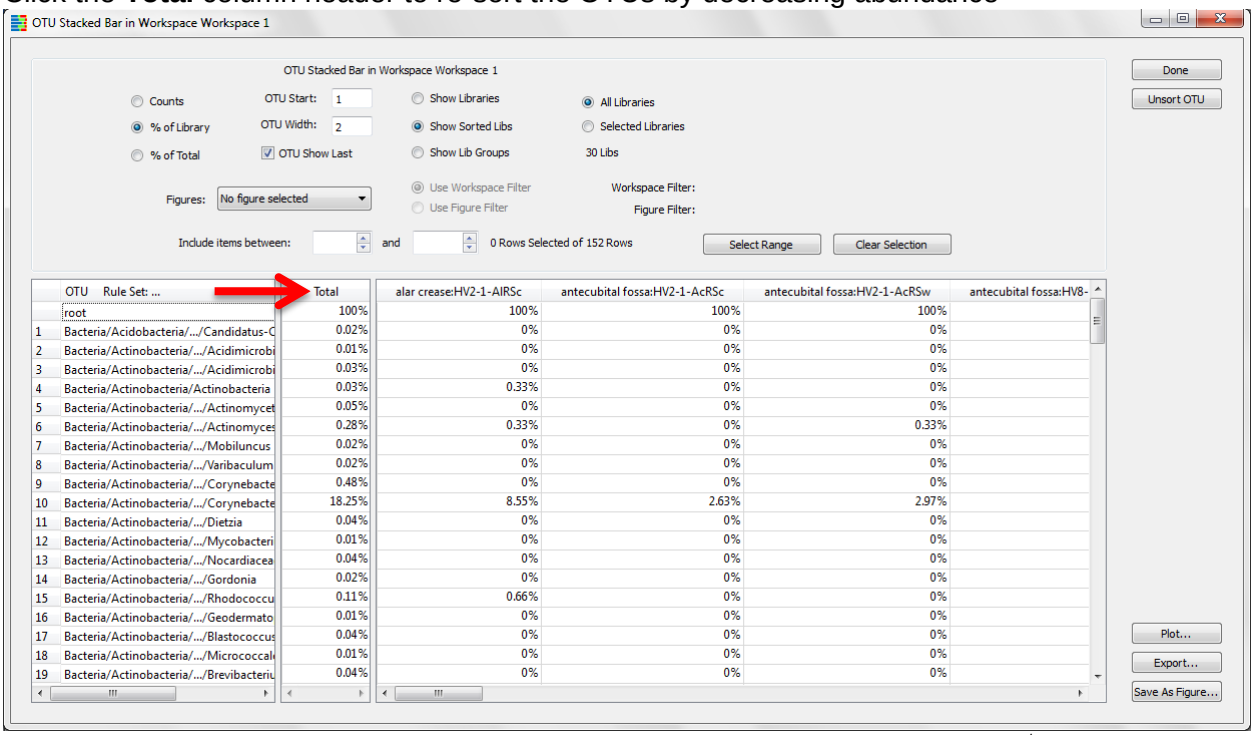
Before diving into a detailed analysis, generating an overview of the dominant organisms that exist in the dataset can be useful. One way to do this is through an OTU stacked bar chart.

A. Create an OTU Stacked Bar Chart of the Top 10 Most Prevalent Taxa

Tools → Plot → OTU Stacked Bar

A new window will appear with the OTU data available in the workspace

Click the **Total** column header to re-sort the OTUs by decreasing abundance



To display only the top 10 taxa in the project, note that the **Total** value of the 1st OTU in the column is 31.35

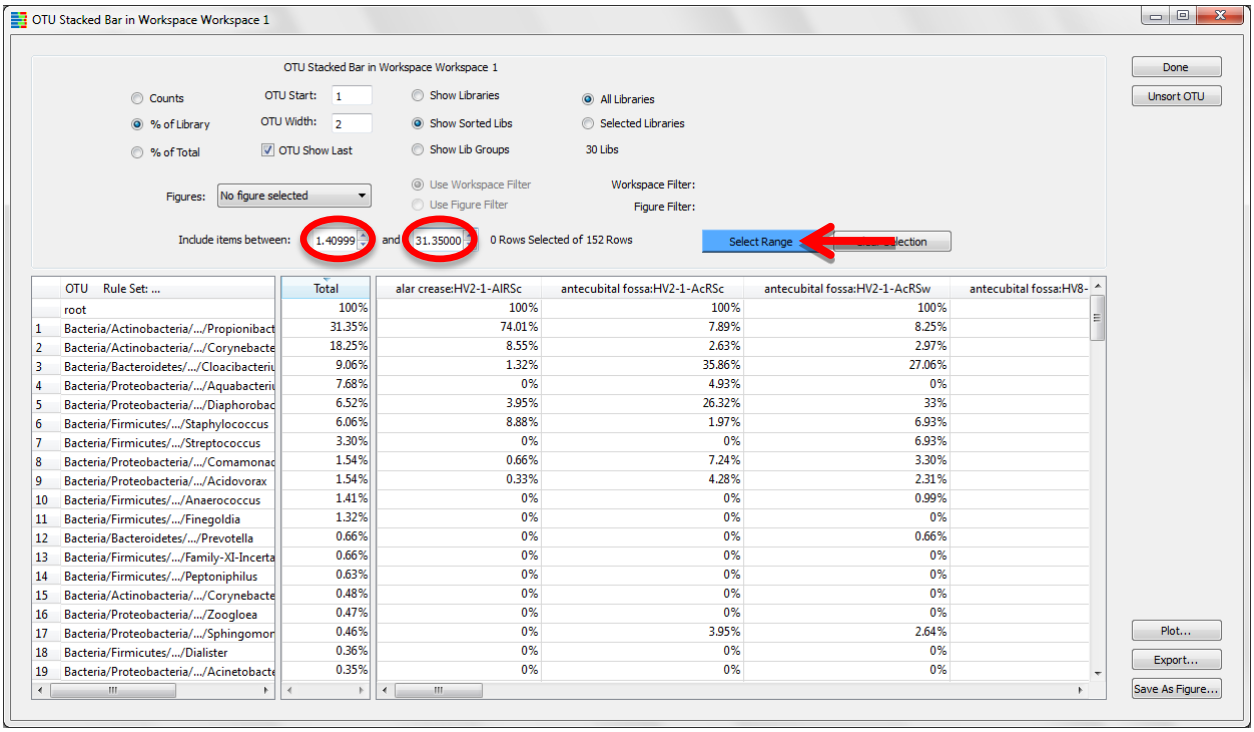
Note that the **Total** value of the 10th OTU in the column is 1.41

In the **Include items between** field, enter “1.41” into the first box (the lower bounding limit)

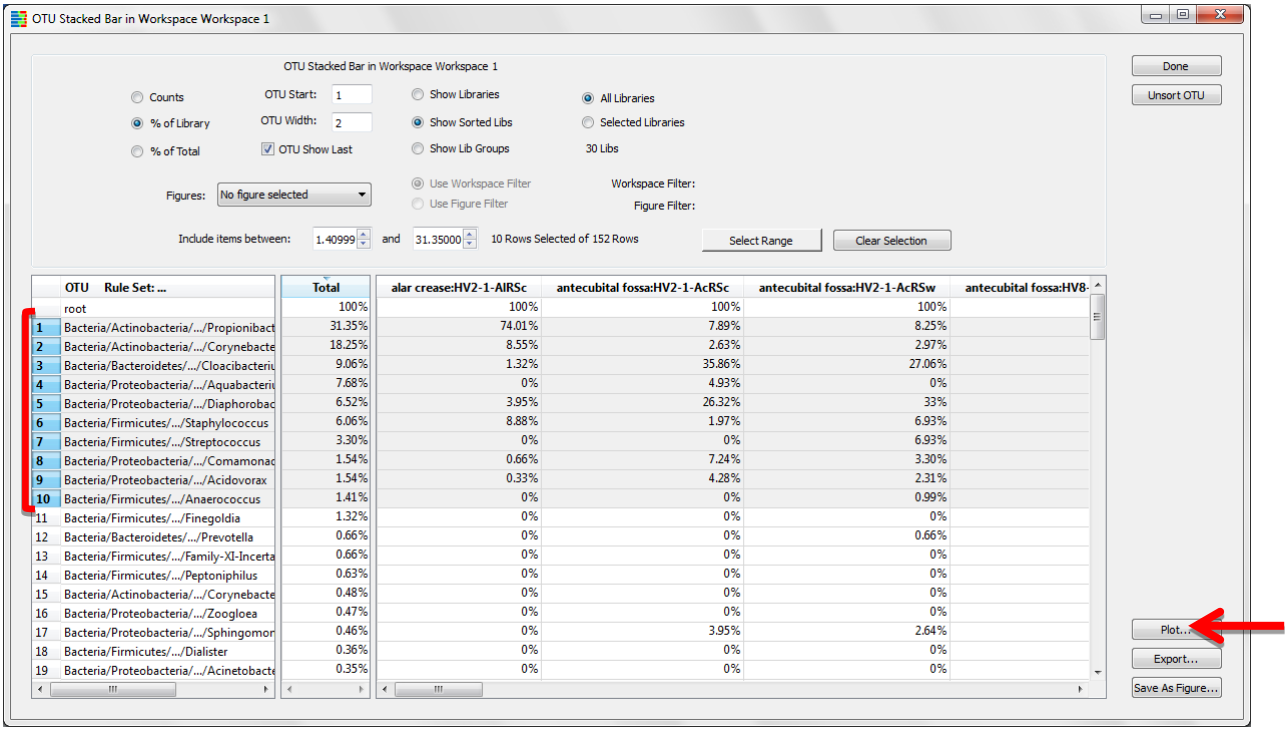
In the **Include items between** field, enter “31.35” into the second box (the upper bounding limit)

Click **Select Range**

© 2012-2013 Charles E. Robertson and Incubix, Inc.

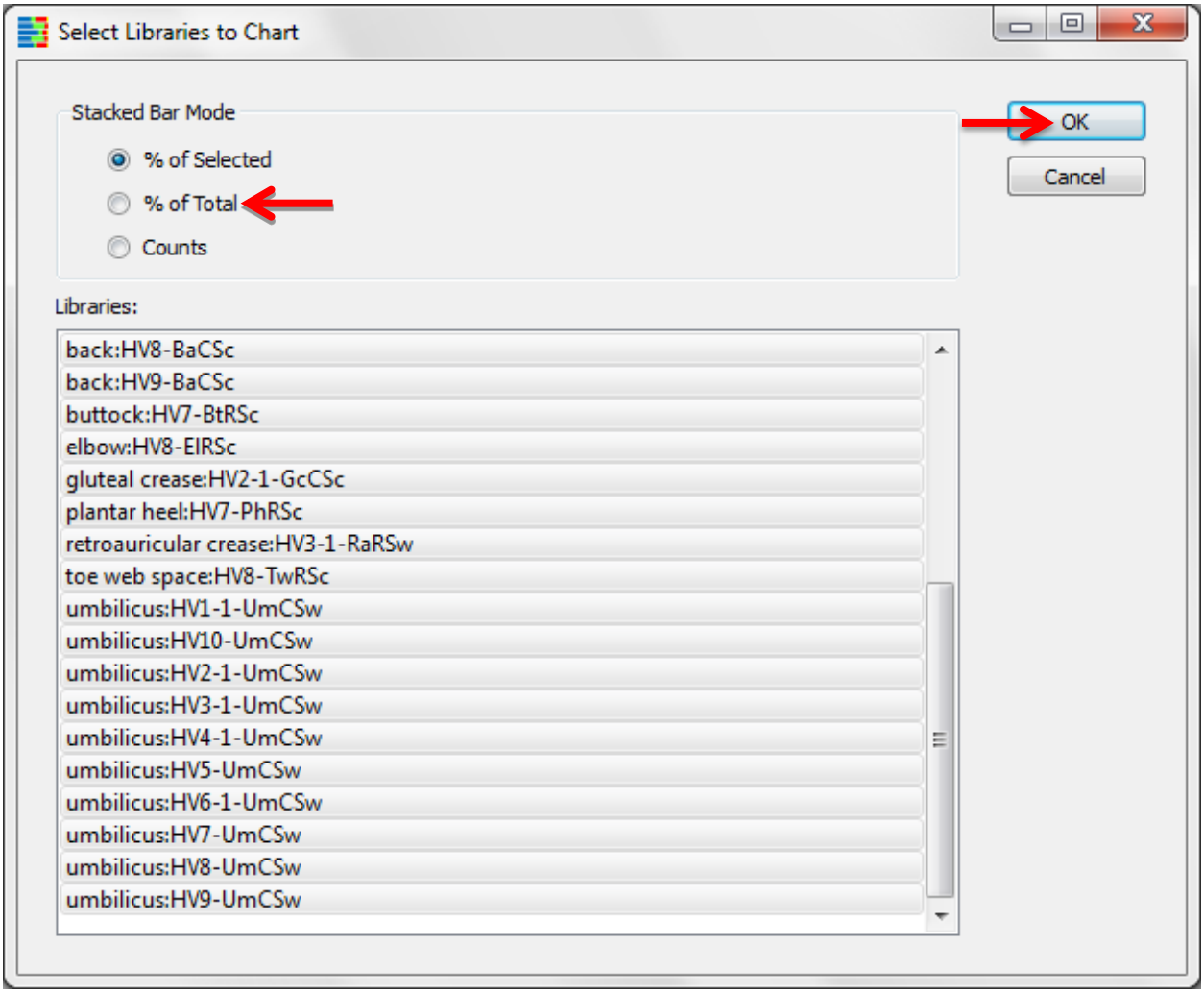


The top 10 OTUs are now highlighted
Click **Plot**

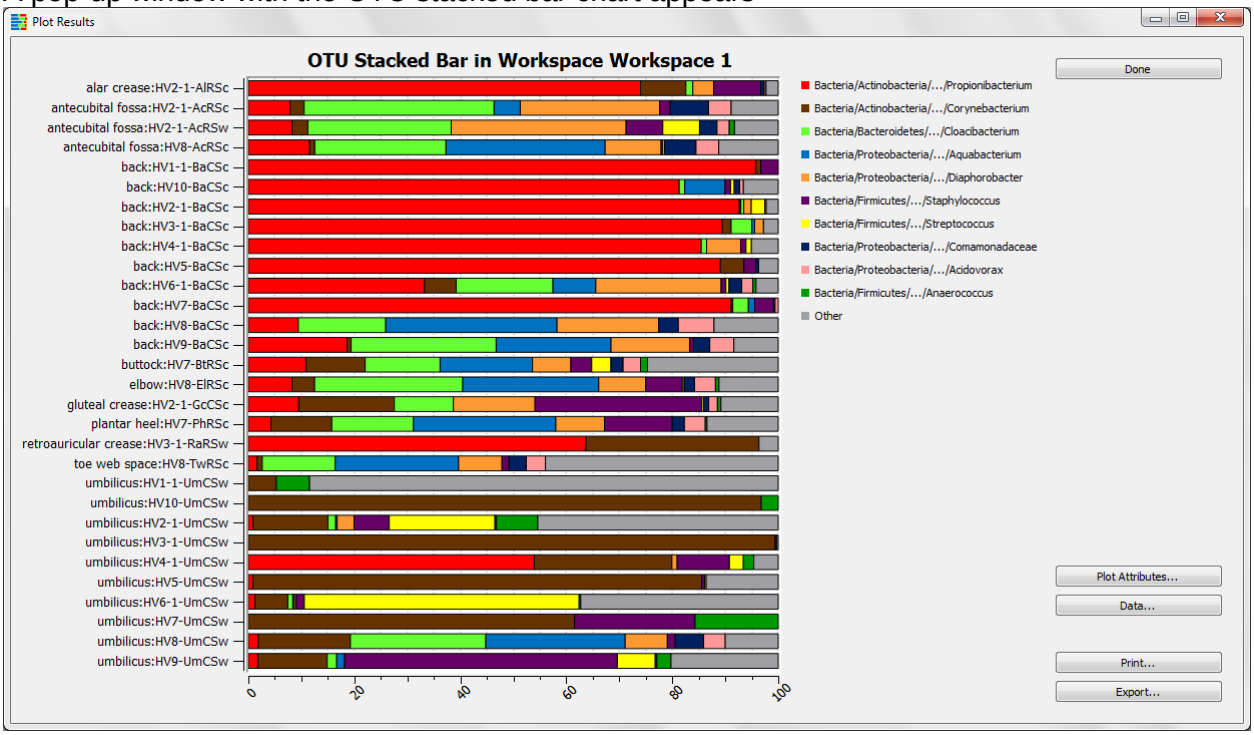


A new window will appear containing stacked bar display options
To create a stacked bar chart which displays a big picture of the project components, select **% of Total**
Click **OK**

© 2012-2013 Charles E. Robertson and Incubix, Inc.



A pop-up window with the OTU stacked bar chart appears

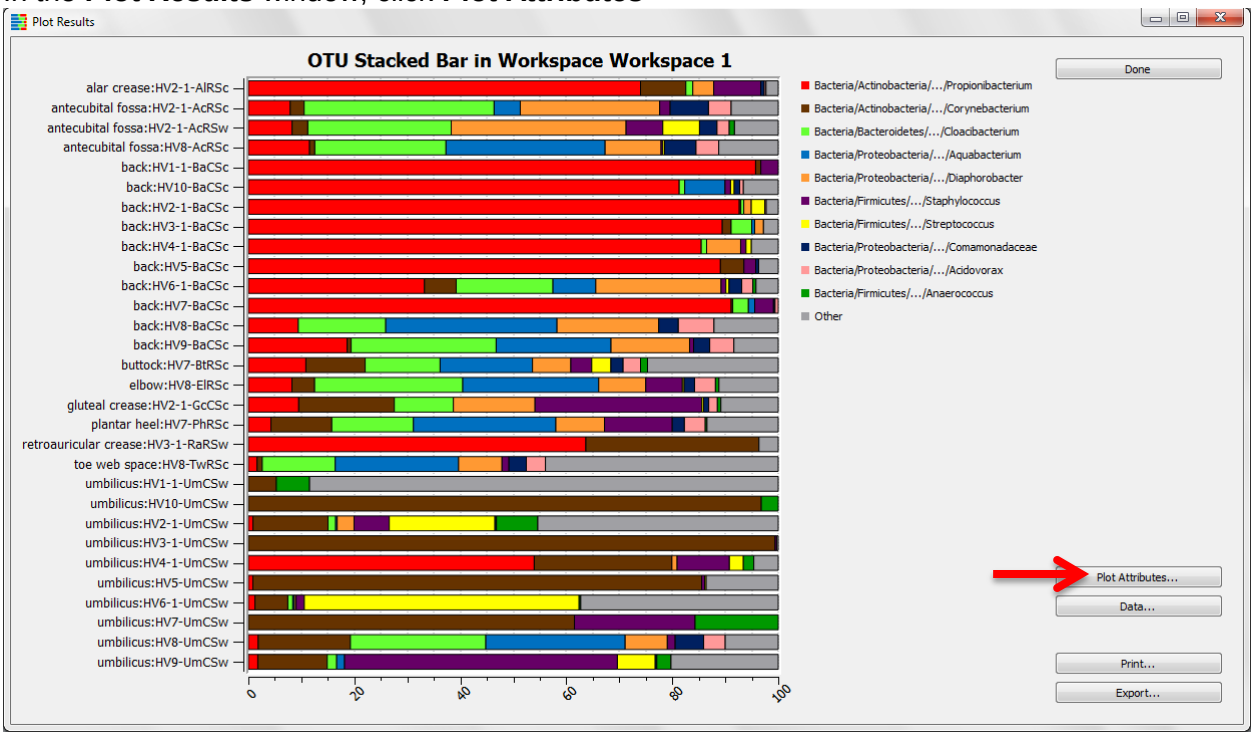


© 2012-2013 Charles E. Robertson and Incubix, Inc.

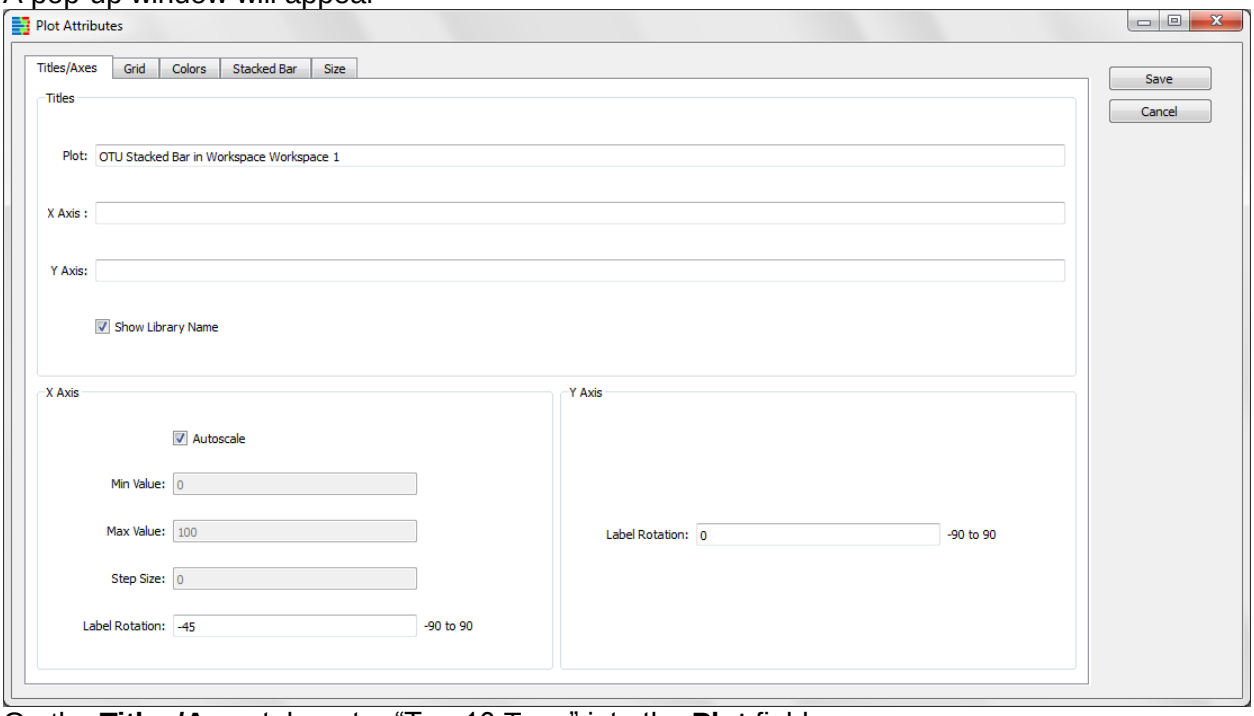
We will now change the default title of the stacked bar chart and add axis labels.

B. Change the Title and Label the Axes

In the **Plot Results** window, click **Plot Attributes**

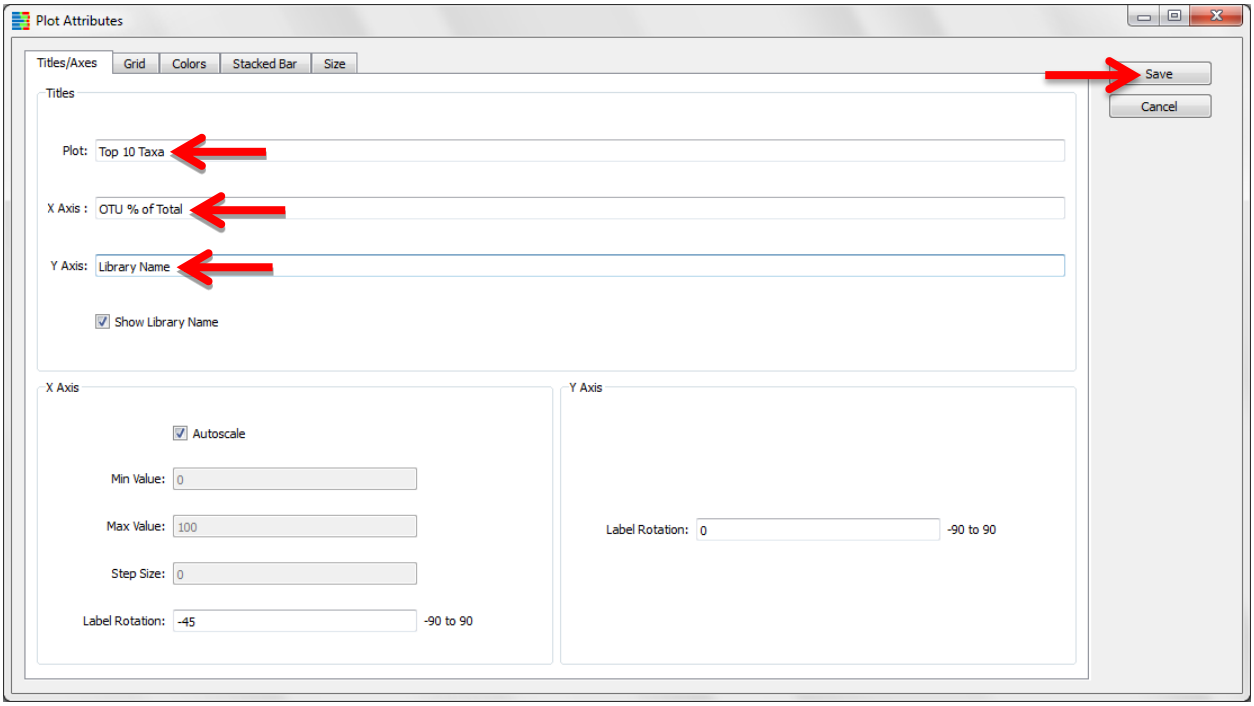


A pop-up window will appear

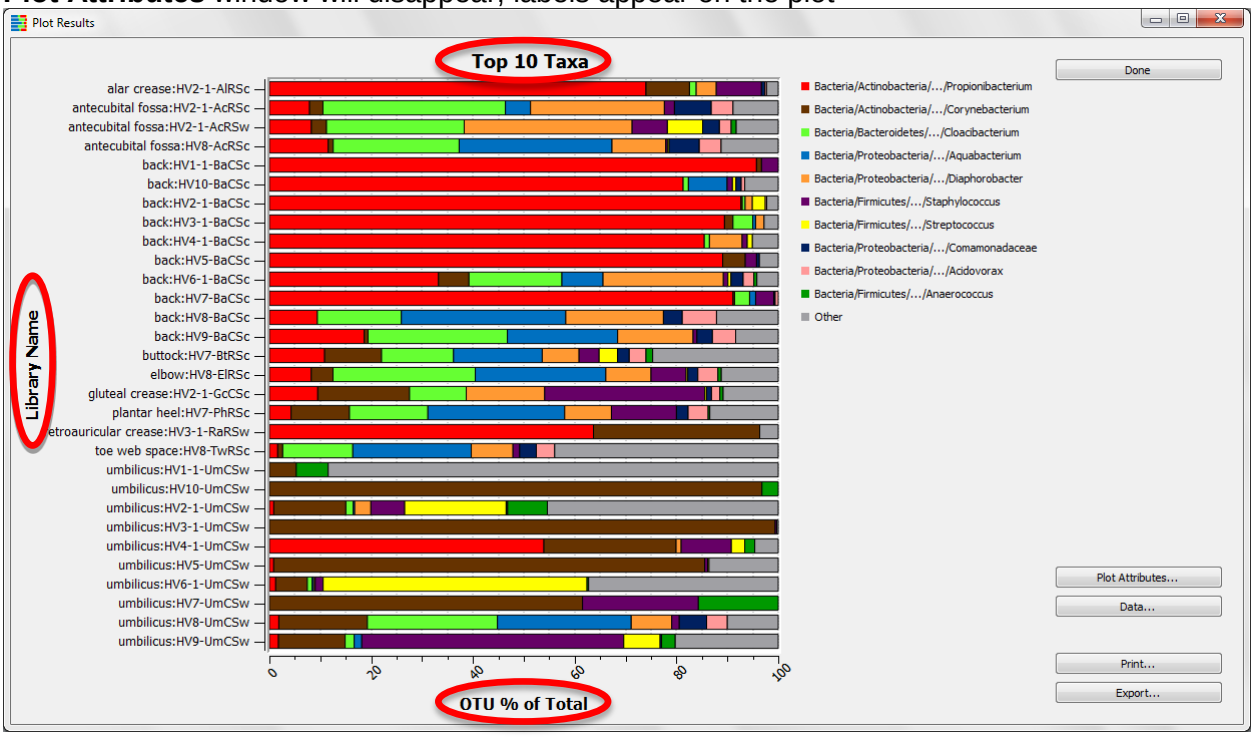


On the **Titles/Axes** tab, enter “Top 10 Taxa” into the **Plot** field
Enter “OTU % of Total” into the **X Axis** field
Enter “Library Name” into the **Y Axis** field
Click **Save**

© 2012-2013 Charles E. Robertson and Incubix, Inc.



Plot Attributes window will disappear; labels appear on the plot



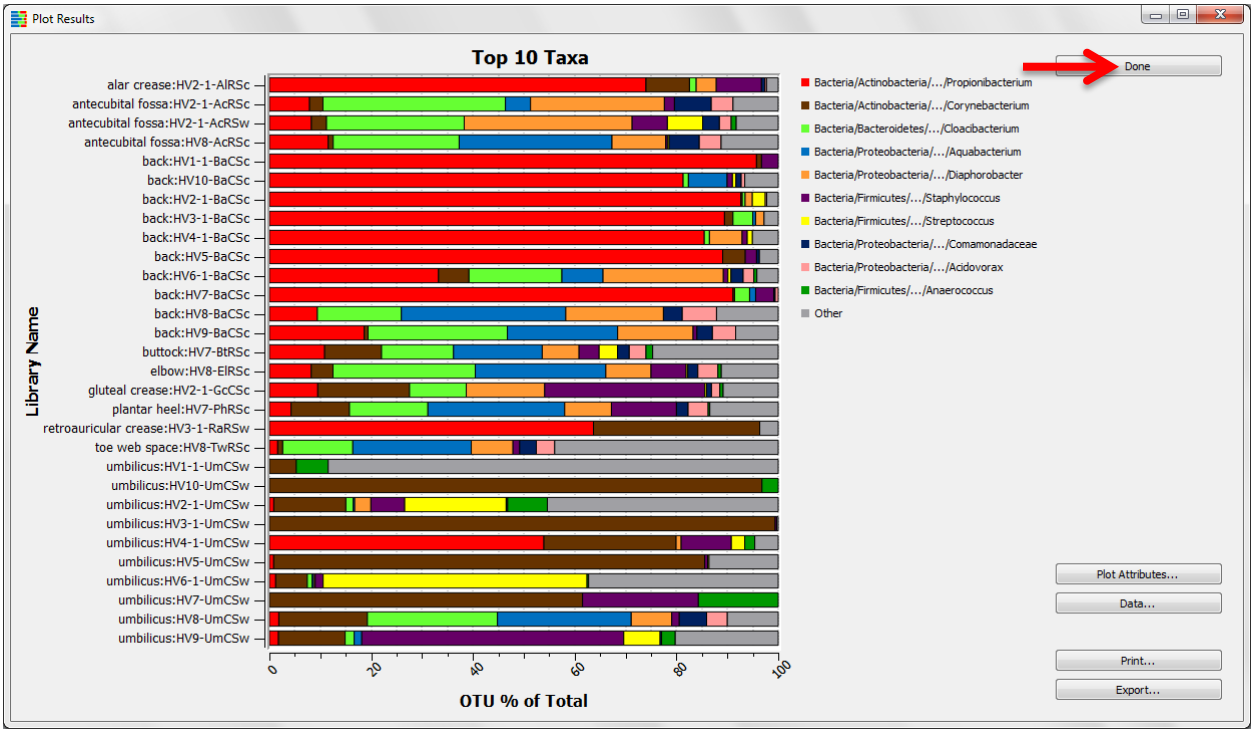
Red and brown appear to be dominant colors in this plot. According to the legend, these colors belong to the "Actinobacteria" phylum. This information may be useful in guiding us toward a hypothesis involving the dominant taxa.

Saving figures in Explicet is easy and convenient. Figures are saved within the larger project, so they stay linked to the data from which they were created and do not create additional files on your computer.

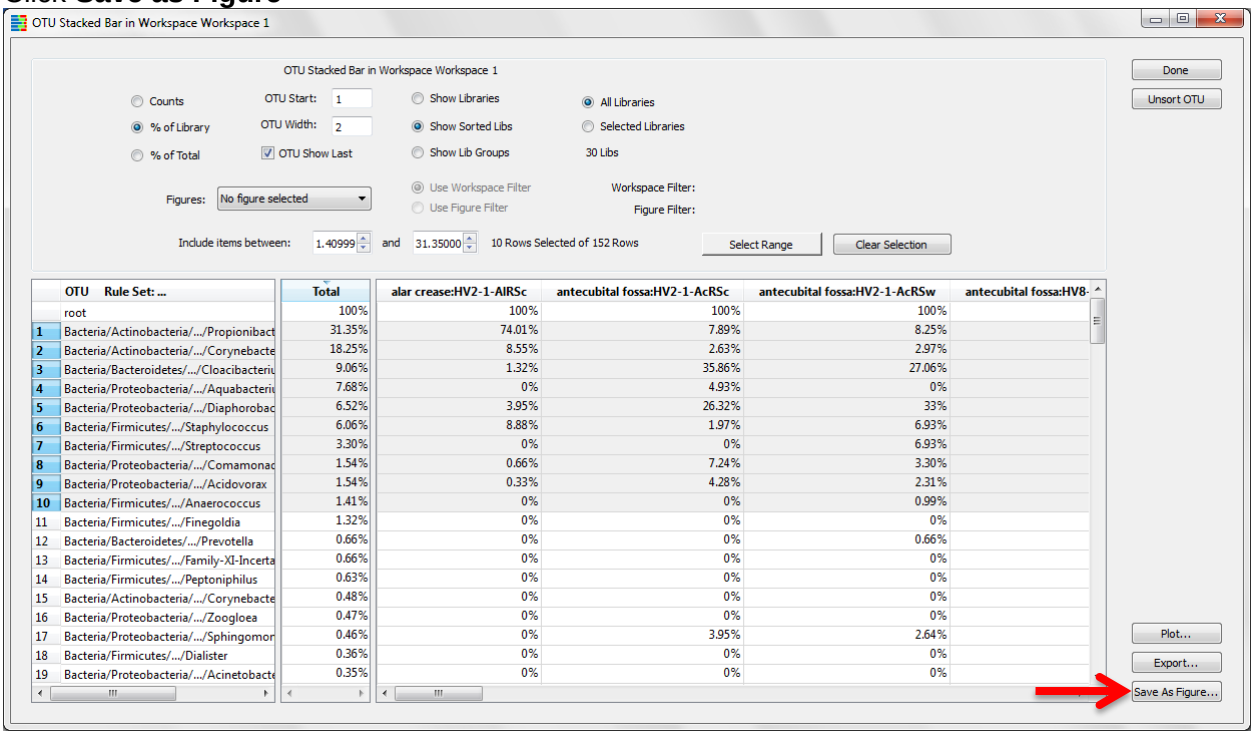
C. Save the OTU Stacked Bar Chart as a Figure

Click **Done** in the stacked bar chart **Plot Results** window

© 2012-2013 Charles E. Robertson and Incubix, Inc.

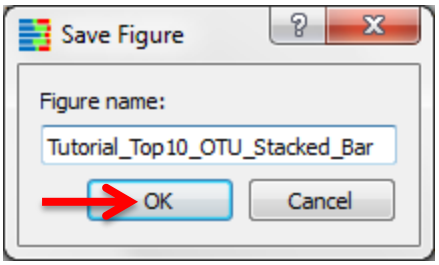


The **OTU Stacked Bar** setup window is back on the screen
Click **Save as Figure**

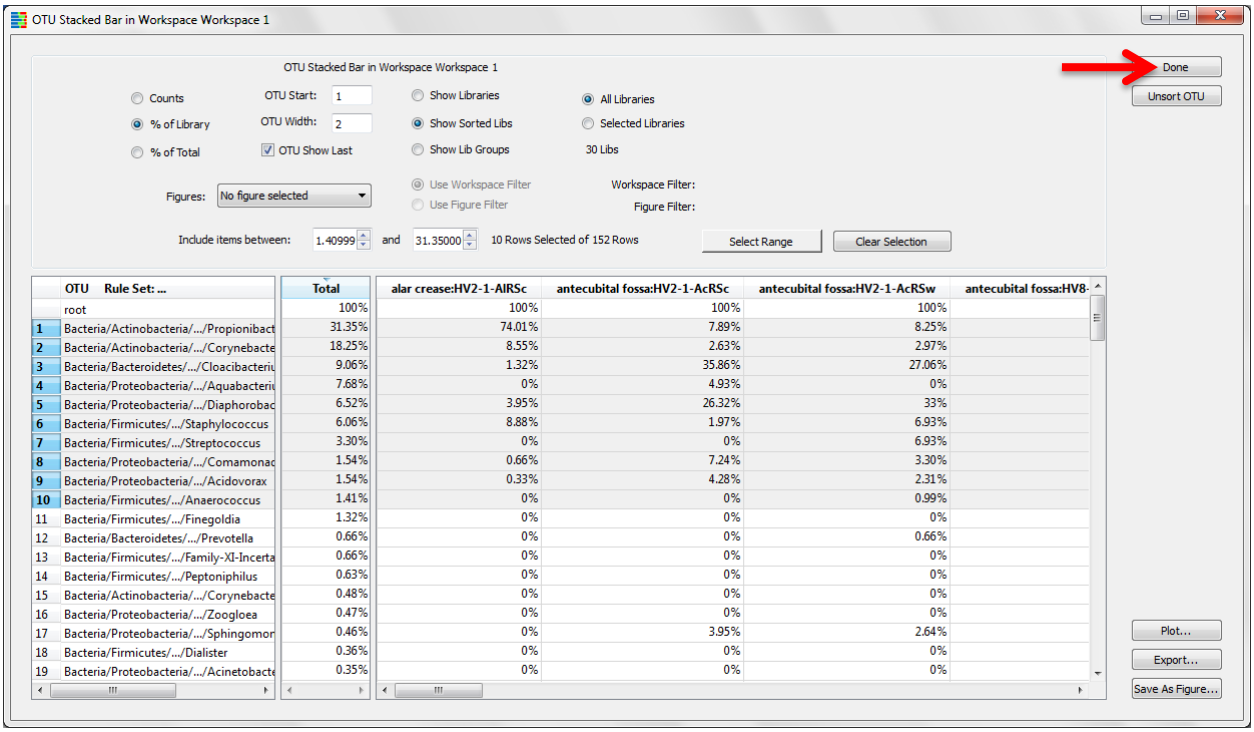


Enter stacked bar chart name in pop-up window
Click **OK**

© 2012-2013 Charles E. Robertson and Incubix, Inc.



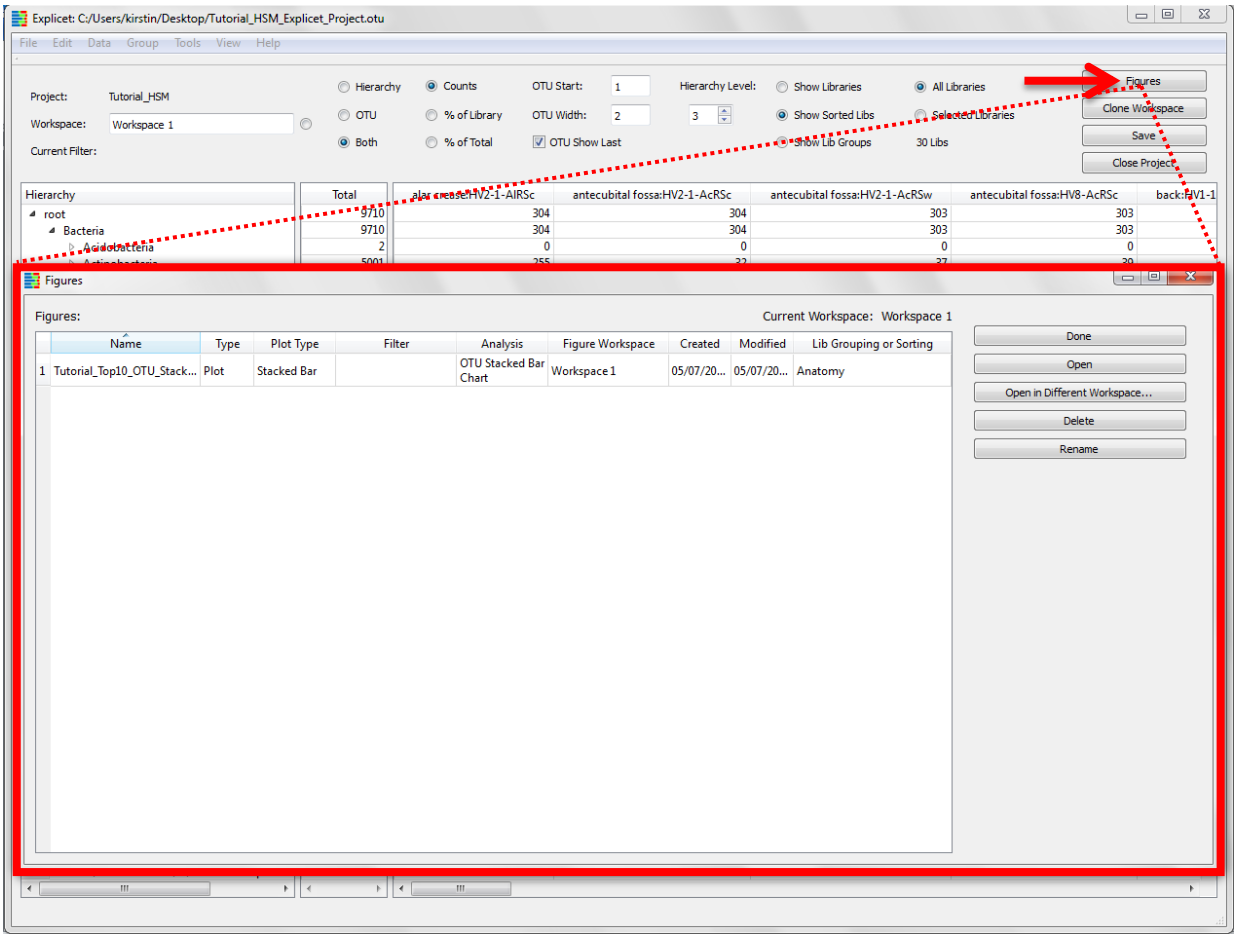
Click **Done**



	OTU Rule Set: ...	Total	alar crease:HV2-1-AIRSc	antecubital fossa:HV2-1-AcRSc	antecubital fossa:HV2-1-AcRSw	antecubital fossa:HV8-
	root	100%	100%	100%	100%	
1	Bacteria/Actinobacteria/.../Propionibact	31.35%	74.01%	7.89%	8.25%	
2	Bacteria/Actinobacteria/.../Corynebacte	18.25%	8.55%	2.63%	2.97%	
3	Bacteria/Bacteroidetes/.../Cloacibacteriu	9.06%	1.32%	35.86%	27.06%	
4	Bacteria/Proteobacteria/.../Aquabacteriu	7.68%	0%	4.93%	0%	
5	Bacteria/Proteobacteria/.../Diaphorobac	6.52%	3.95%	26.32%	33%	
6	Bacteria/Firmicutes/.../Staphylococcus	6.06%	8.88%	1.97%	6.93%	
7	Bacteria/Firmicutes/.../Streptococcus	3.30%	0%	0%	6.93%	
8	Bacteria/Proteobacteria/.../Comamonac	1.54%	0.66%	7.24%	3.30%	
9	Bacteria/Proteobacteria/.../Acidovorax	1.54%	0.33%	4.28%	2.31%	
10	Bacteria/Firmicutes/.../Anaerococcus	1.41%	0%	0%	0.99%	
11	Bacteria/Firmicutes/.../Finegoldia	1.32%	0%	0%	0%	
12	Bacteria/Bacteroidetes/.../Prevotella	0.66%	0%	0%	0.66%	
13	Bacteria/Firmicutes/.../Family-XI-Incerta	0.66%	0%	0%	0%	
14	Bacteria/Firmicutes/.../Peptoniphilus	0.63%	0%	0%	0%	
15	Bacteria/Actinobacteria/.../Corynebacte	0.48%	0%	0%	0%	
16	Bacteria/Proteobacteria/.../Zoogloea	0.47%	0%	0%	0%	
17	Bacteria/Proteobacteria/.../Sphingomor	0.46%	0%	3.95%	2.64%	
18	Bacteria/Firmicutes/.../Dialister	0.36%	0%	0%	0%	
19	Bacteria/Proteobacteria/.../Acinetobacte	0.35%	0%	0%	0%	

Once saved, the stacked bar chart and associated figure data can be recalled at any point by clicking the **Figures** button on the workspace window. This provides a convenient mechanism for editing figures during manuscript preparation. Figures can also be exported in a format suitable for further modification in dedicated drawing software.

© 2012-2013 Charles E. Robertson and Incubix, Inc.



© 2012-2013 Charles E. Robertson and Incubix, Inc.

VII. Make a Pie Chart

Another useful way to generate an overview of the organisms that exist in the dataset is through a pie chart, which allows graphical depictions of the taxonomic hierarchy.

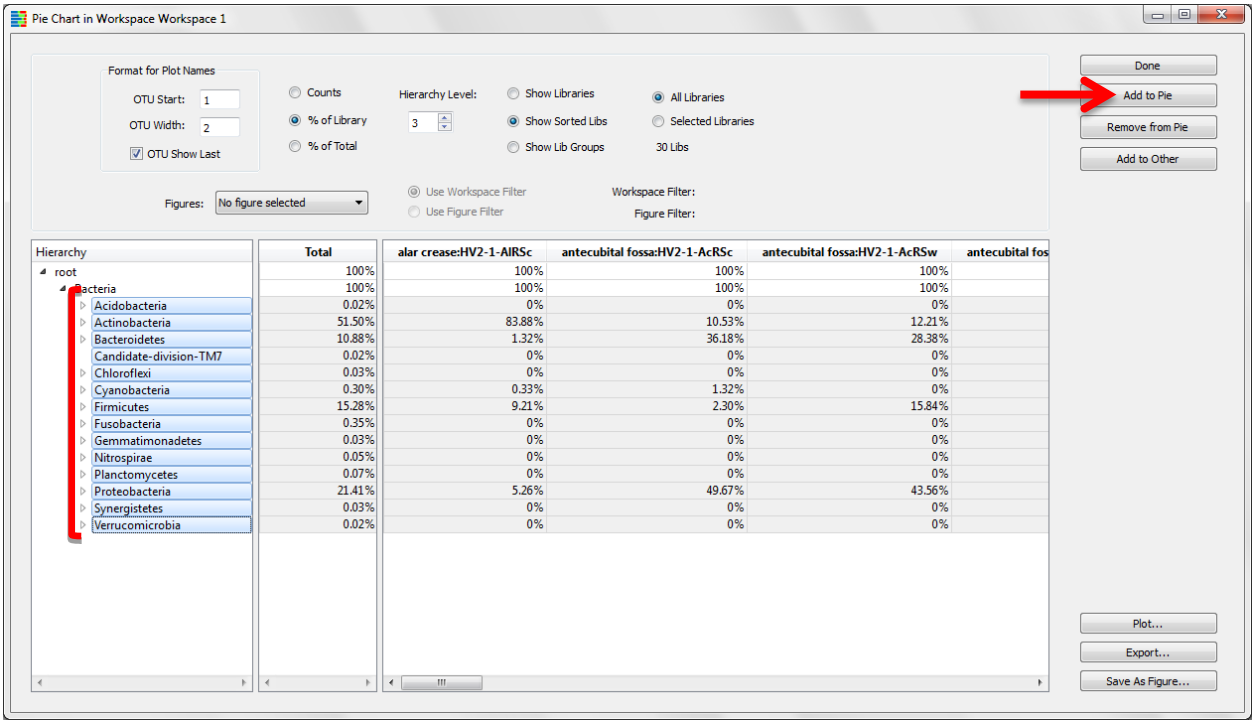
A. Create a Pie Chart of the Project Components

Tools → Plot → Pie Chart

A new window will appear with the hierarchical data available in the workspace

Shift-click all of the phyla in the list

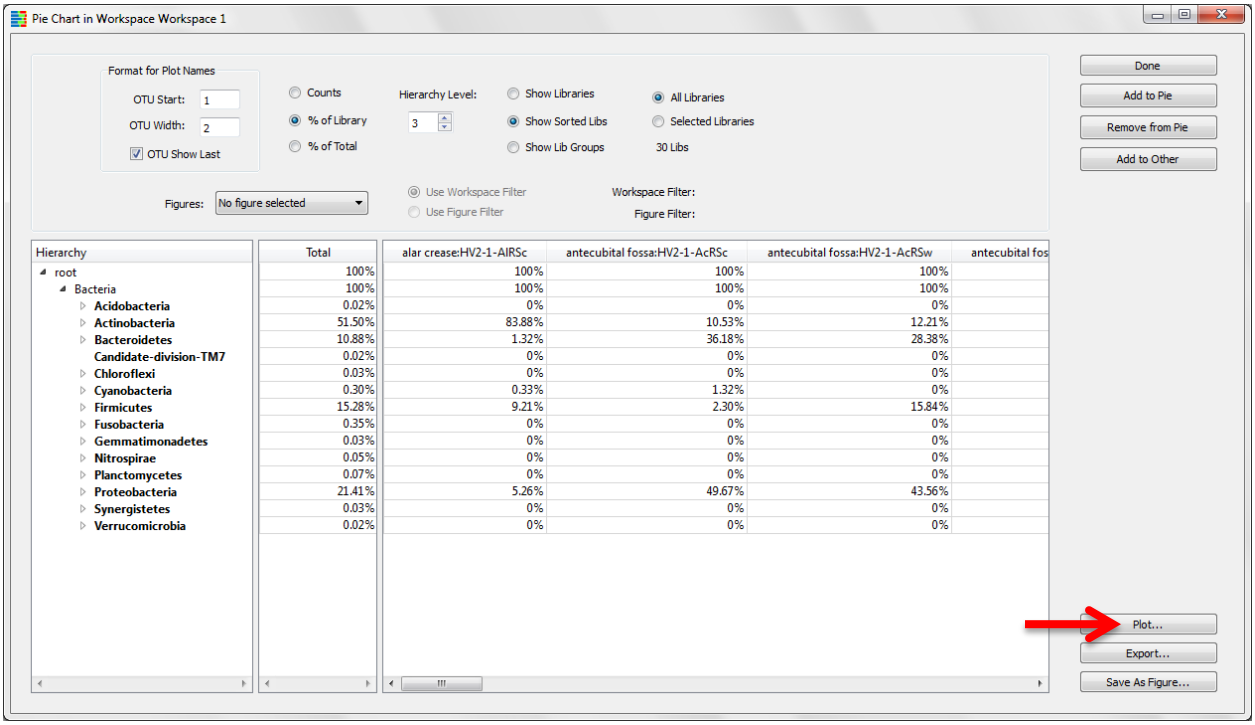
Click **Add to Pie**



The selected phyla which were added to the pie are now bold

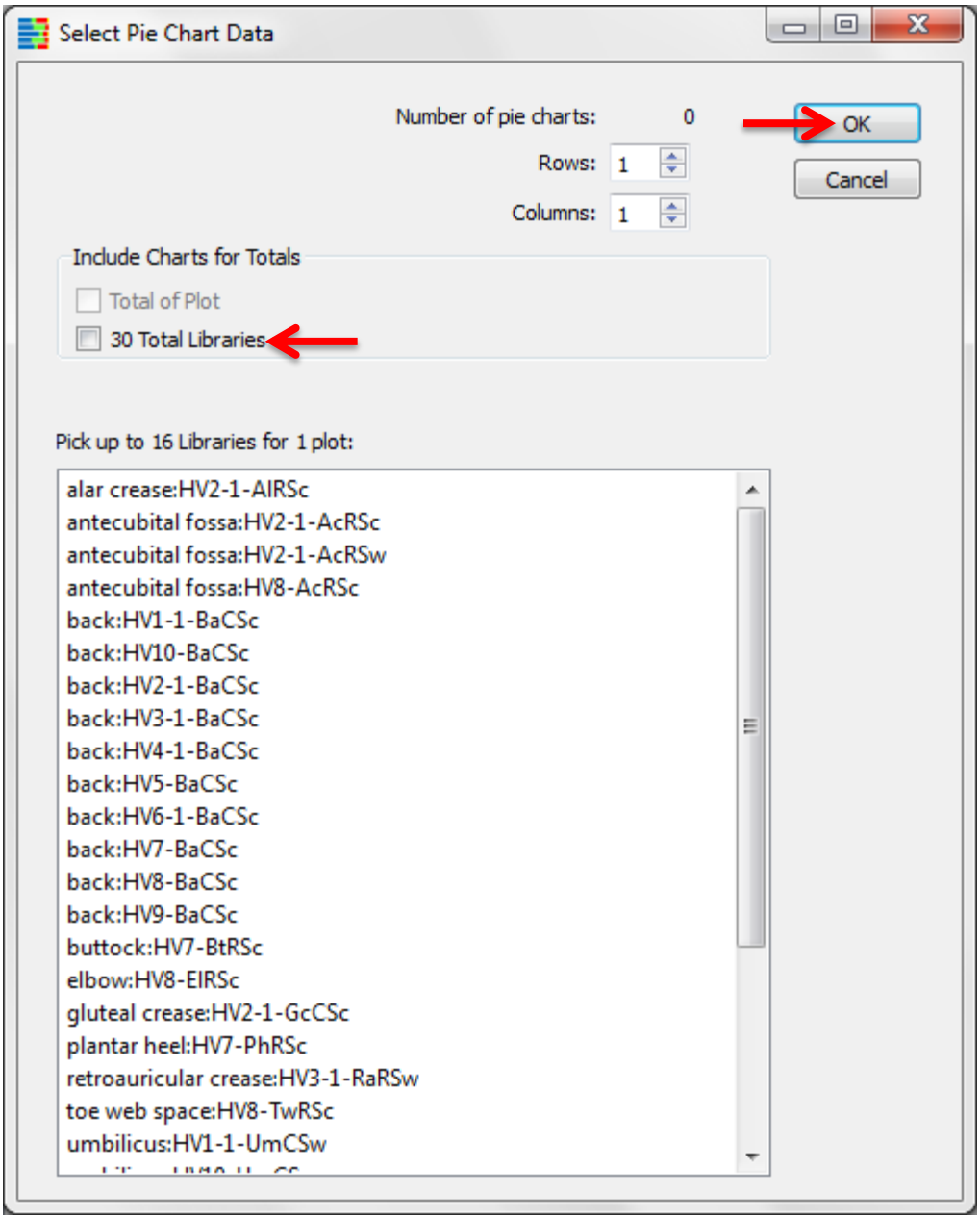
Click **Plot**

© 2012-2013 Charles E. Robertson and Incubix, Inc.



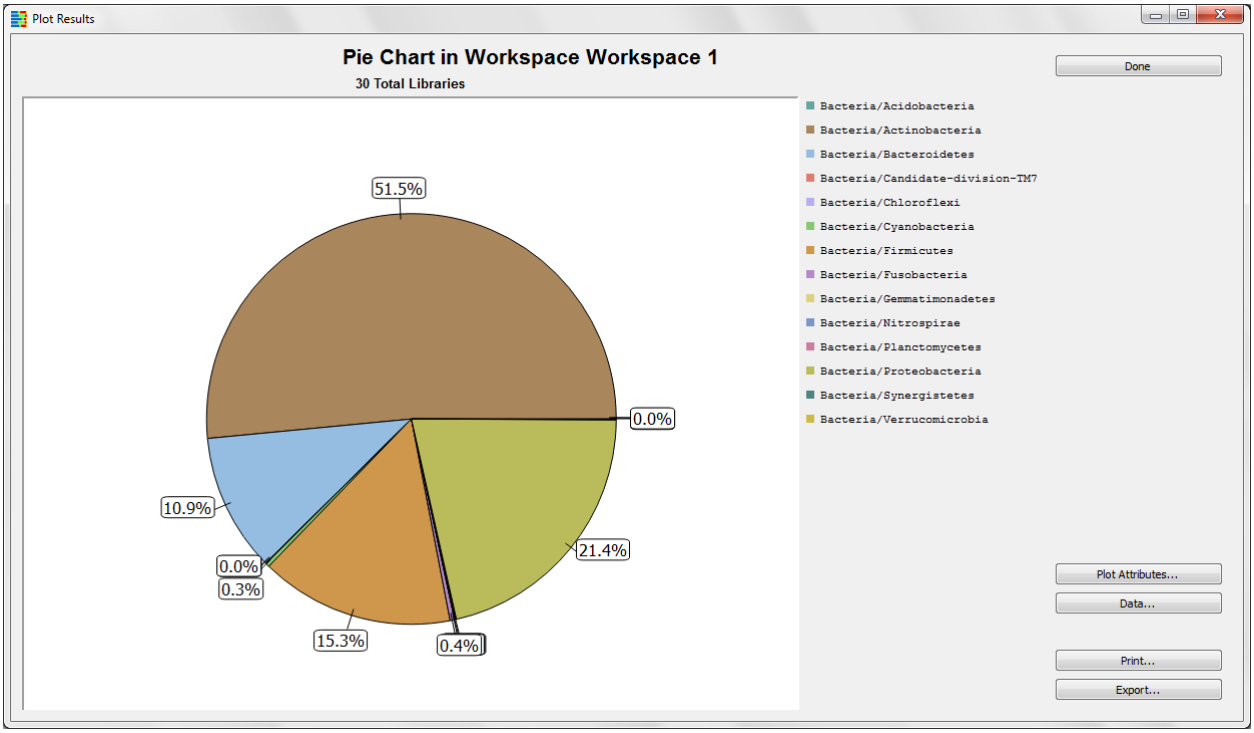
A new window will appear containing pie chart display options
To create only a single pie chart displaying the combined libraries' data, select **30 Total Libraries**
Click **OK**

© 2012-2013 Charles E. Robertson and Incubix, Inc.



A pop-up window with the pie chart appears

© 2012-2013 Charles E. Robertson and Incubix, Inc.

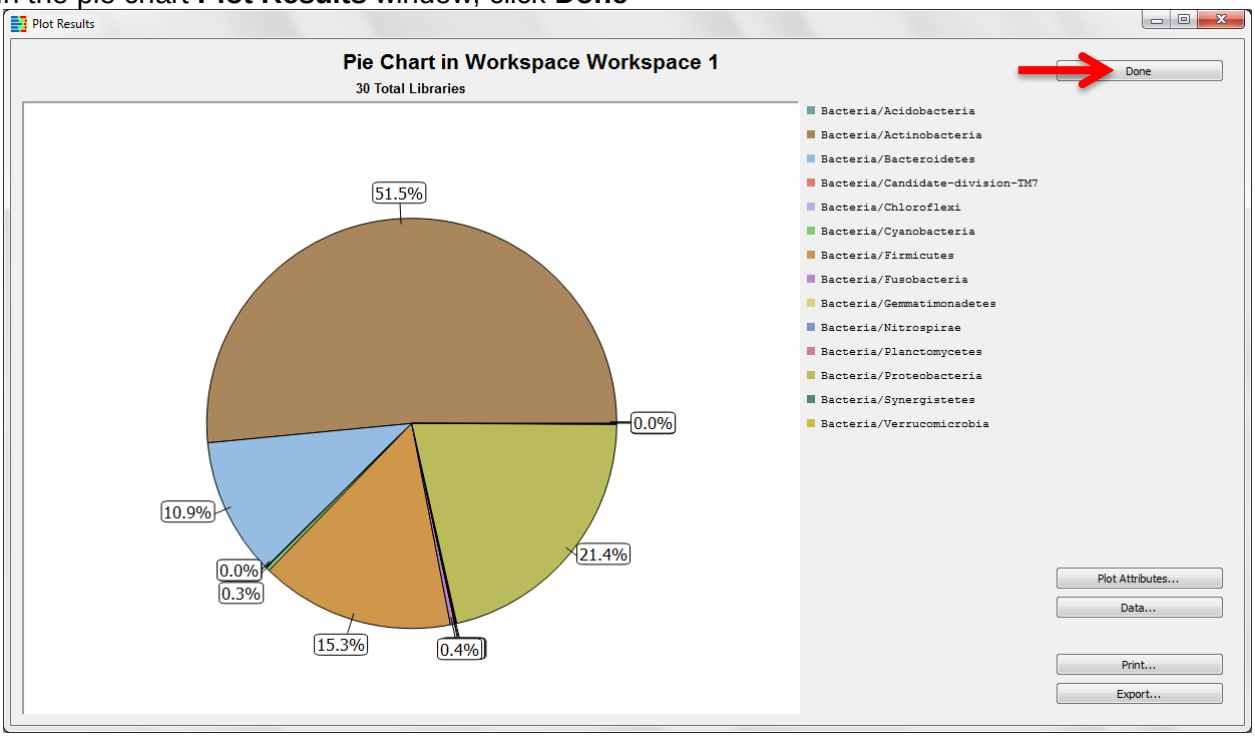


By looking at the pie chart of the phyla, it is clear that the brown wedge, Actinobacteria, is the most prevalent phylum in the data.

Additionally, we can see that the green wedge, Proteobacteria, makes up the second largest portion of the total. To visualize the classes present within the Proteobacteria phylum, we can create pie chart sub-wedges.

B. Make a Pie Chart with Sub-Wedges

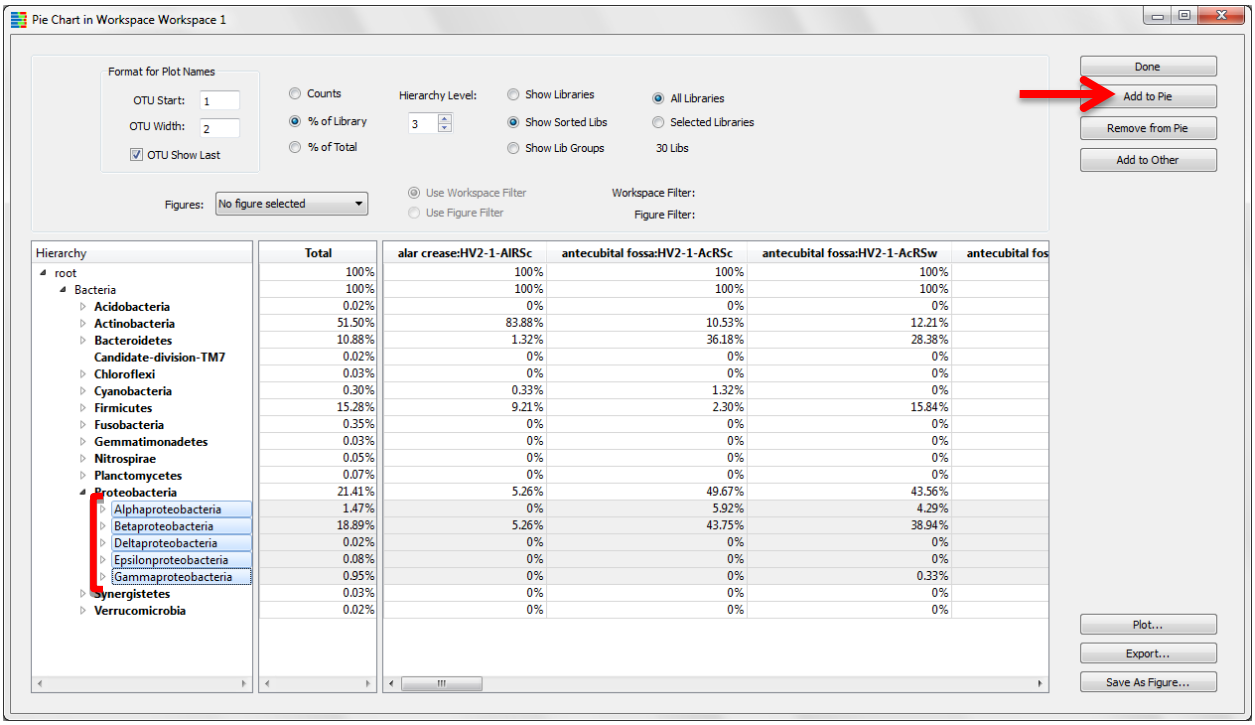
In the pie chart **Plot Results** window, click **Done**



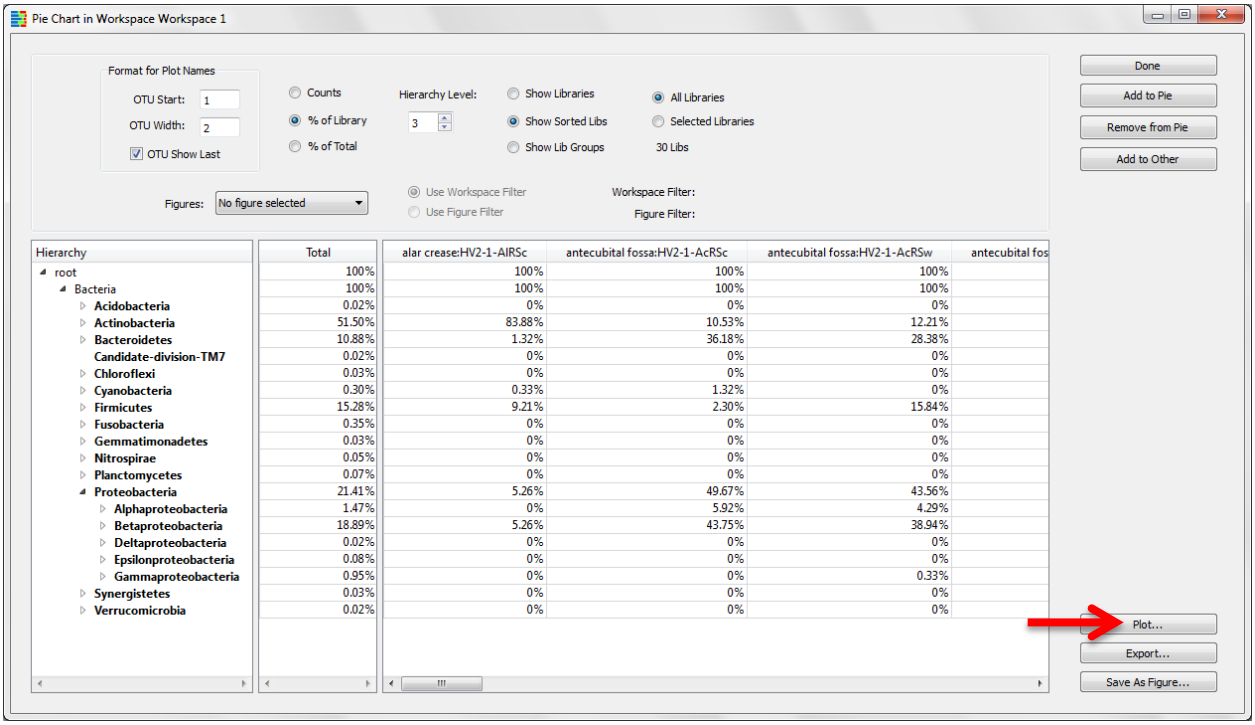
The **Taxonomy Pie Chart** setup window is back on the screen
Use the drop down arrow to the left of “Proteobacteria” to find the classes within the phylum

© 2012-2013 Charles E. Robertson and Incubix, Inc.

Shift-click all of the classes in the list
Click **Add to Pie**

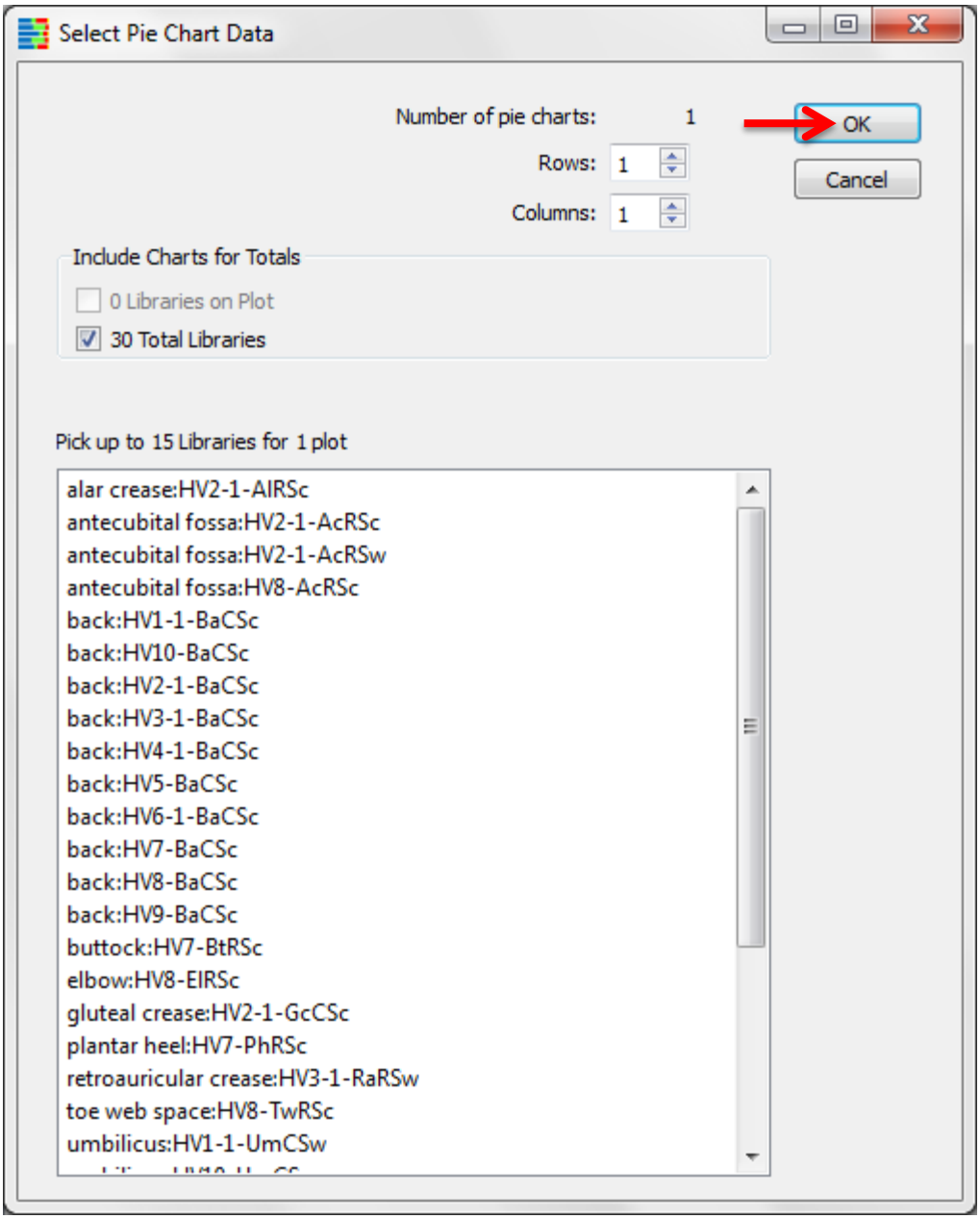


The selected classes that were added to the pie are now bold
Click **Plot**



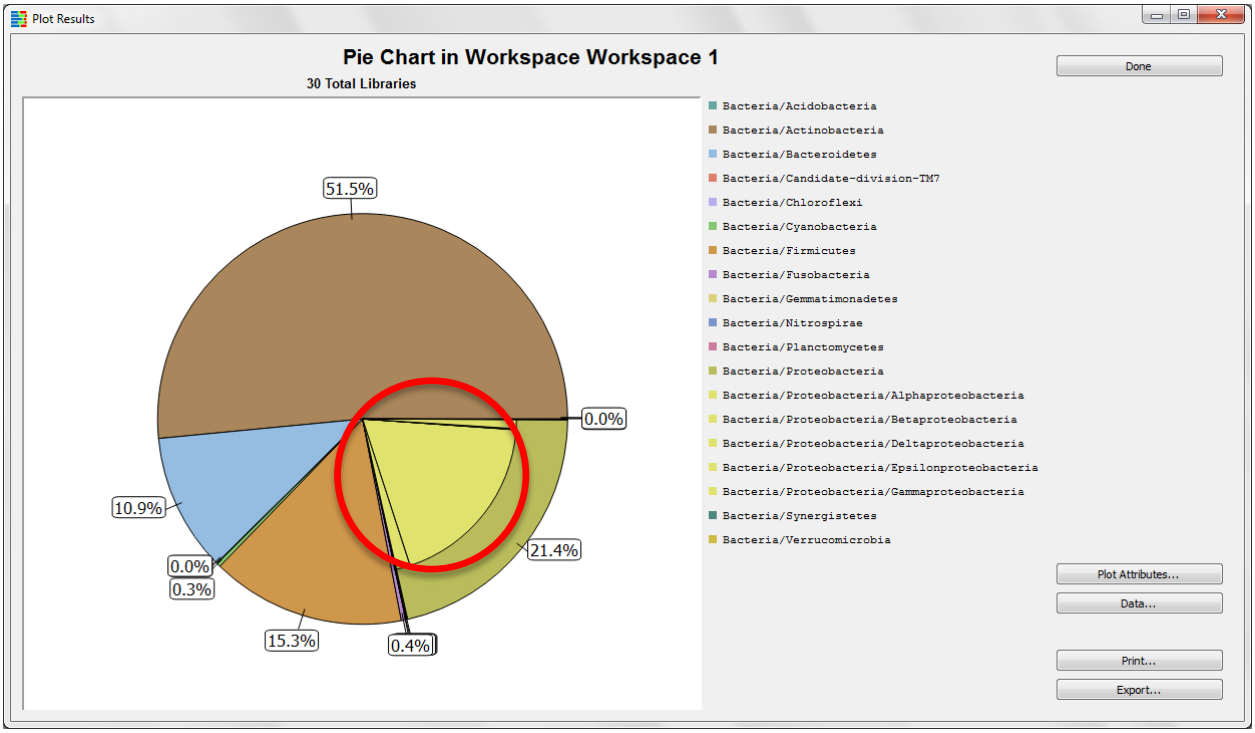
A new window will appear containing pie chart display options
Again, we will create only a single pie chart displaying the combined libraries' data, so click **OK**

© 2012-2013 Charles E. Robertson and Incubix, Inc.



A pop-up window with the pie chart appears. We now see the classes within Proteobacteria represented as sub-wedges.

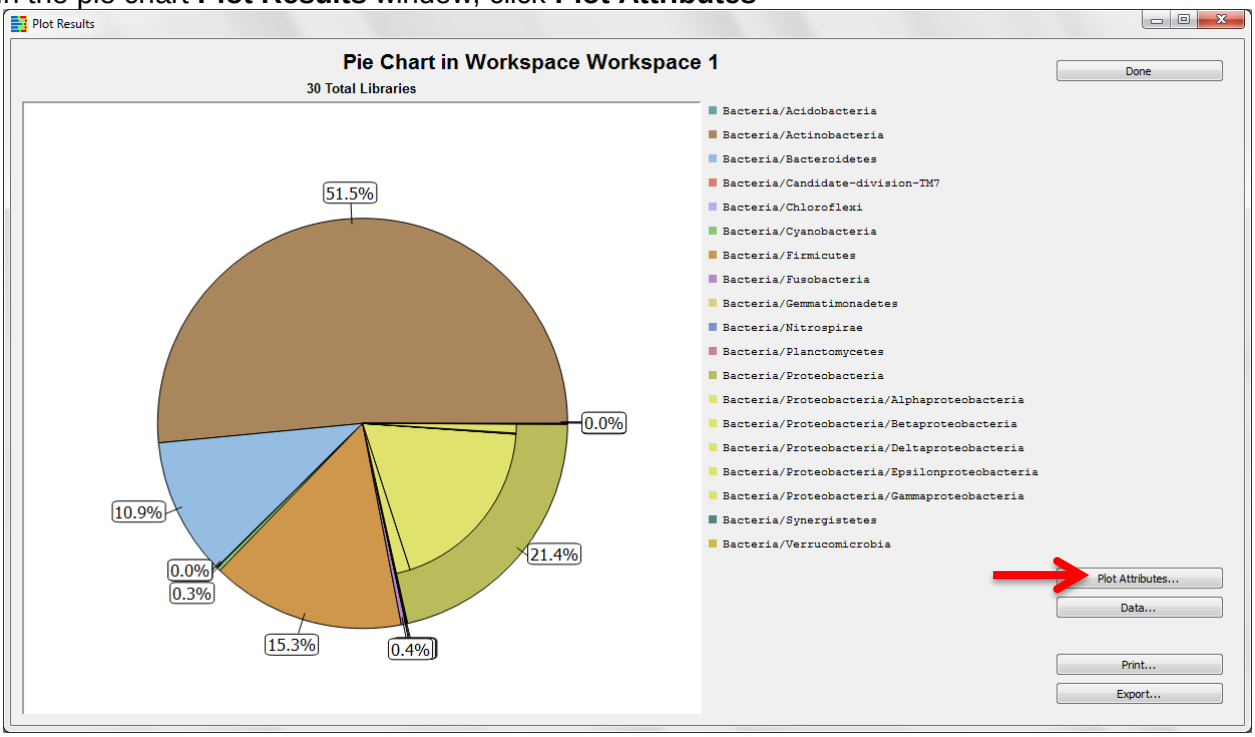
© 2012-2013 Charles E. Robertson and Incubix, Inc.



In order to better differentiate between the different classes, we can change the color of the sub-wedges.

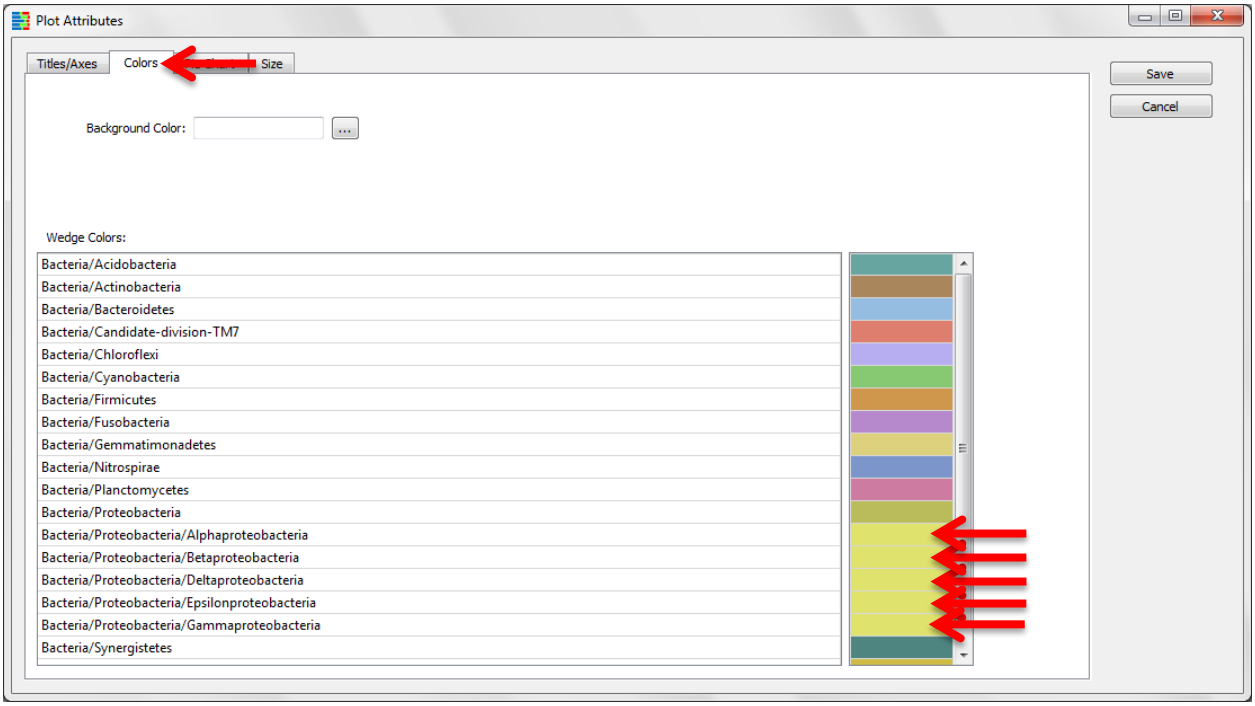
C. Change Wedge Colors in the Pie Chart

In the pie chart **Plot Results** window, click **Plot Attributes**

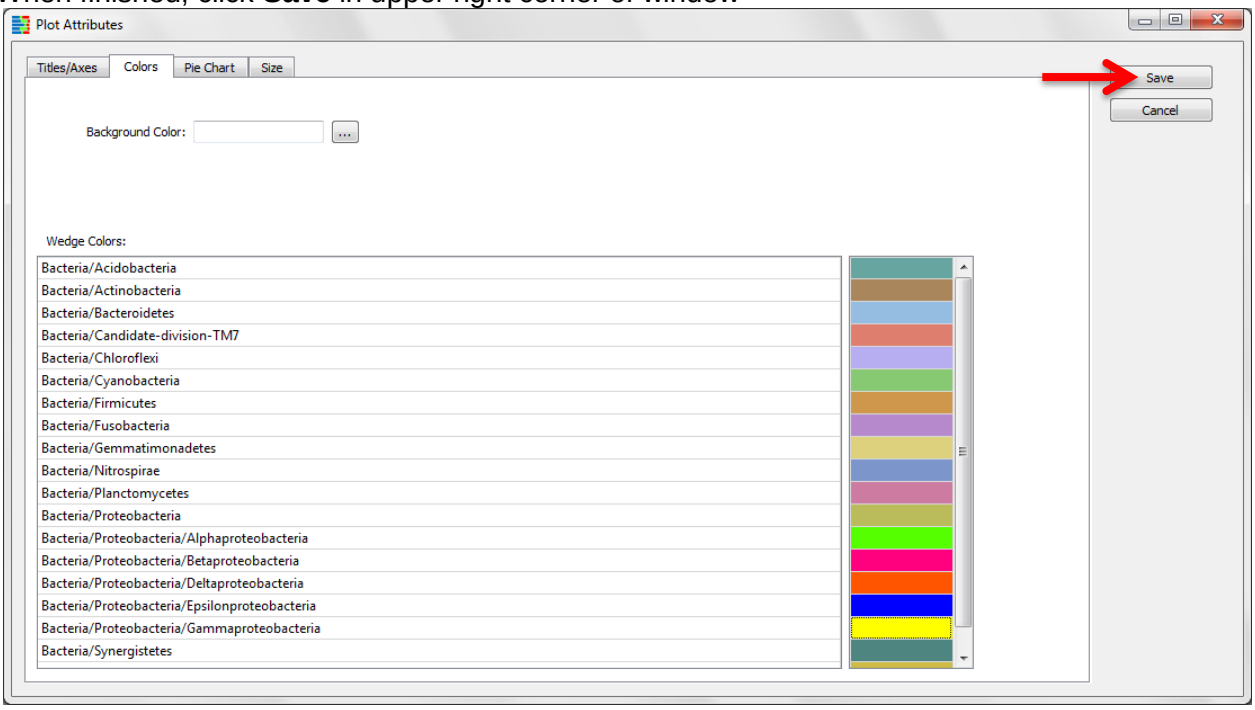


A pop-up window will appear
Click on the **Colors** tab
To pick a different wedge color, click on the color, and select a new color from the pop-up display

© 2012-2013 Charles E. Robertson and Incubix, Inc.

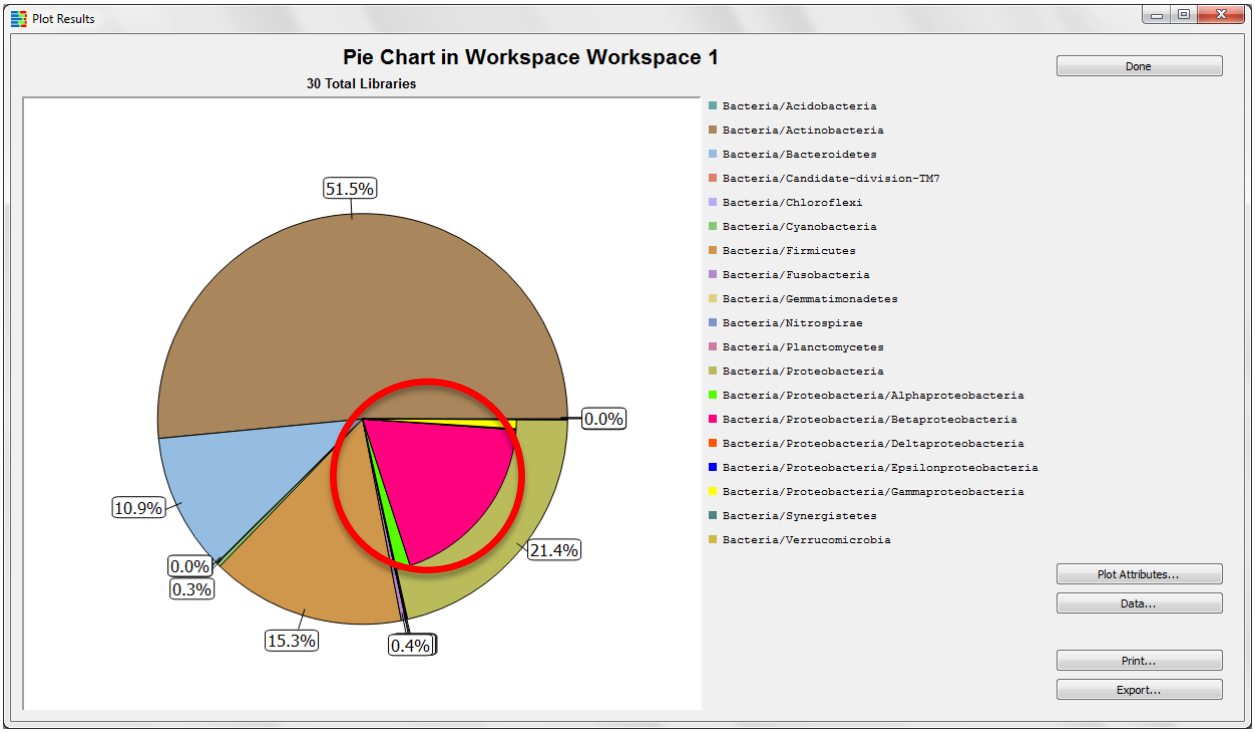


When finished, click **Save** in upper right corner of window



Plot Attributes window will disappear; changes will be shown on the plot

© 2012-2013 Charles E. Robertson and Incubix, Inc.



You may choose to save the pie chart as a figure. To do so, continue as shown earlier in the stacked bar chart example; close the graphics window, and select **Save As Figure** in the **Pie Chart** window.

© 2012-2013 Charles E. Robertson and Incubix, Inc.

VIII. Create a Workspace

A workspace is a way for users to make experiments on copies or subsets of their entire data set, while keeping the original data fully intact.

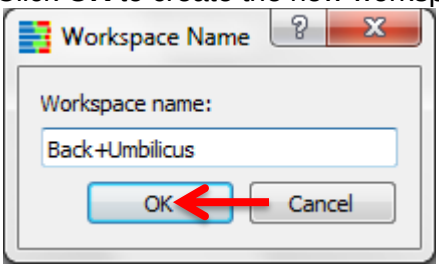
Although the skin is a single organ, it harbors microbial communities that live in a range of physiologically and topographically distinct niches. The back is typically a sebaceous region, whereas the umbilicus is often a moist region of the body. Therefore, these two niches may have different taxa present. We will create a workspace for a mini-experiment to compare data from only these two anatomical positions.

A. Create a New Workspace

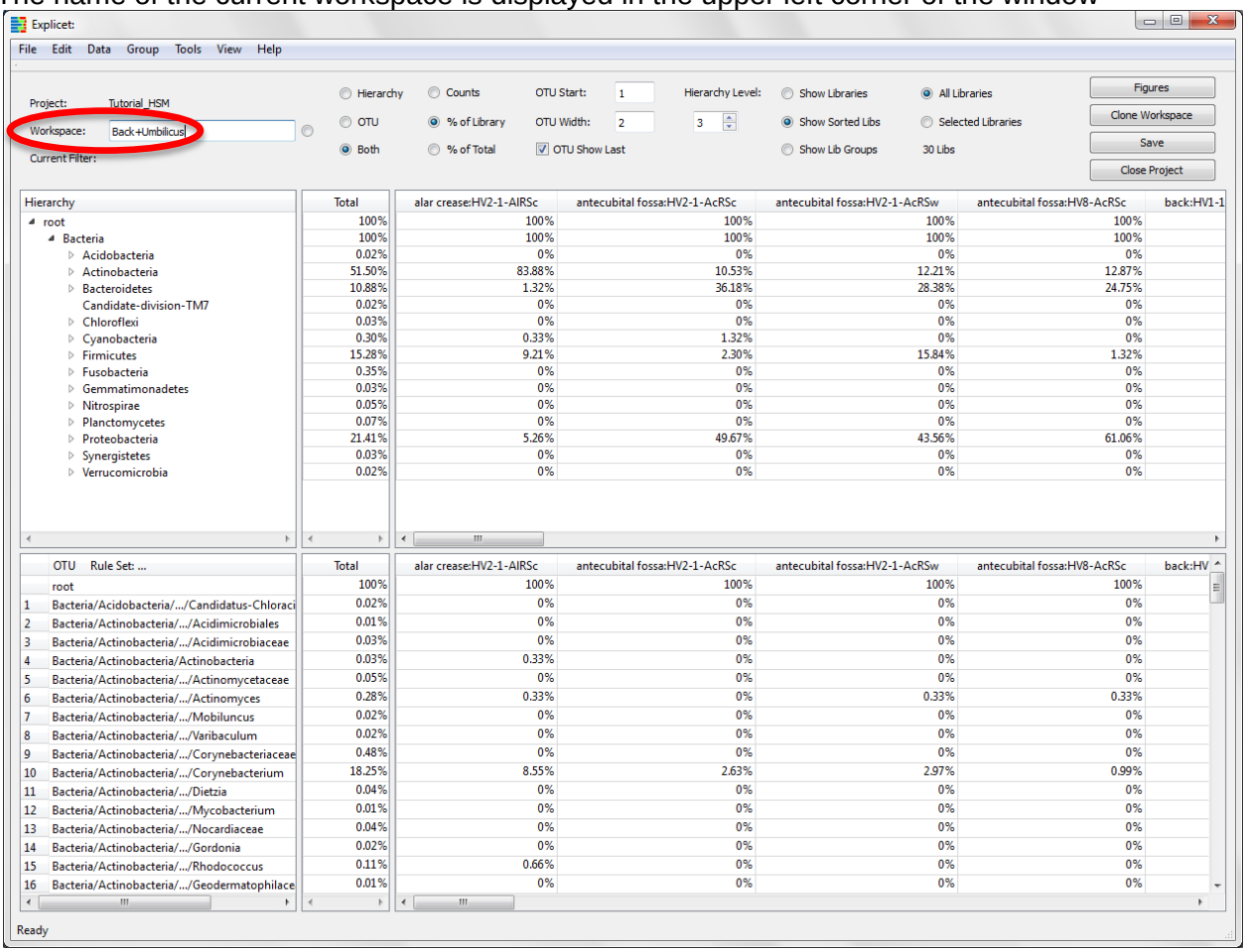
File → New → Workspace from Current Workspace

- "from Current Workspace" allows us to copy all of the display changes we've already made to the new workspace.

Enter desired workspace name in the pop-up window
Click **OK** to create the new workspace



The name of the current workspace is displayed in the upper left corner of the window



© 2012-2013 Charles E. Robertson and Incubix, Inc.

IX. Apply a Filter

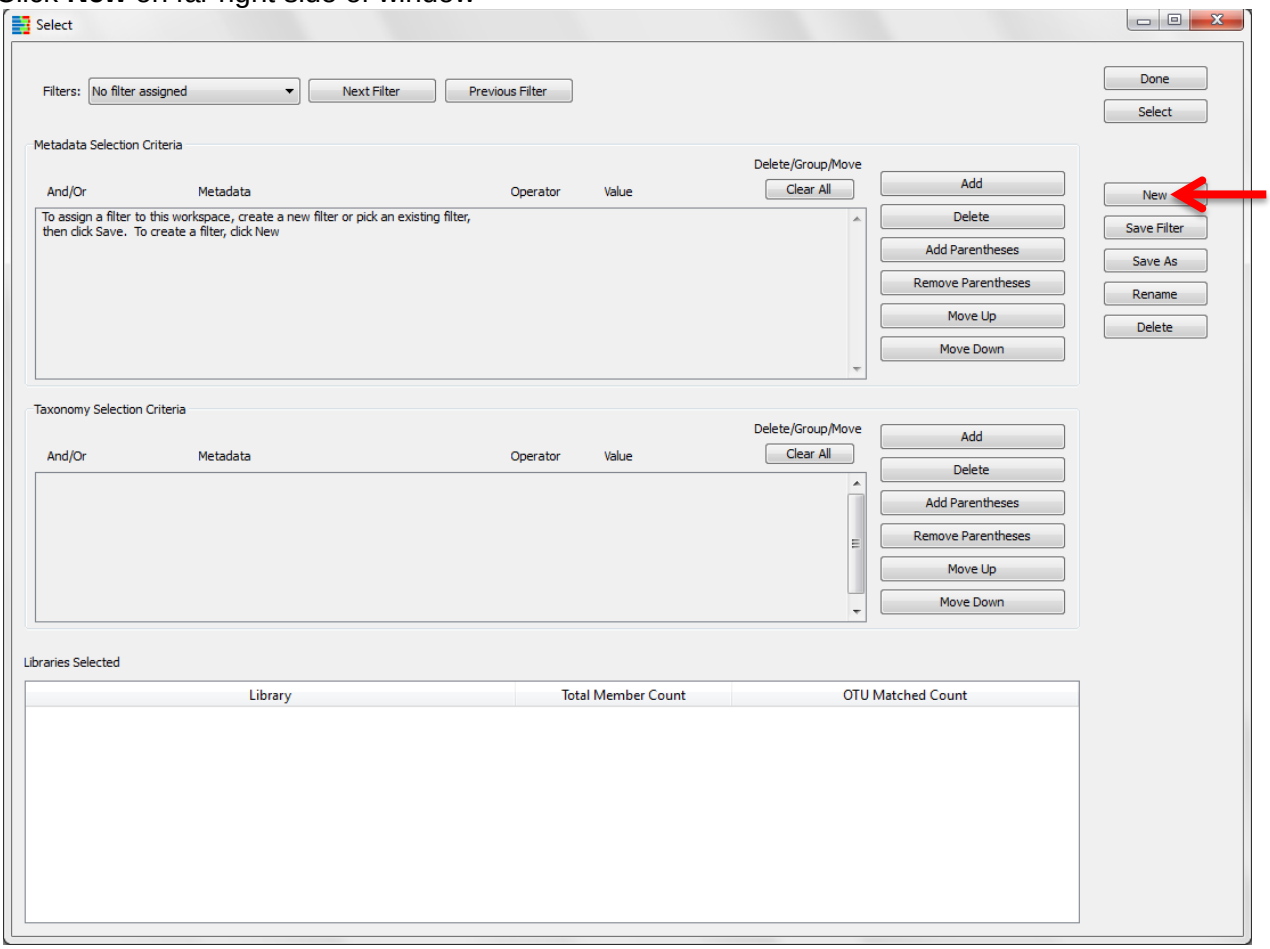
To compare data from only the back and umbilicus, we need to separate these libraries from the other body parts. This is done in Explicet via “filters”.

A. Create a Filter

Data → Select Libraries

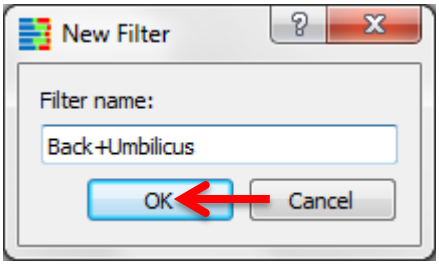
New pop-up window appears for creation of filters

Click **New** on far right side of window



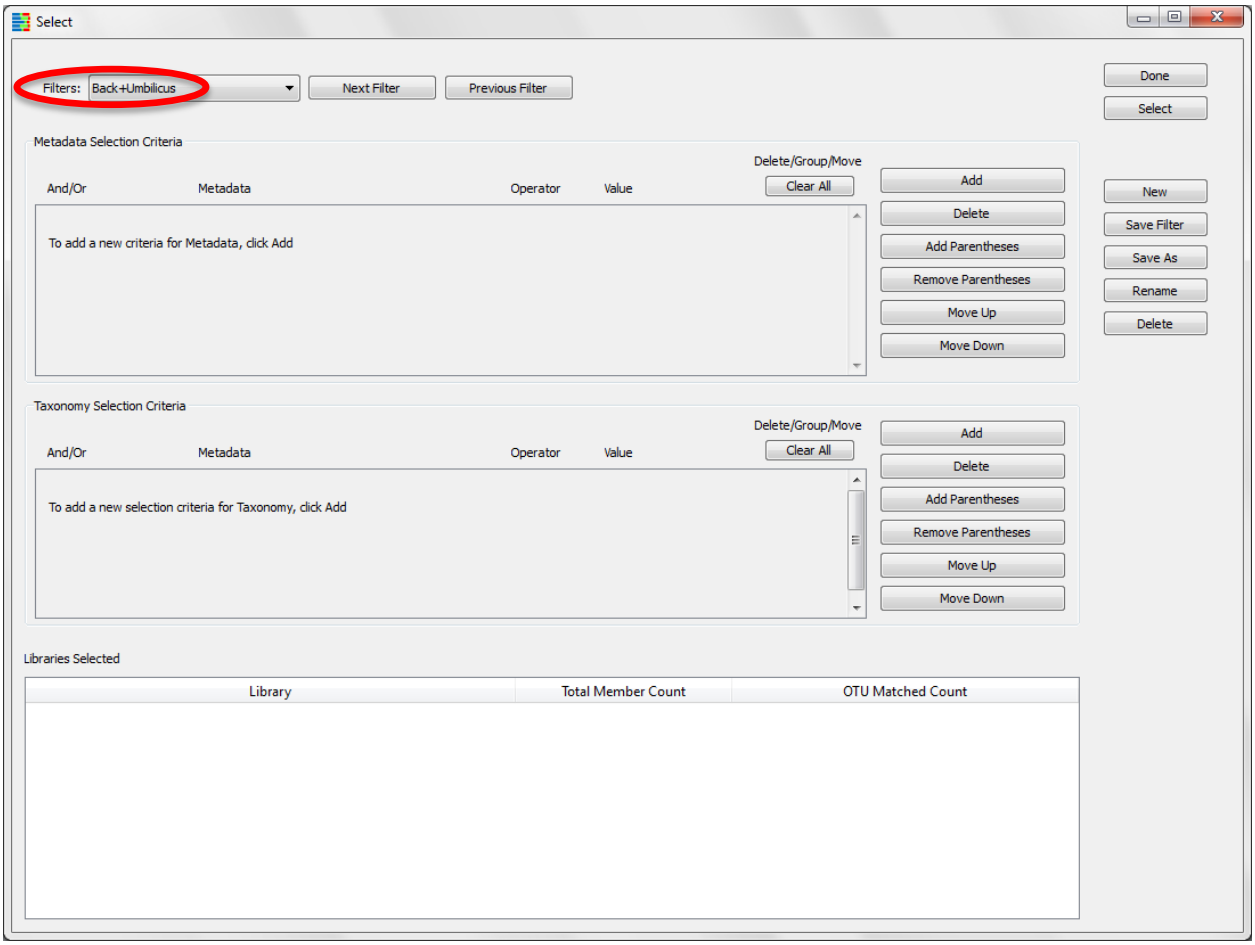
Enter desired filter name in the pop-up window

Click **OK**



The filter name will appear in upper left corner of window

© 2012-2013 Charles E. Robertson and Incubix, Inc.

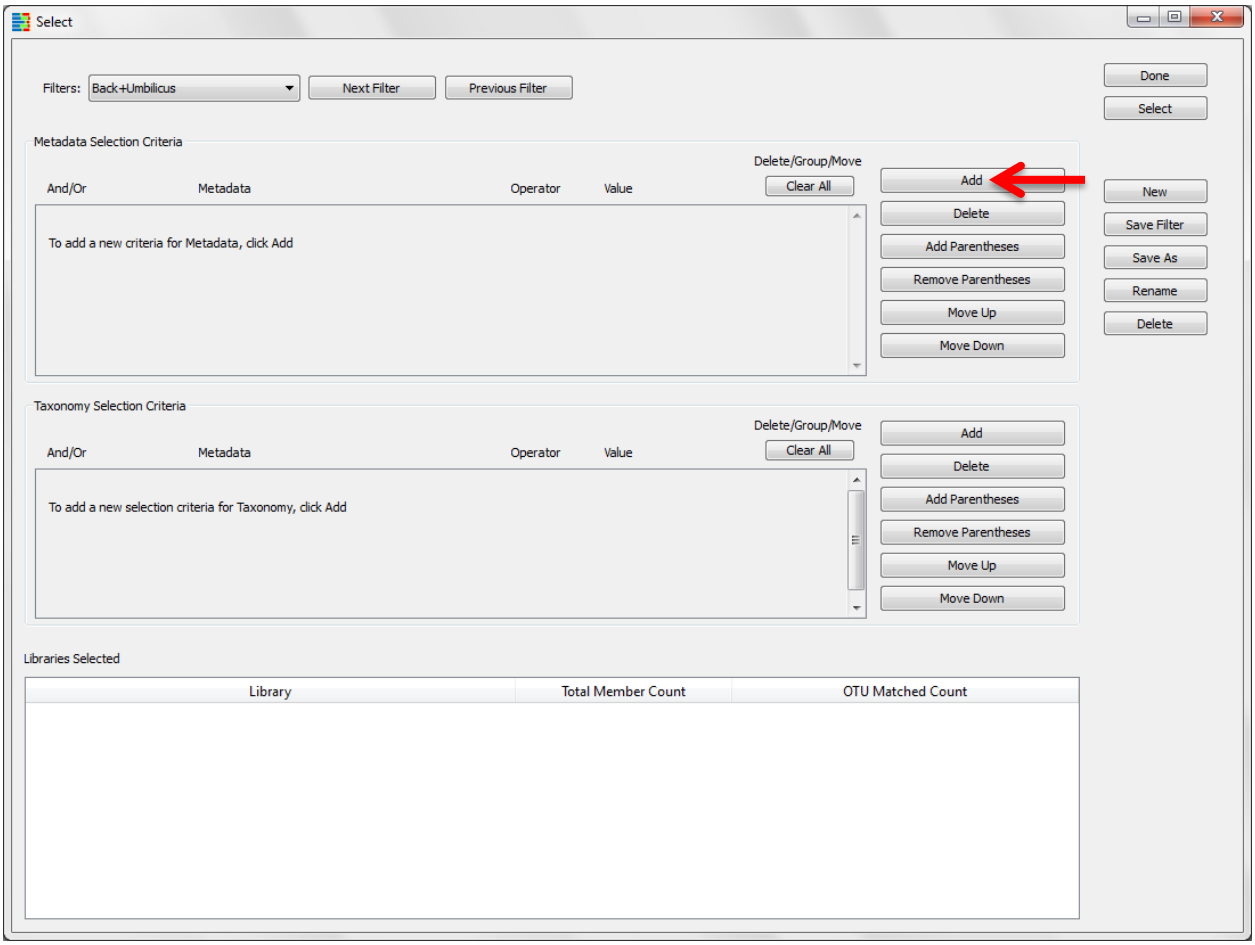


Now that we have created a new filter, we need to set up the parameters to filter by. We will select for all libraries that were sampled from the "back" or "umbilicus" anatomical sites.

B. Set Up the Filter Parameters

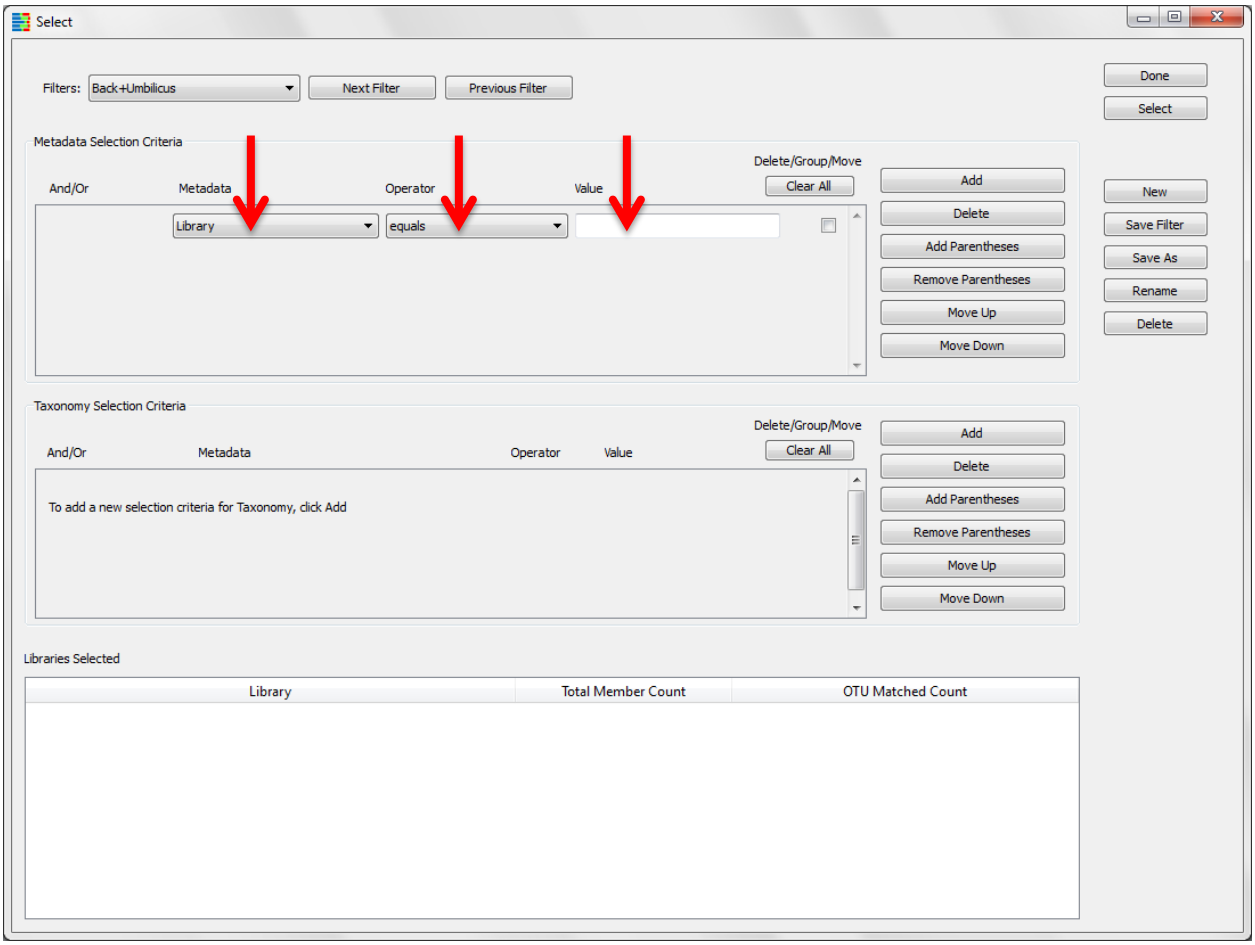
Click **Add** in the Metadata Criteria pane

© 2012-2013 Charles E. Robertson and Incubix, Inc.



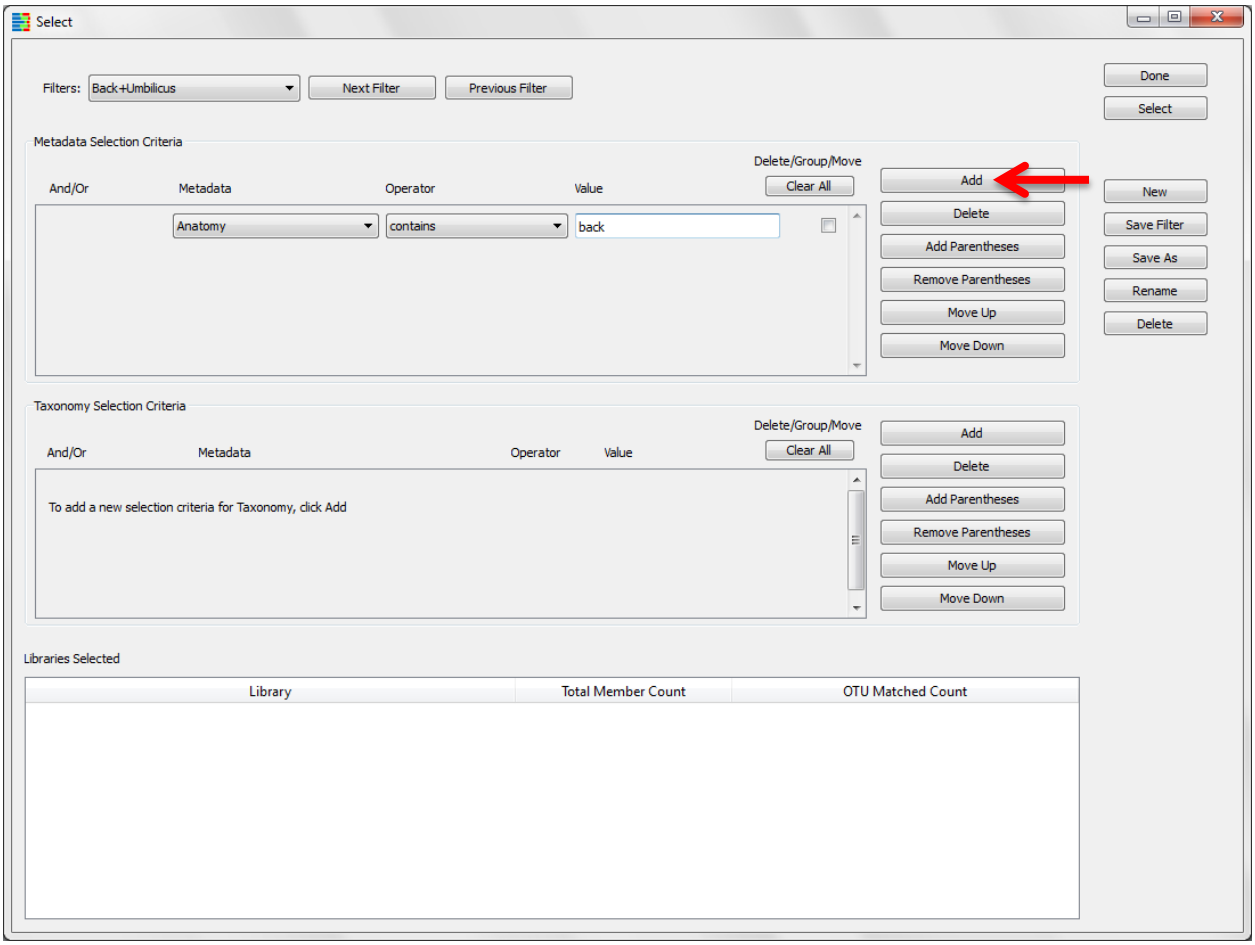
Use the first pull-down menu to select “Anatomy” (**Metadata** to filter by)
Use the second pull-down menu to select “contains” (filter **Operator**)
Enter “back” into **Value**

© 2012-2013 Charles E. Robertson and Incubix, Inc.



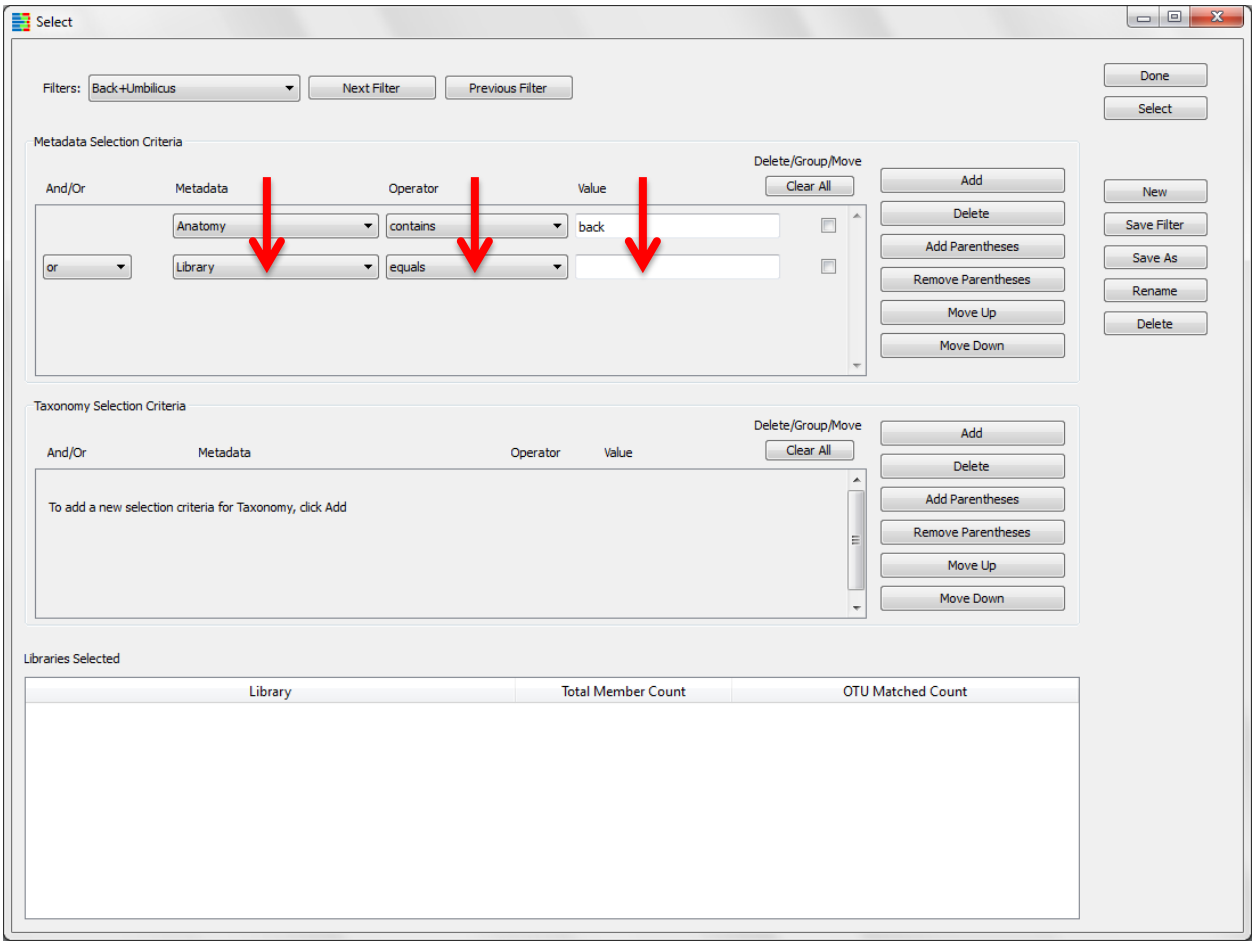
Click **Add** in the **Metadata Criteria** pane

© 2012-2013 Charles E. Robertson and Incubix, Inc.



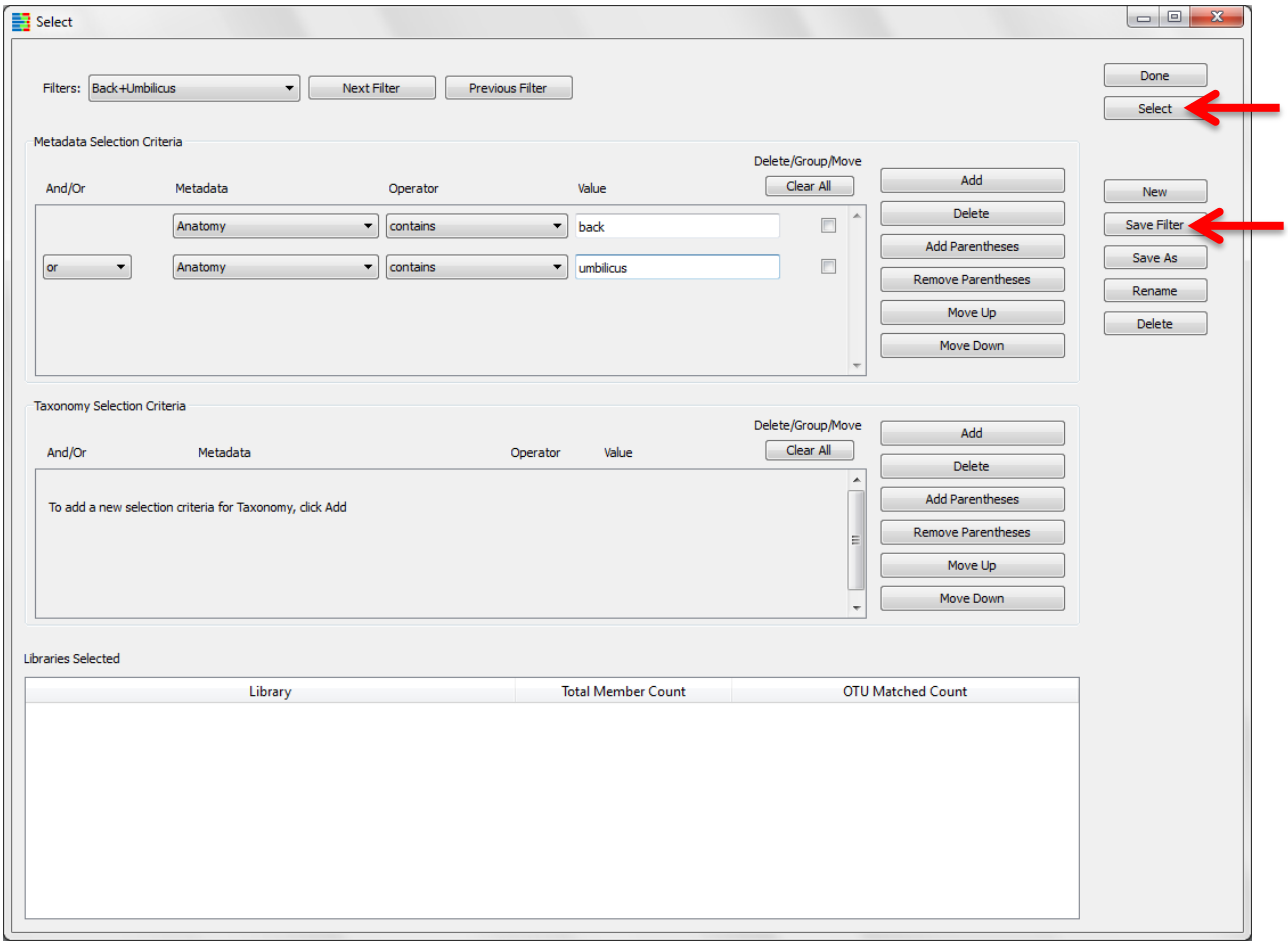
Use the first pull-down menu to select “Or”
Use the second pull-down menu to select “Anatomy” (**Metadata** to filter by)
Use the third pull-down menu to select “contains” (filter **Operator**)
Enter “umbilicus” into **Value**

© 2012-2013 Charles E. Robertson and Incubix, Inc.



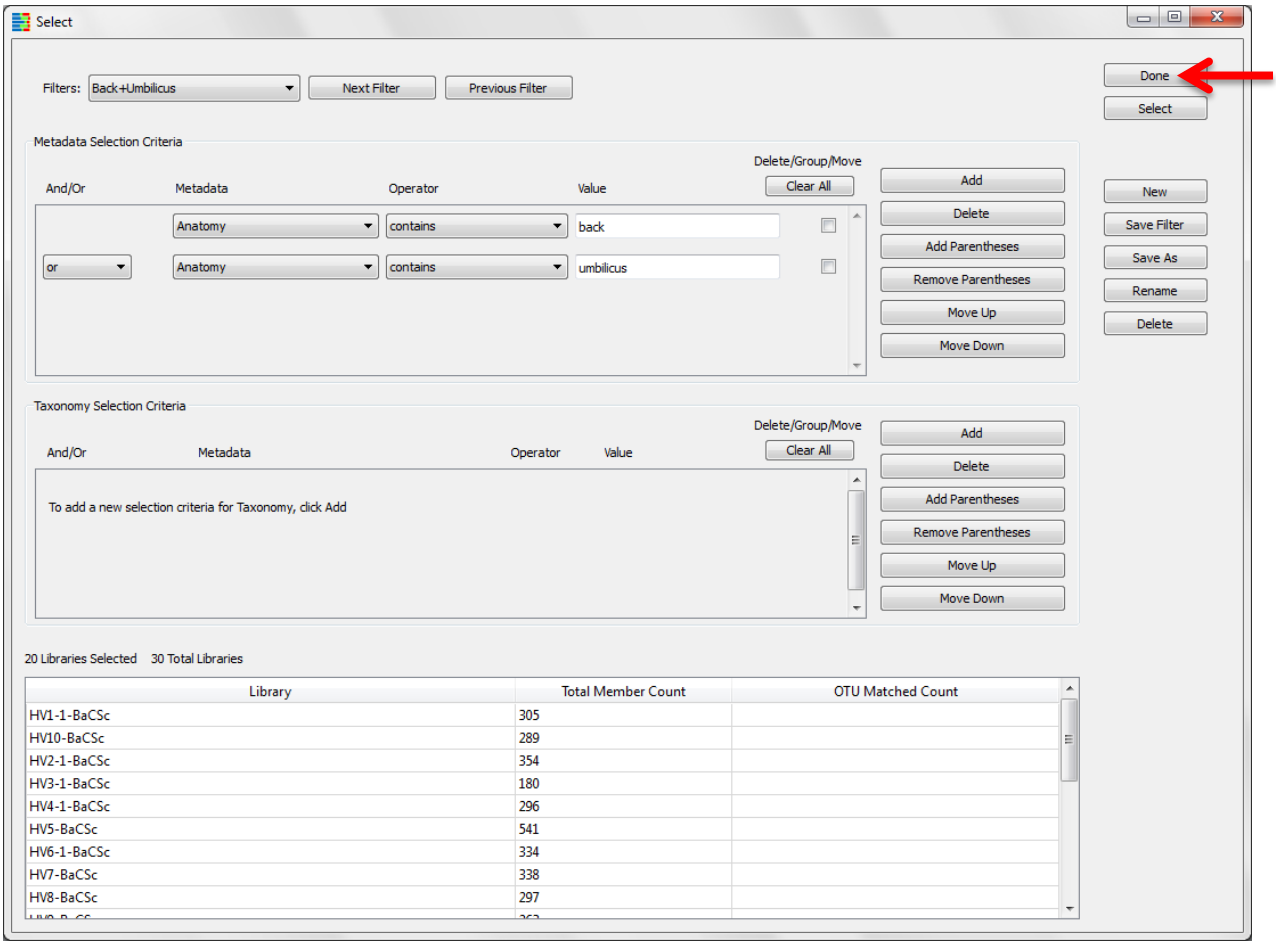
To apply filter, click **Select** in upper right corner of window
Click **Save Filter** on far right side of window to keep the filter

© 2012-2013 Charles E. Robertson and Incubix, Inc.



Click **Done** in upper right corner of window

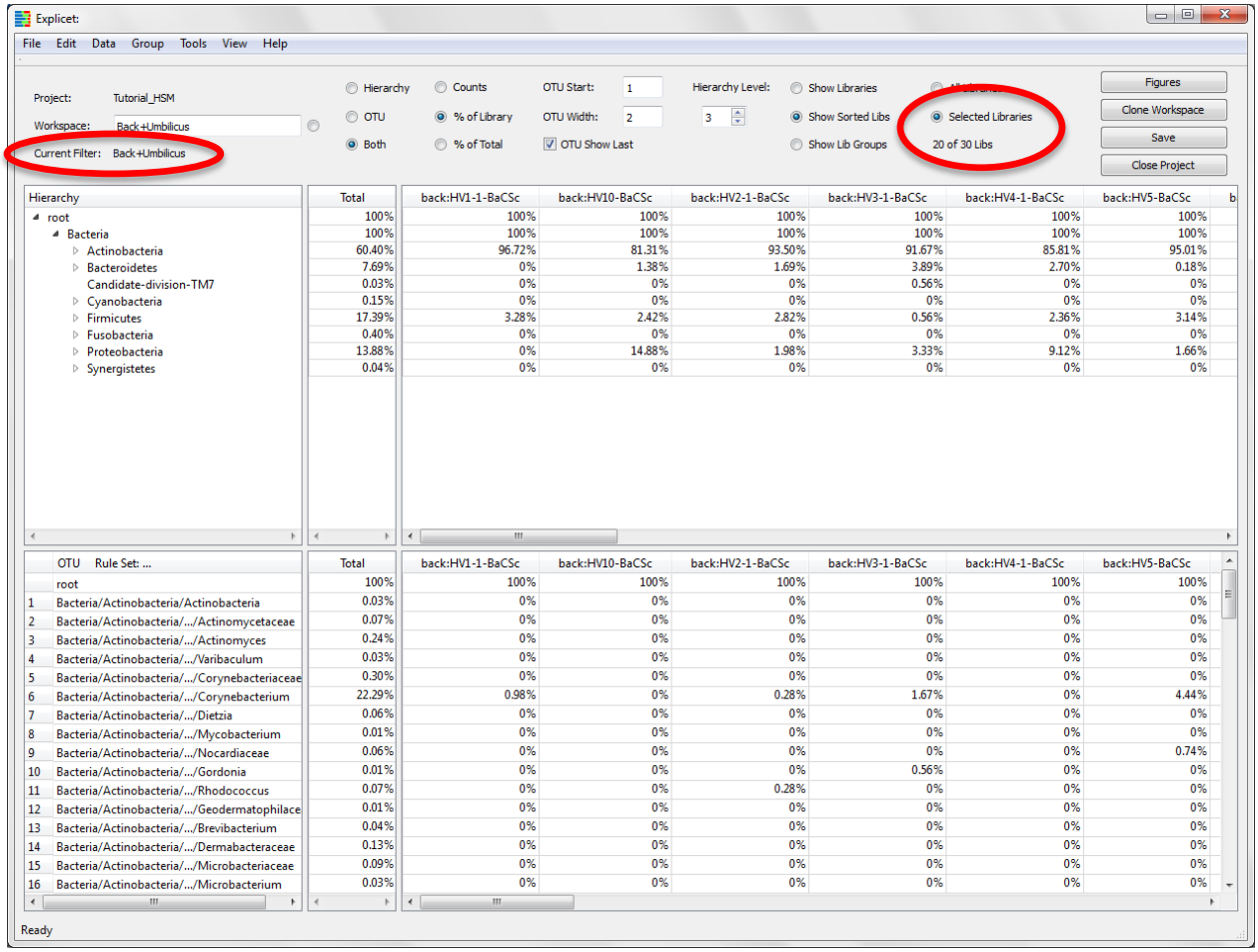
© 2012-2013 Charles E. Robertson and Incubix, Inc.



Pop-up window will disappear

On the current workspace window, **Selected Libraries** is now selected, and the name of the **Current Filter** is displayed in the upper left corner of the window. The workspace window now only displays libraries from the 20 back and umbilicus samples.

© 2012-2013 Charles E. Robertson and Incubix, Inc.



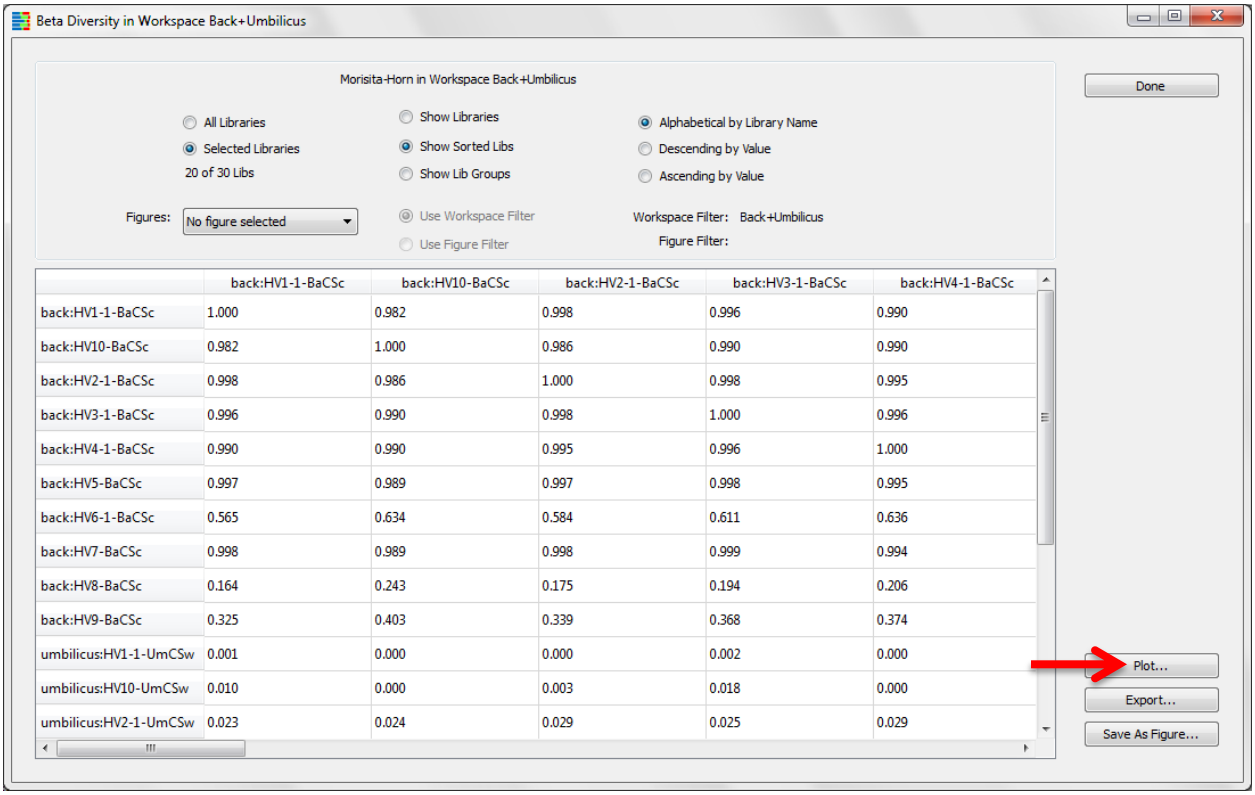
© 2012-2013 Charles E. Robertson and Incubix, Inc.

X. Beta Diversity (Morisita-Horn)

By viewing our libraries in a Morisita-Horn heatmap, we can estimate the similarity of the microbial communities present in the samples at these two anatomical positions. Morisita-Horn is an often used metric that can give insight into how similar or how different sets of samples are from each other by looking at the patterns of all of the different OTUs at the same time.

A. Create a Morisita-Horn Heatmap

Tools → Analyze → Beta Diversity → Morisita-Horn
A new window will appear with a table of the sequence variant counts
Click **Plot**

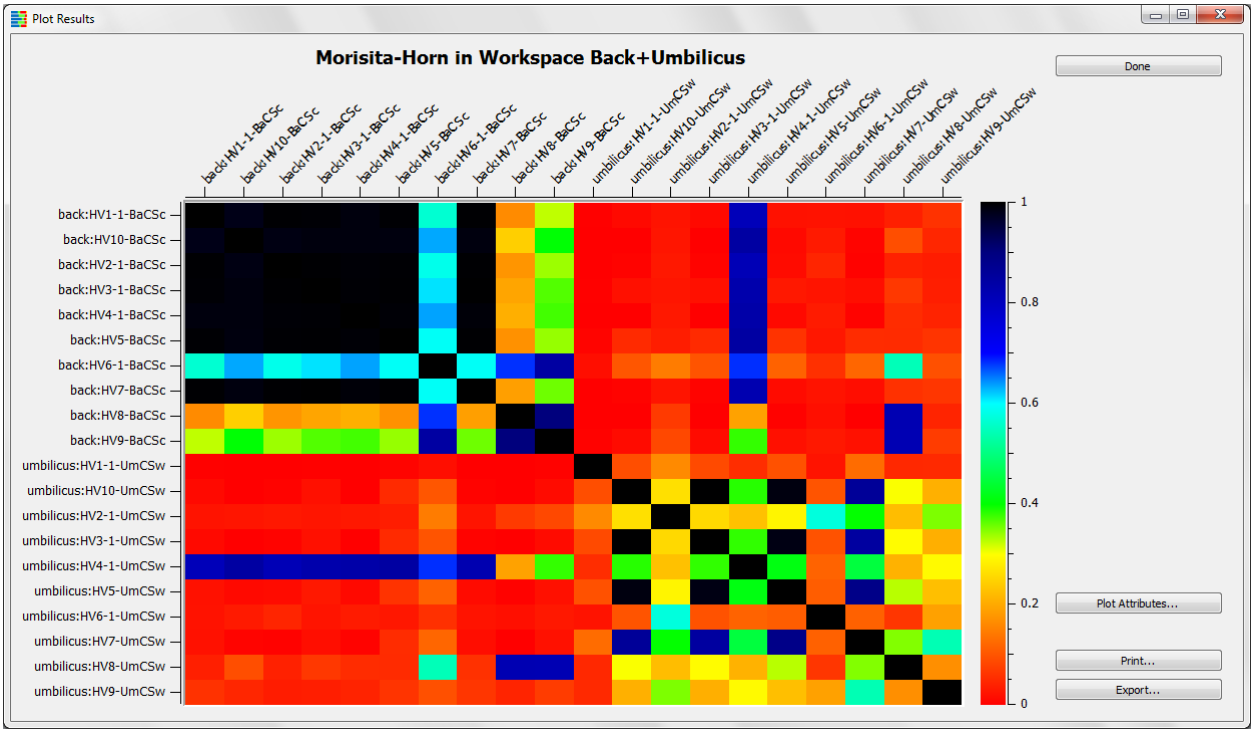


	back:HV1-1-BaCSc	back:HV10-BaCSc	back:HV2-1-BaCSc	back:HV3-1-BaCSc	back:HV4-1-BaCSc
back:HV1-1-BaCSc	1.000	0.982	0.998	0.996	0.990
back:HV10-BaCSc	0.982	1.000	0.986	0.990	0.990
back:HV2-1-BaCSc	0.998	0.986	1.000	0.998	0.995
back:HV3-1-BaCSc	0.996	0.990	0.998	1.000	0.996
back:HV4-1-BaCSc	0.990	0.990	0.995	0.996	1.000
back:HV5-BaCSc	0.997	0.989	0.997	0.998	0.995
back:HV6-1-BaCSc	0.565	0.634	0.584	0.611	0.636
back:HV7-BaCSc	0.998	0.989	0.998	0.999	0.994
back:HV8-BaCSc	0.164	0.243	0.175	0.194	0.206
back:HV9-BaCSc	0.325	0.403	0.339	0.368	0.374
umbilicus:HV1-1-UmCSw	0.001	0.000	0.000	0.002	0.000
umbilicus:HV10-UmCSw	0.010	0.000	0.003	0.018	0.000
umbilicus:HV2-1-UmCSw	0.023	0.024	0.029	0.025	0.029

A new window will appear containing the heatmap of Morisita-Horn sequence variant counts

Note: In our workspace, we have **Selected Libraries** selected, so the heatmap will only display results from our libraries of interest (only those libraries sampled from the back or umbilicus).

© 2012-2013 Charles E. Robertson and Incubix, Inc.



Anatomical positions with Morisita-Horn values near 1 (implying the samples' constituent taxonomy patterns are very *similar*) appear black. Anatomical positions with Morisita-Horn values near 0 (implying the samples' constituent taxonomy patterns are very *different*) appear red. Based on this data, the back is more similar across subjects than the umbilicus. Plot attributes allow control of plot characteristics and color usage as described earlier.

You may choose to save the Morisita-Horn heatmap as a figure. To do so, continue as shown earlier in the stacked bar chart example; close the graphics window, and select **Save As Figure** in the **OTU Heatmap** window.

© 2012-2013 Charles E. Robertson and Incubix, Inc.

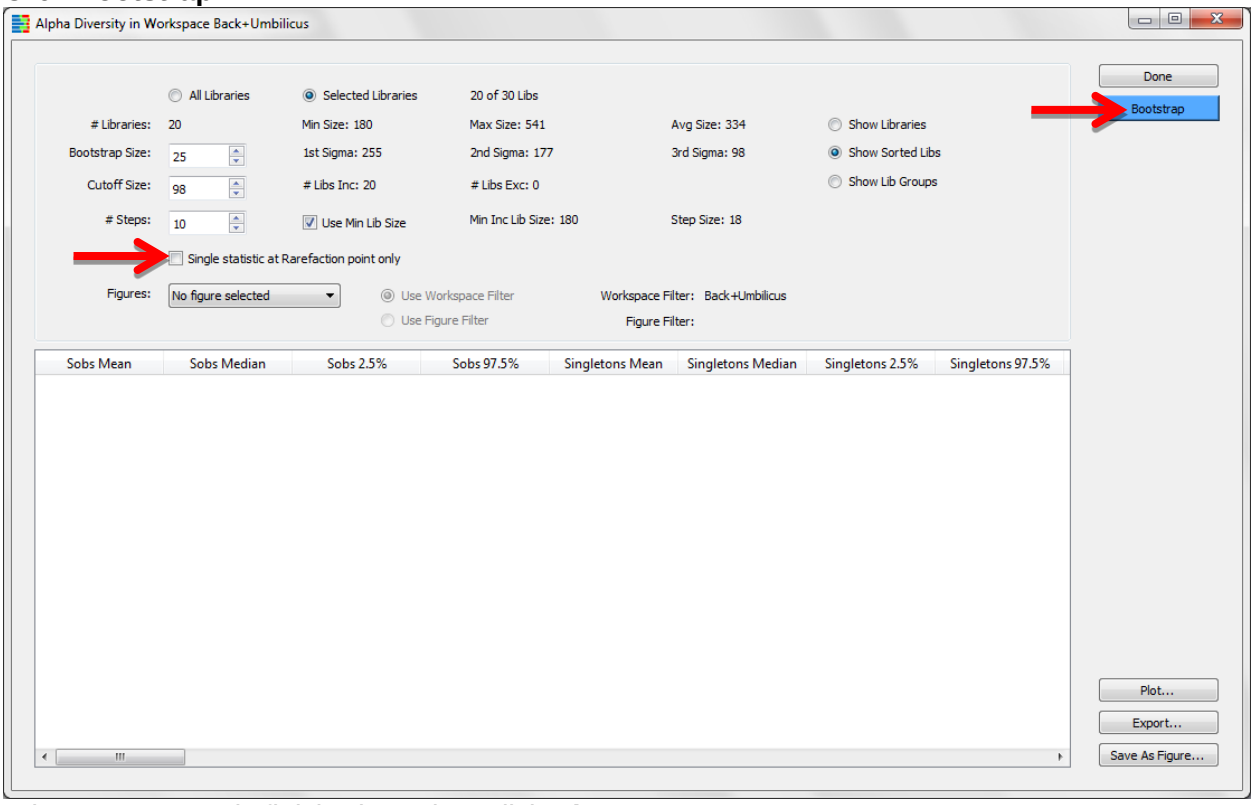
XI. Alpha Diversity

The alpha diversity statistics computed by Explicet are generally shown in one of two ways: either as a single value calculated at the size of the smallest library (known as the rarefaction point) or as multiple values plotted as collector's curves for each library. Collector's curves are the classic way to evaluate the impact of increasing sample size (i.e., more sequencing) on the information content of the dataset. All collector's curves in Explicet are computed with rarefaction, meaning all libraries are resampled to allow fair comparison between libraries of greatly different size. The higher the resolution of the calculations (large number of bootstrap iterations, large number of steps), the slower the computations will proceed. It is recommended that users start with the defaults and then increase as needed to get the curves to smooth out. Very large bootstrap iterations and a large number of steps may result in a run of multiple days… So, start small and work up.

The alpha diversity metrics are often quick, reliable ways to determine if samples in a dataset are sequenced adequately. Since we have a workspace set up to run mini-experiments on a subset of our data, we should make sure that the data is representative. We need to make sure that enough sequences were generated from the back and umbilicus samples to be considered representative of the anatomical position for a subject. We can test this by running an alpha diversity test called Good's Coverage.

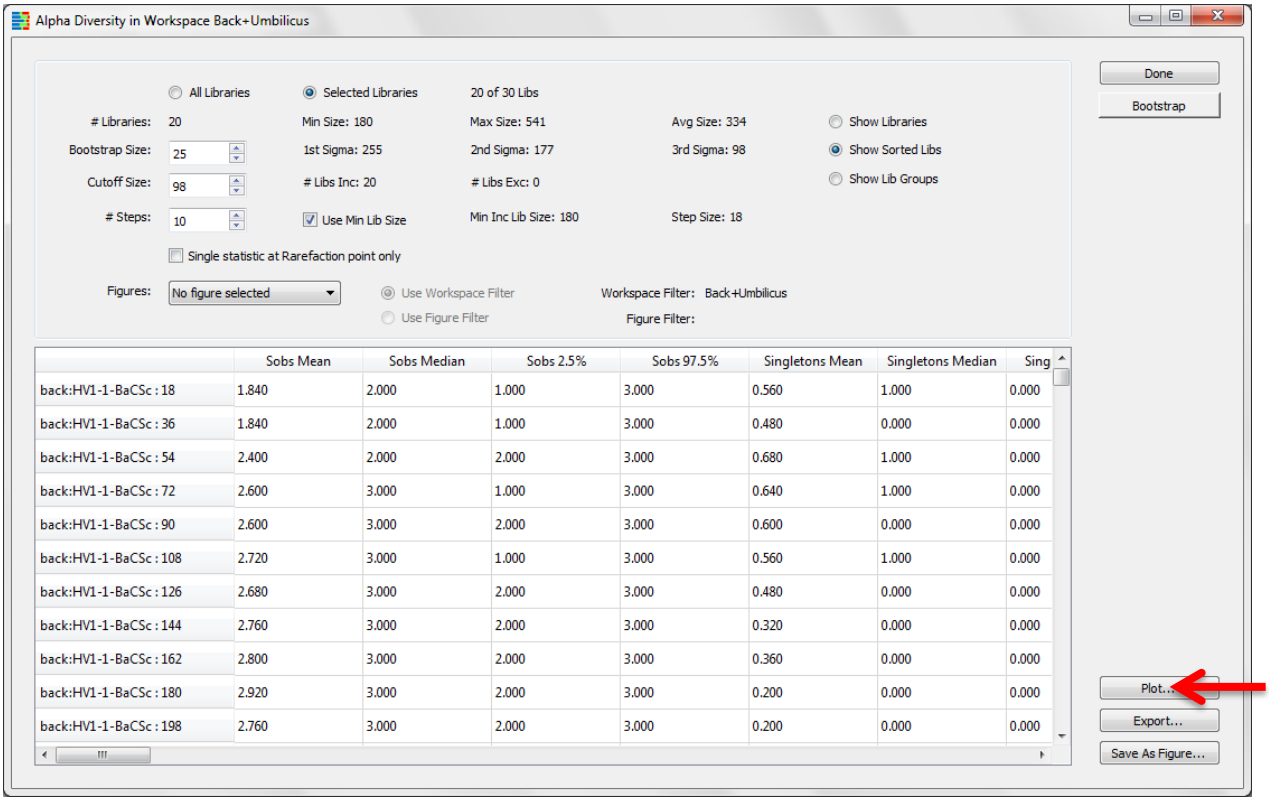
A. Run a Good's Coverage Test

Tools → Analyze → Alpha Diversity
New pop-up window appears
To create curves, deselect **Single statistic at Rarefaction point only**
Click **Bootstrap**



When **Bootstrap** is finished running, click **Plot**

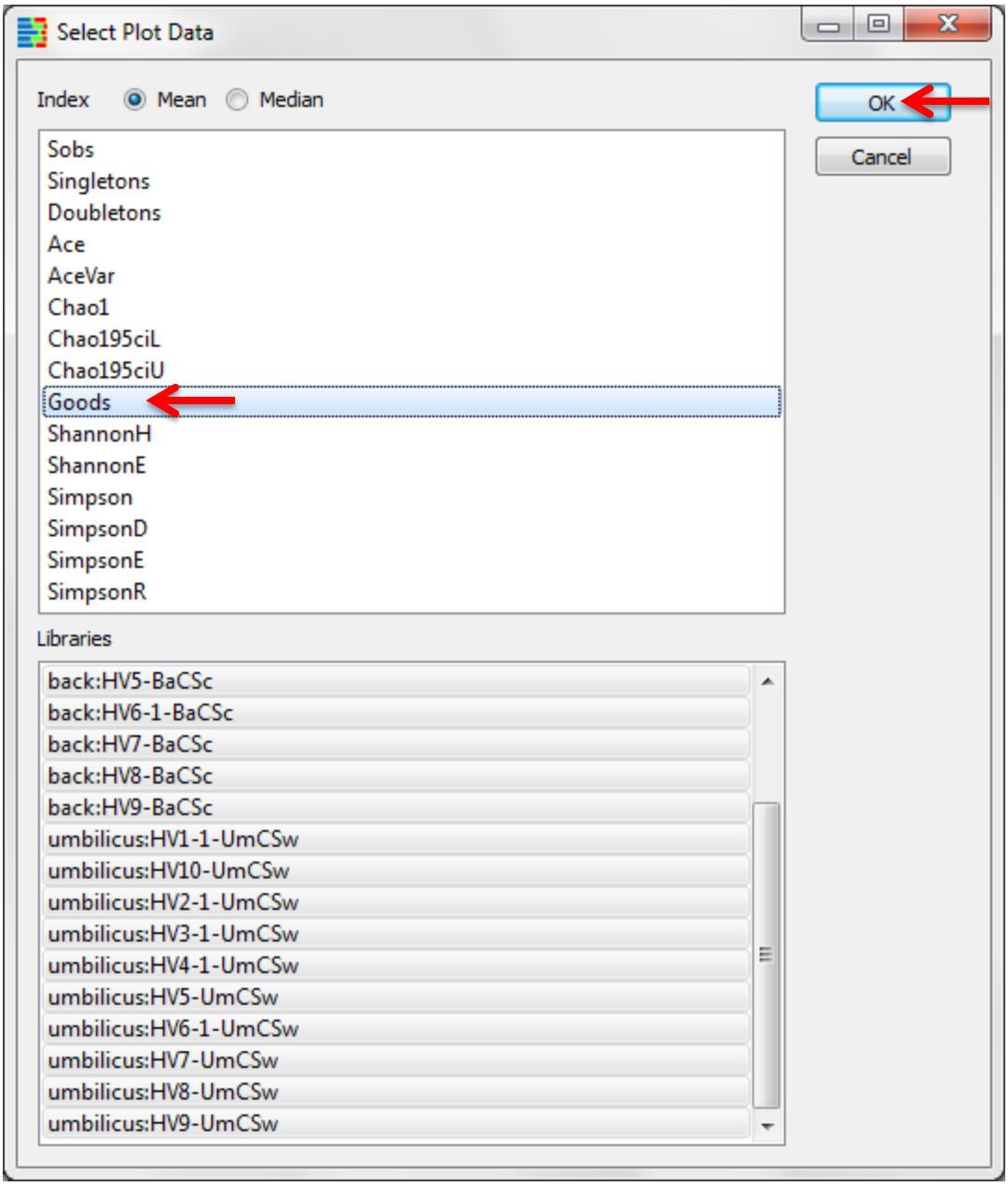
© 2012-2013 Charles E. Robertson and Incubix, Inc.



	Sobs Mean	Sobs Median	Sobs 2.5%	Sobs 97.5%	Singletons Mean	Singletons Median	Sing
back:HV1-1-BaCSc : 18	1.840	2.000	1.000	3.000	0.560	1.000	0.000
back:HV1-1-BaCSc : 36	1.840	2.000	1.000	3.000	0.480	0.000	0.000
back:HV1-1-BaCSc : 54	2.400	2.000	2.000	3.000	0.680	1.000	0.000
back:HV1-1-BaCSc : 72	2.600	3.000	1.000	3.000	0.640	1.000	0.000
back:HV1-1-BaCSc : 90	2.600	3.000	2.000	3.000	0.600	0.000	0.000
back:HV1-1-BaCSc : 108	2.720	3.000	1.000	3.000	0.560	1.000	0.000
back:HV1-1-BaCSc : 126	2.680	3.000	2.000	3.000	0.480	0.000	0.000
back:HV1-1-BaCSc : 144	2.760	3.000	2.000	3.000	0.320	0.000	0.000
back:HV1-1-BaCSc : 162	2.800	3.000	2.000	3.000	0.360	0.000	0.000
back:HV1-1-BaCSc : 180	2.920	3.000	2.000	3.000	0.200	0.000	0.000
back:HV1-1-BaCSc : 198	2.760	3.000	2.000	3.000	0.200	0.000	0.000

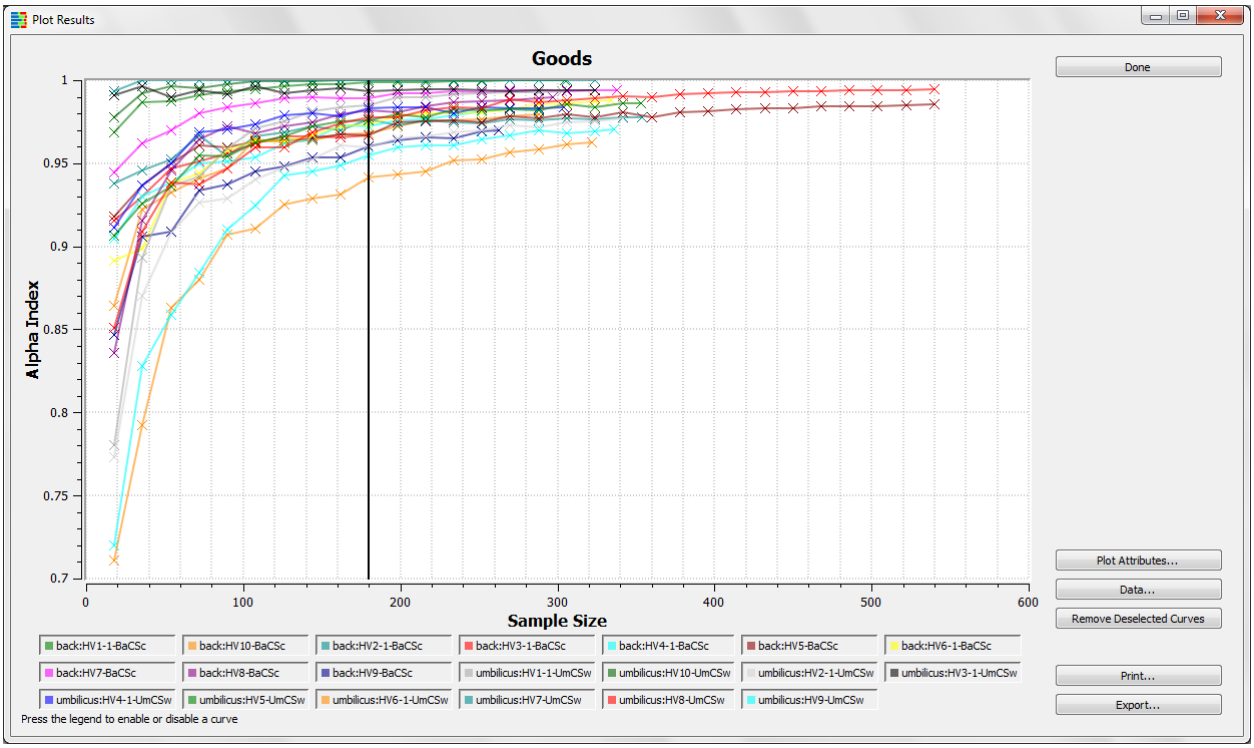
A new pop-up window appears which lists the various alpha diversity tests
Select **Goods**
Click **OK**

© 2012-2013 Charles E. Robertson and Incubix, Inc.



A new pop-up window appears showing the Good's Coverage plot

© 2012-2013 Charles E. Robertson and Incubix, Inc.



Since the curves on the plot generally reach asymptotes, we conclude that both sites were sampled reasonably well to be considered representative of the anatomical positions.

You may choose to save your Good's Coverage plot as a figure. To do so, continue as shown earlier in the stacked bar chart example; close the graphics window, and select **Save As Figure** in the **Alpha Diversity** window.

© 2012-2013 Charles E. Robertson and Incubix, Inc.

XII. Two-Part Test

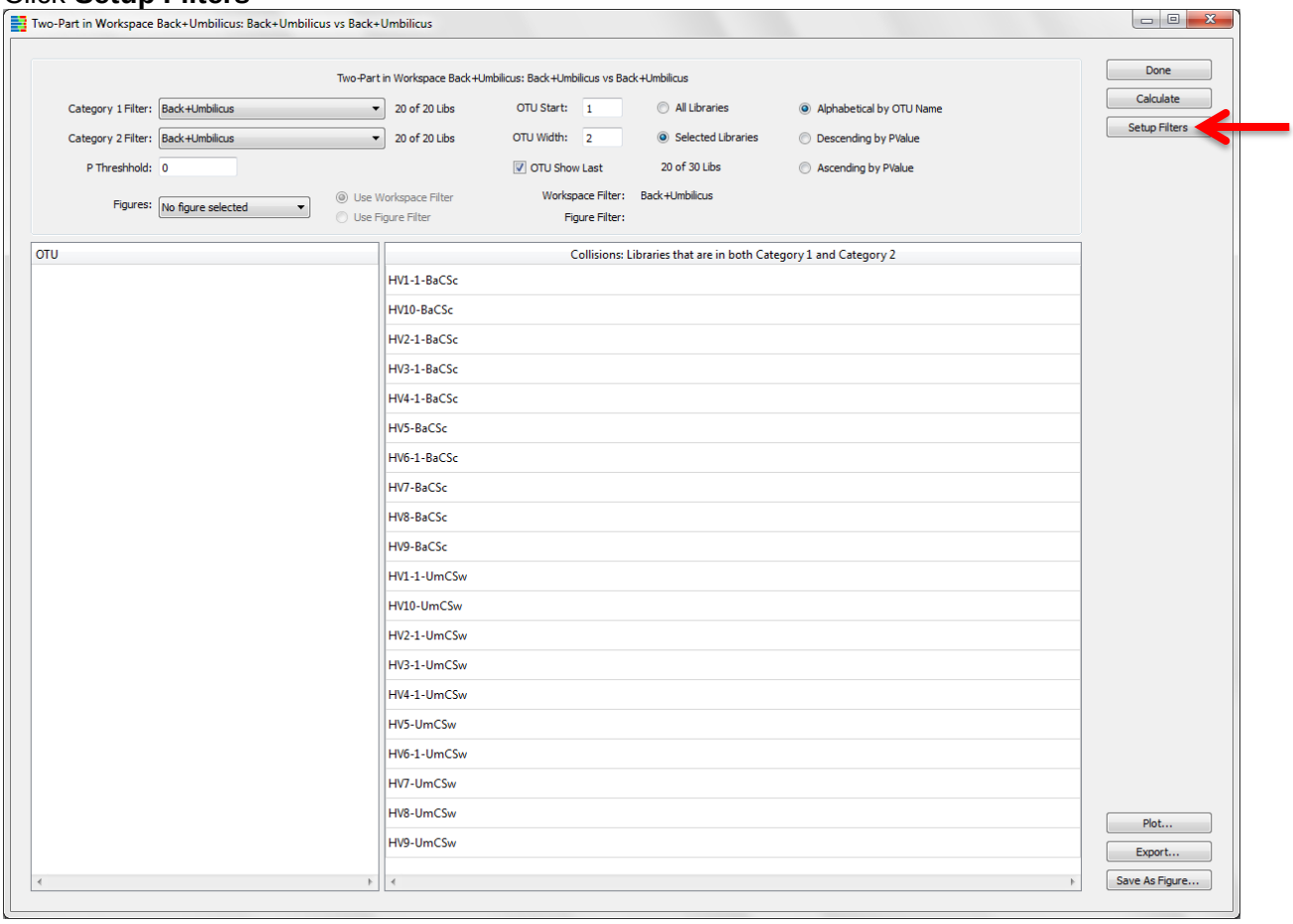
Now that we know our data are representative, we will continue with another statistical test. A Two-Part statistical test can identify taxa that differ between two groups. We will use the Two-Part test to compare sequence counts between the back and umbilicus. The Two-Part Test is a combined statistic that examines both the proportion of the samples that contain a given OTU and the median relative abundance of the OTU across two categories. Because microbiome data often are non-normally distributed, parametric tests such as the familiar t-test may not be appropriate. Consequently, we use a non-parametric Wilcoxon test to examine percent abundance data. For more information on the Two-Part Test, please see:
Wagner BD, Robertson CE, Harris JK (2011) Application of Two-Part Statistics for Comparison of Sequence Variant Counts. *PLoS ONE* 6(5): e20296.

A. Run a Two-Part Test

Tools → Analyze → Two-Part
A new pop-up window appears

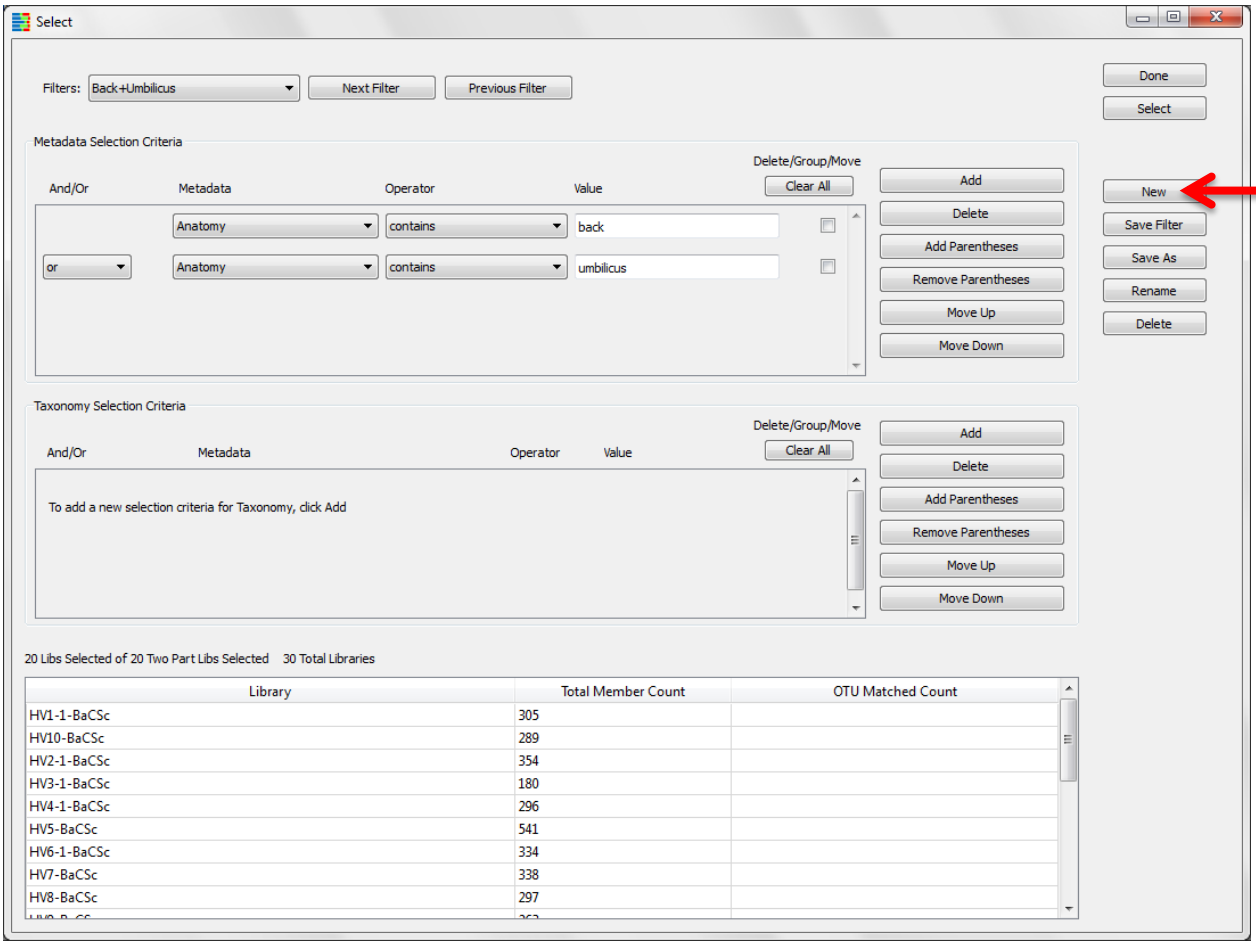
In order to compare the back data against the umbilicus data, we need to set up individual filters for each anatomical position. To do so, we will proceed as discussed earlier in "To create a filter…".

Click **Setup Filters**



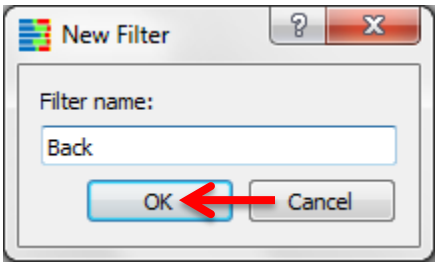
New pop-up window appears for creation of filters
Click **New** on far right side of window

© 2012-2013 Charles E. Robertson and Incubix, Inc.



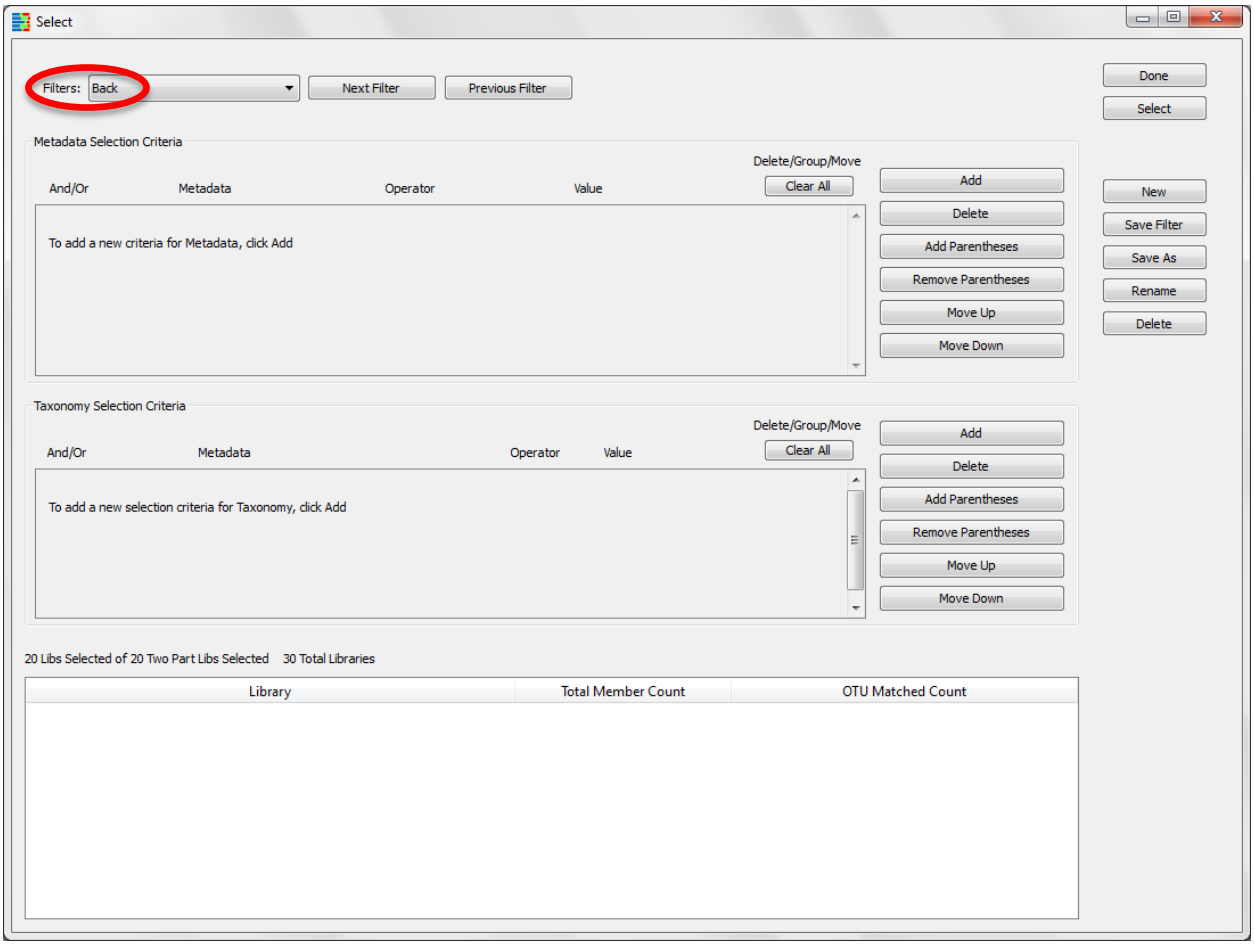
Enter desired filter name in the pop-up window

Click **OK**



The filter name will appear in upper left corner of window

© 2012-2013 Charles E. Robertson and Incubix, Inc.

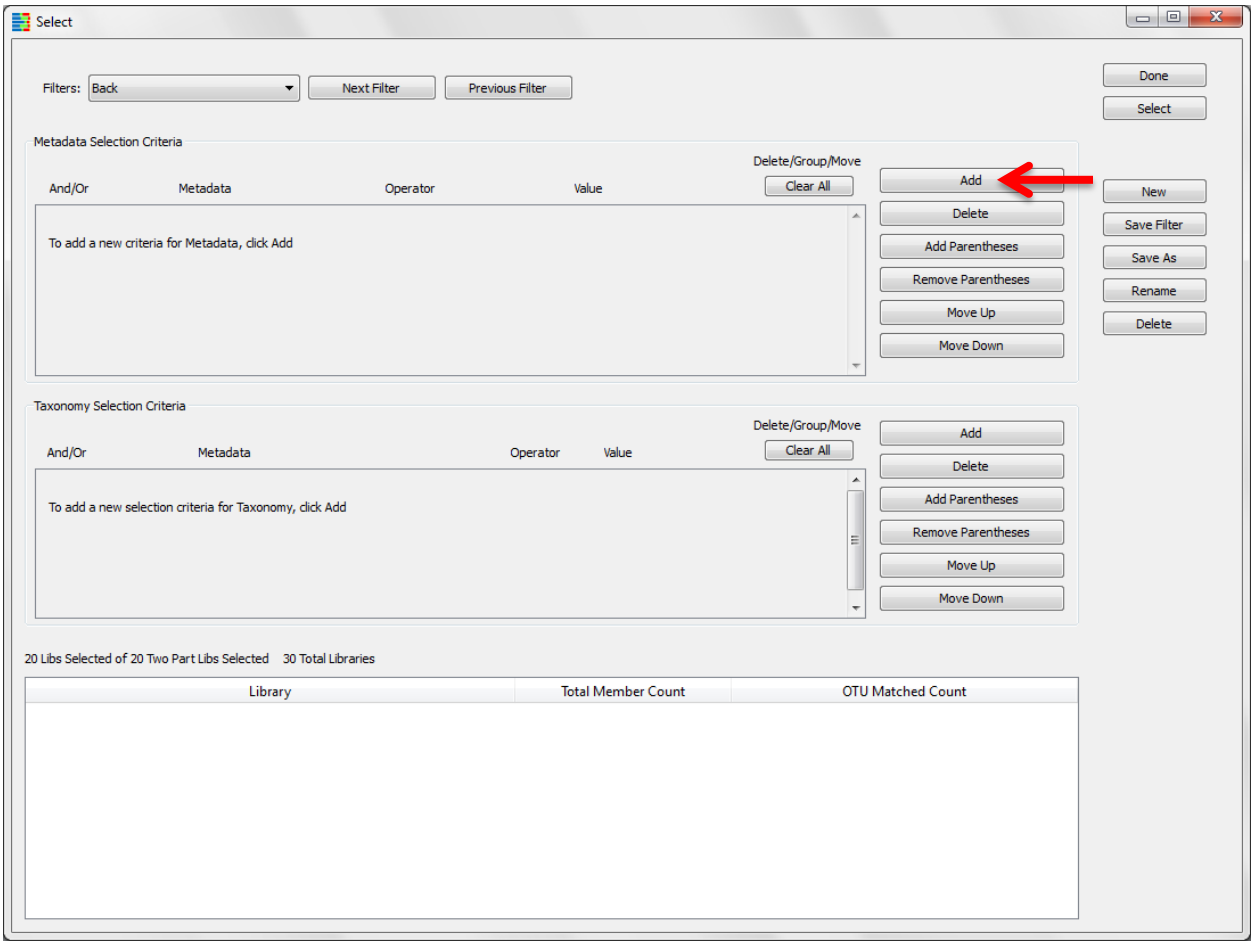


Now that we have created a new filter, we need to set up the parameters to filter by. We will select for all libraries which were sampled from the “back”.

B. Set Up Filter Parameters

Click **Add** in the **Metadata Criteria** pane

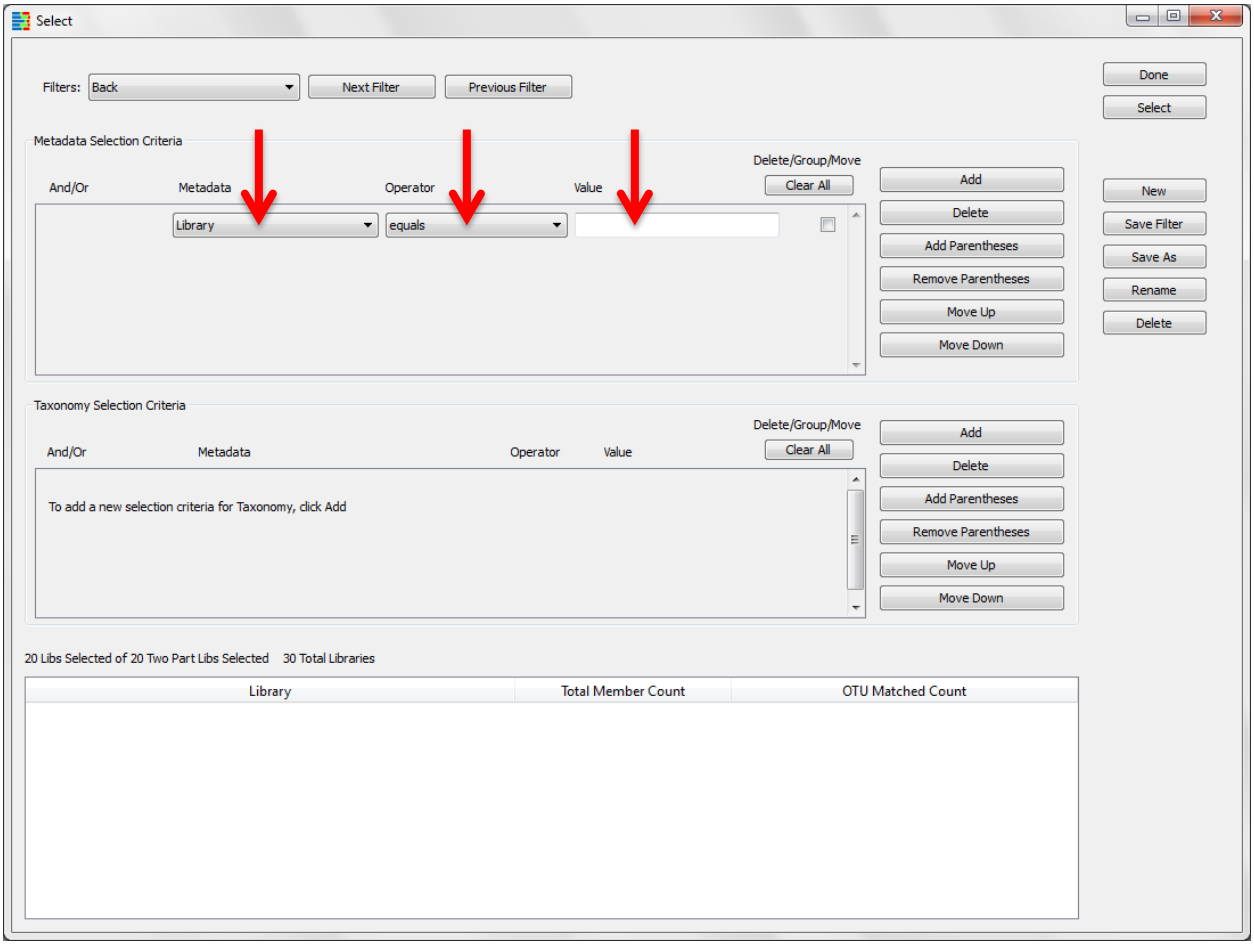
© 2012-2013 Charles E. Robertson and Incubix, Inc.



Use the first pull-down menu to select “Anatomy” (**Metadata** to filter by)
Use the second pull-down menu to select “contains” (filter **Operator**)

© 2012-2013 Charles E. Robertson and Incubix, Inc.

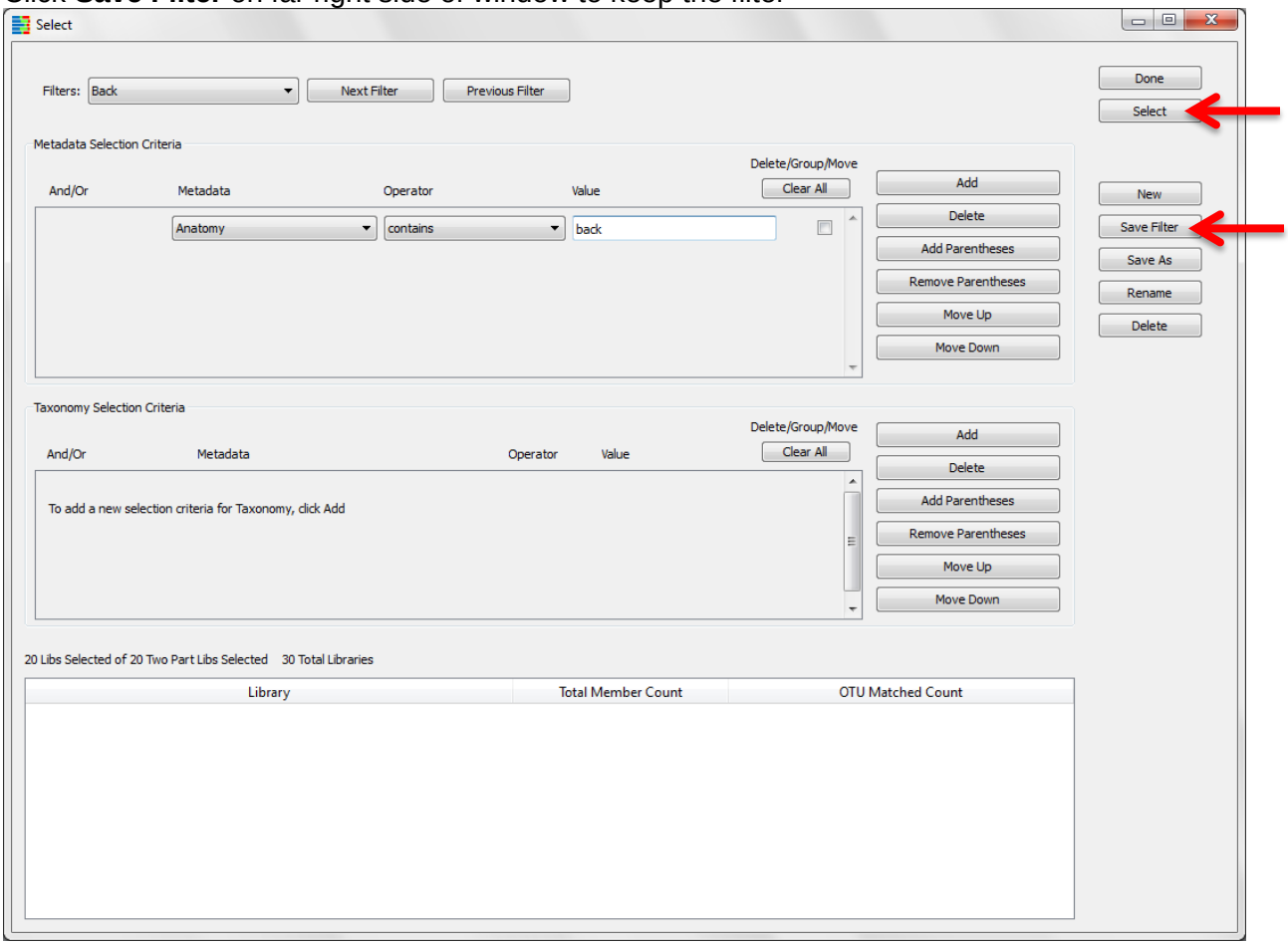
Enter “back” into **Value**



To apply filter, click **Select** in upper right corner of window

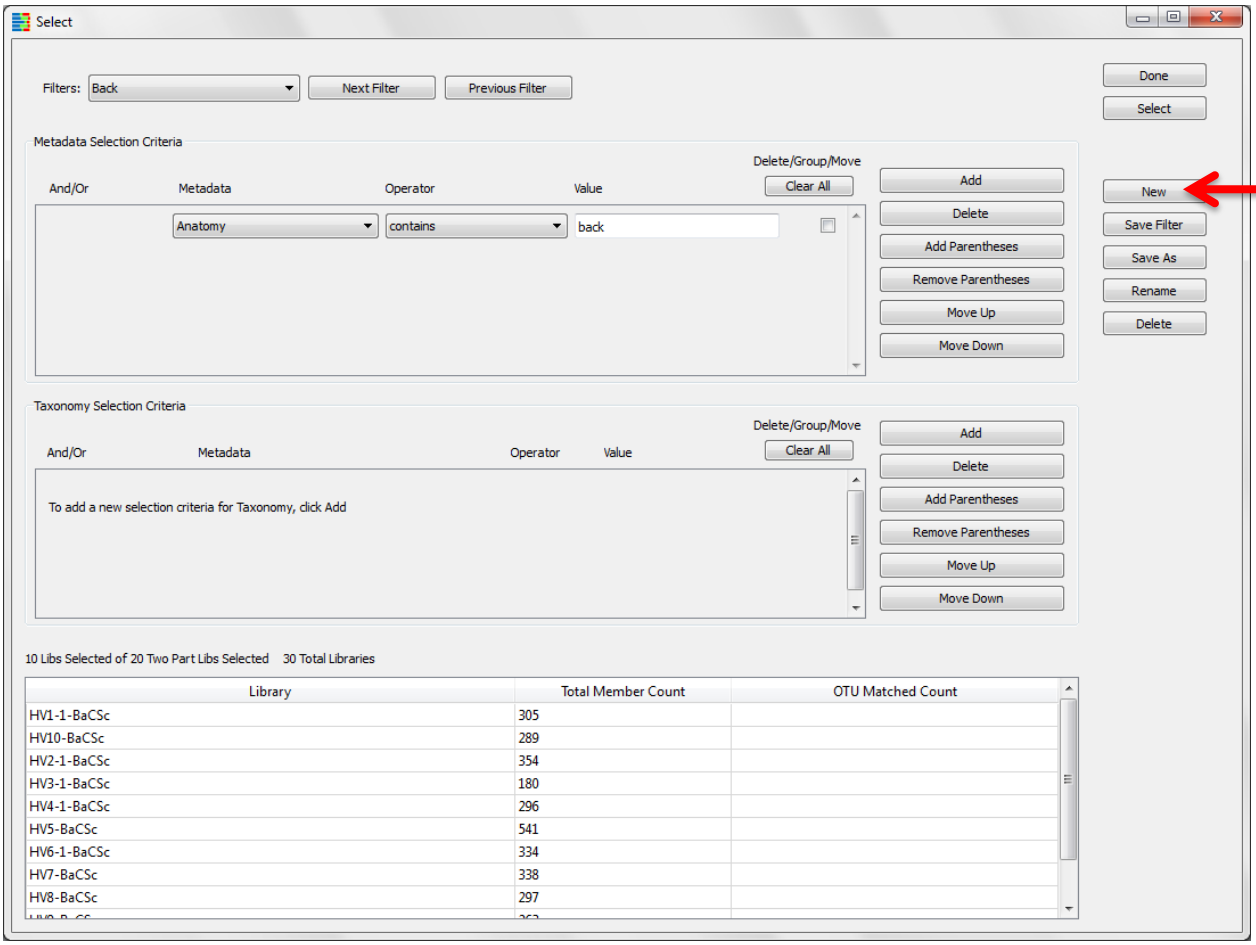
© 2012-2013 Charles E. Robertson and Incubix, Inc.

Click **Save Filter** on far right side of window to keep the filter



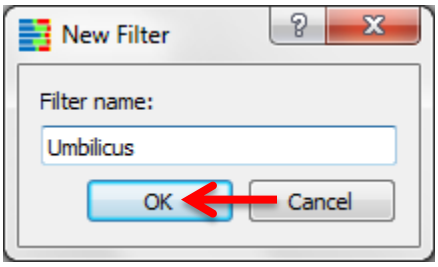
Now we will create a separate filter for the umbilicus
Click **New** on far right side of window

© 2012-2013 Charles E. Robertson and Incubix, Inc.



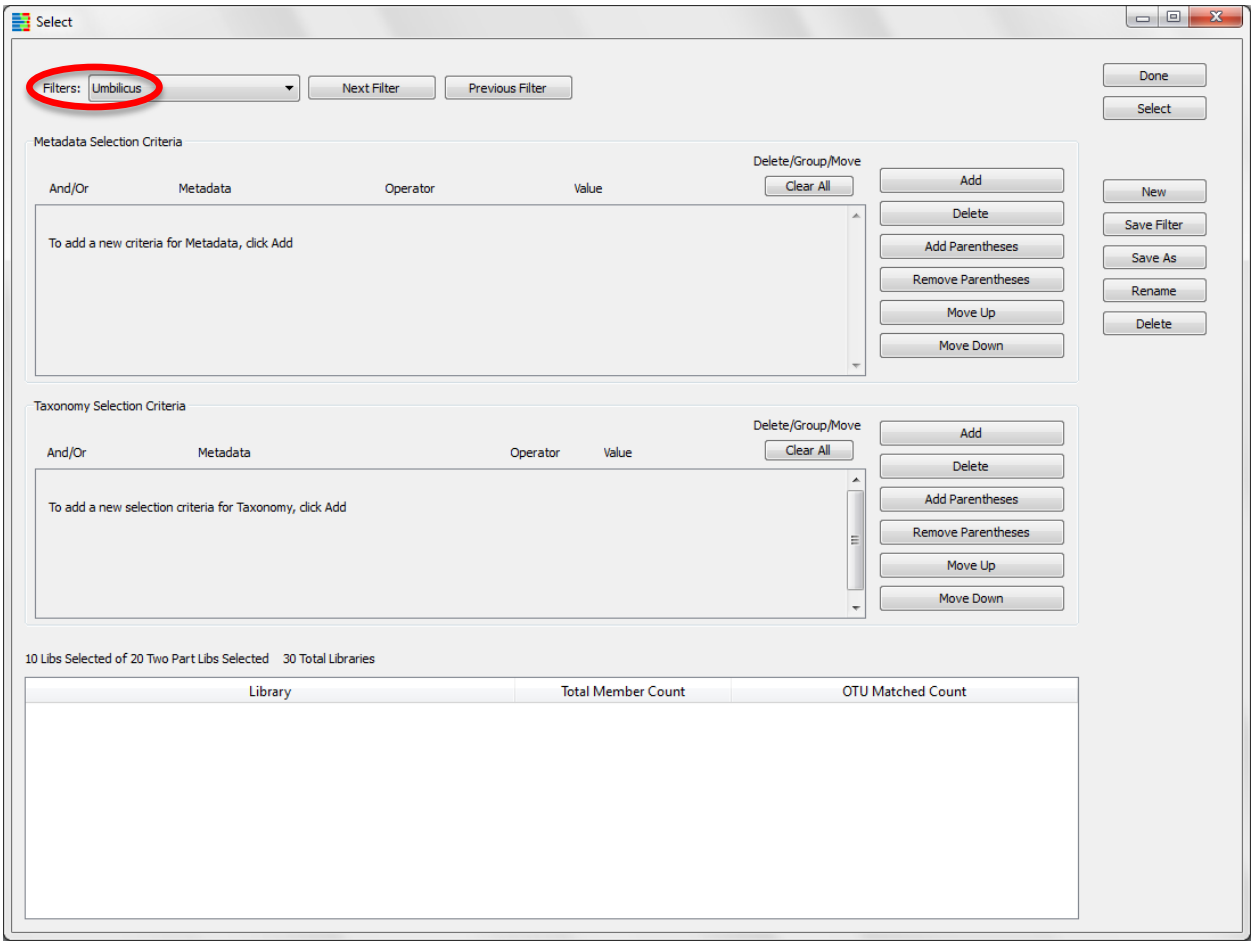
Enter desired filter name in the pop-up window

Click **OK**



The filter name will appear in upper left corner of window.

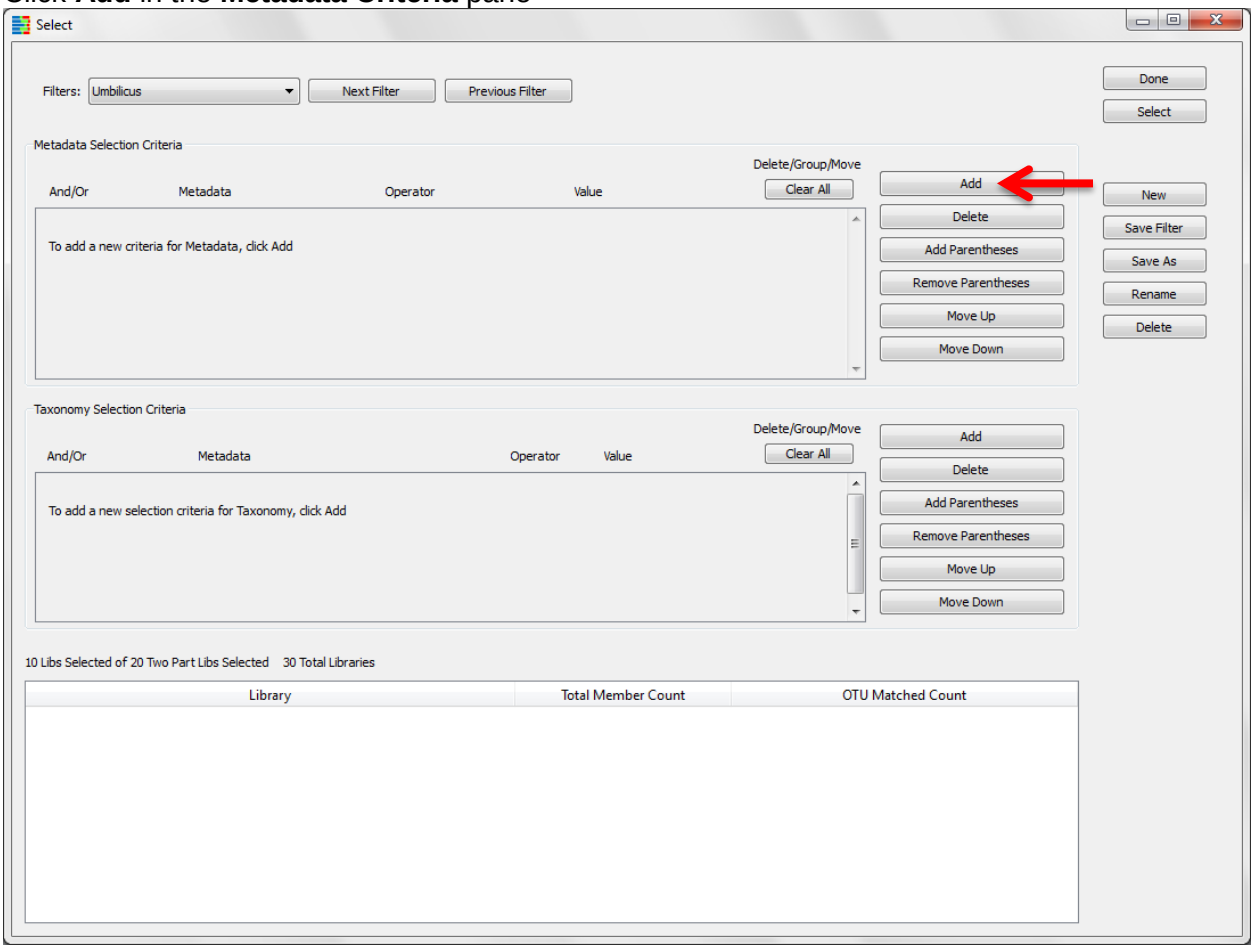
© 2012-2013 Charles E. Robertson and Incubix, Inc.



Now that we have created a new filter, we need to set up the parameters to filter by. We will select for all libraries which were sampled from the “umbilicus”.

© 2012-2013 Charles E. Robertson and Incubix, Inc.

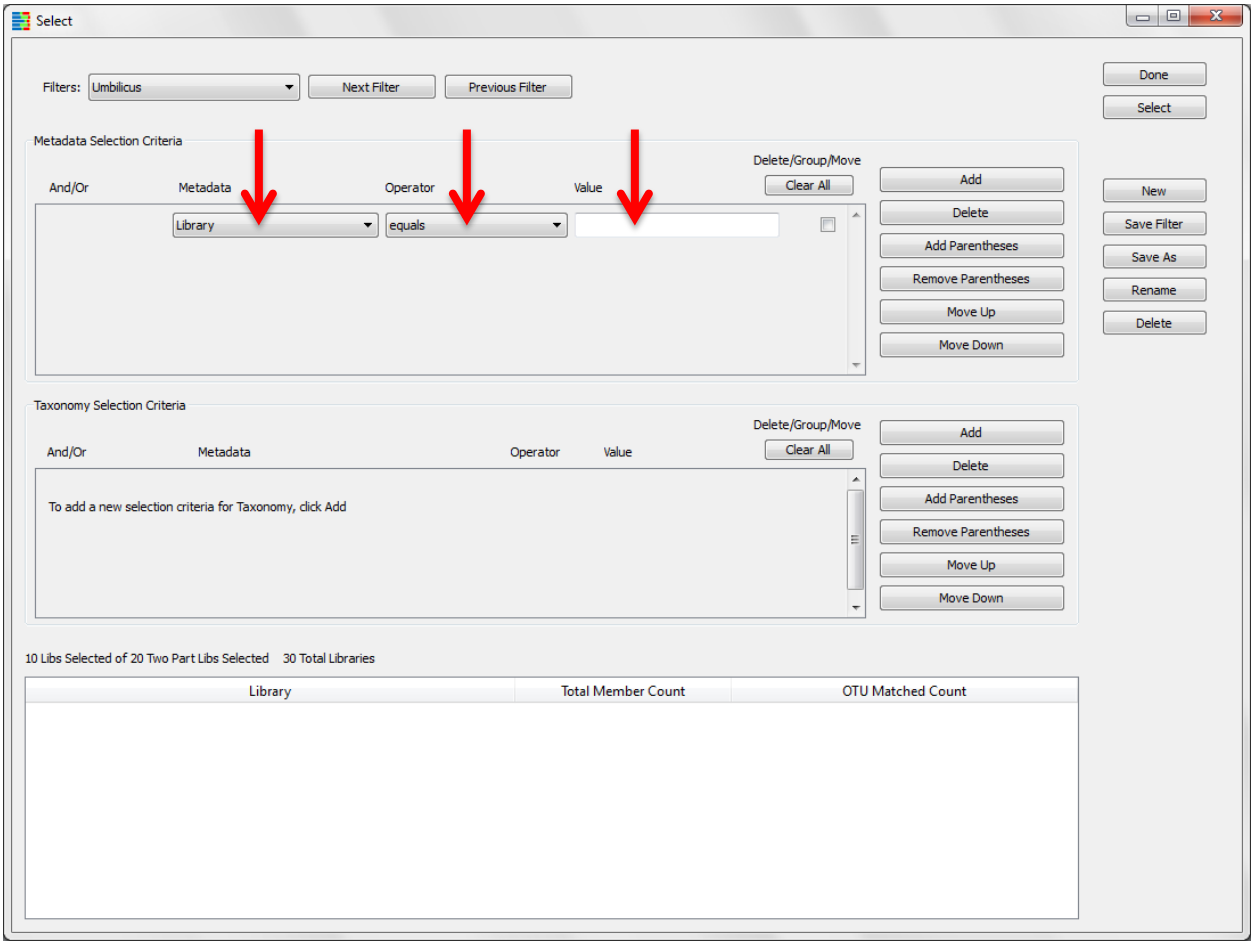
Click **Add** in the **Metadata Criteria** pane



Use the first pull-down menu to select “Anatomy” (**Metadata** to filter by)
Use the second pull-down menu to select “contains” (filter **Operator**)

© 2012-2013 Charles E. Robertson and Incubix, Inc.

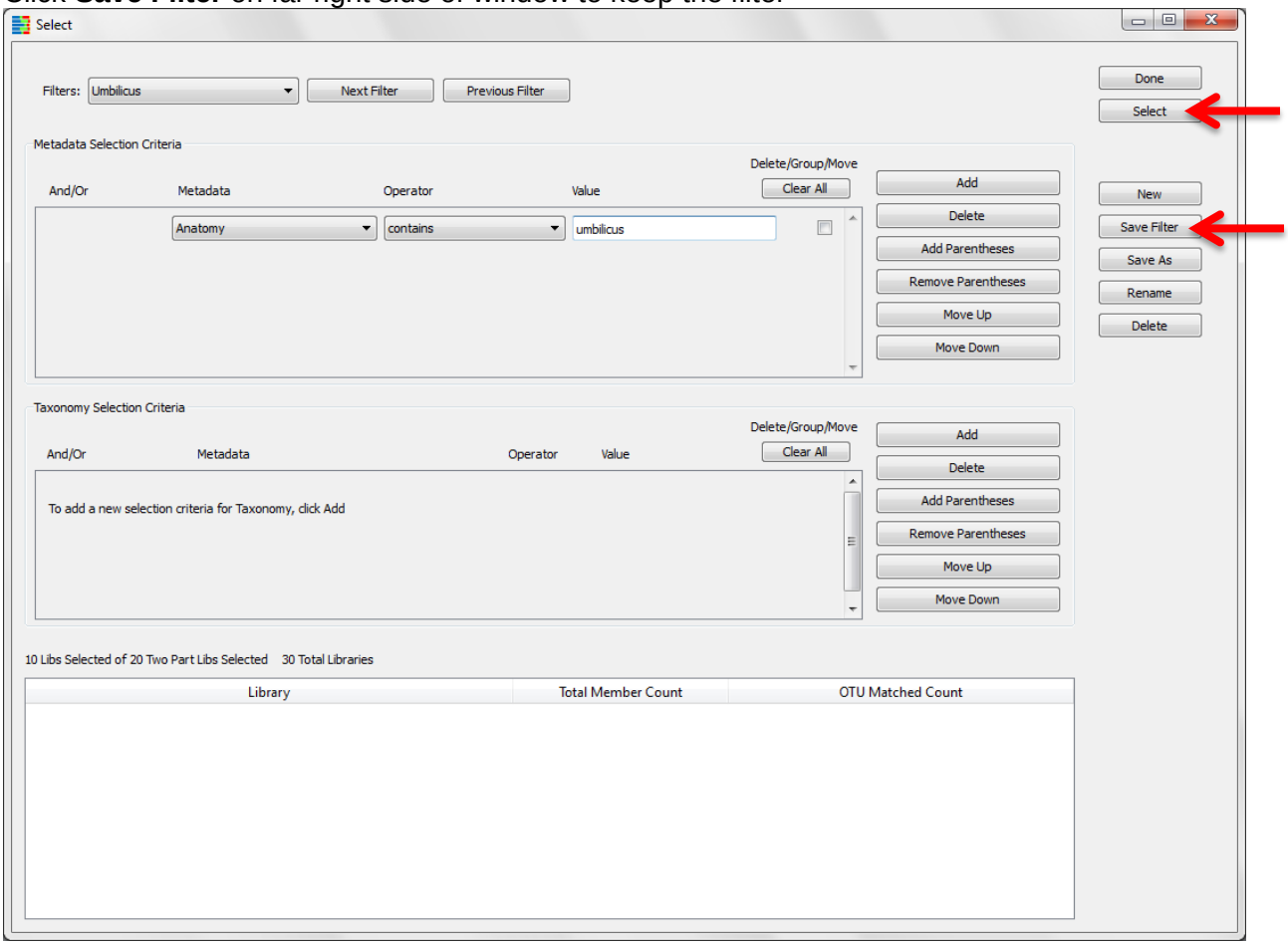
Enter “umbilicus” into **Value**



To apply filter, click **Select** in upper right corner of window

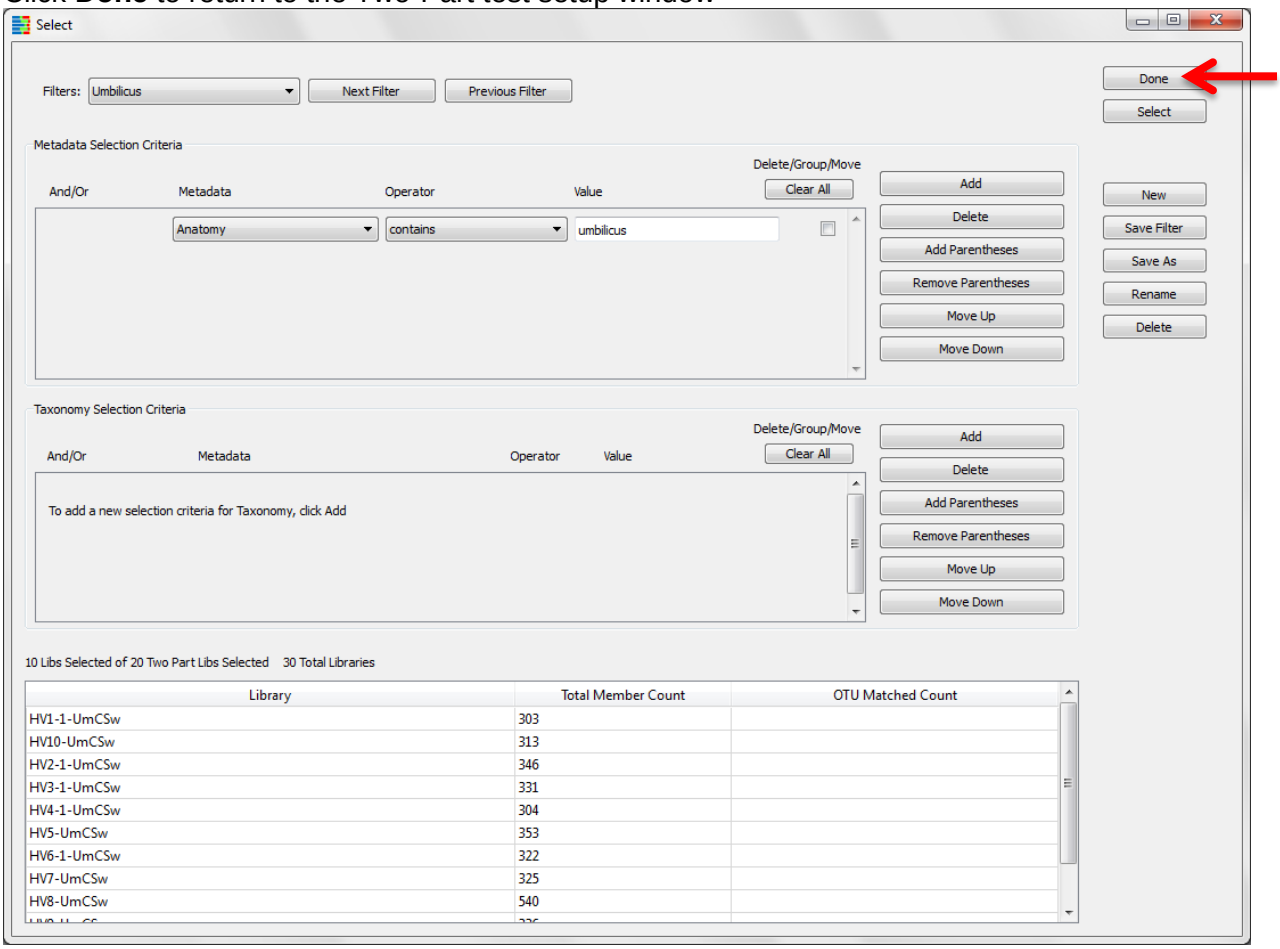
© 2012-2013 Charles E. Robertson and Incubix, Inc.

Click **Save Filter** on far right side of window to keep the filter



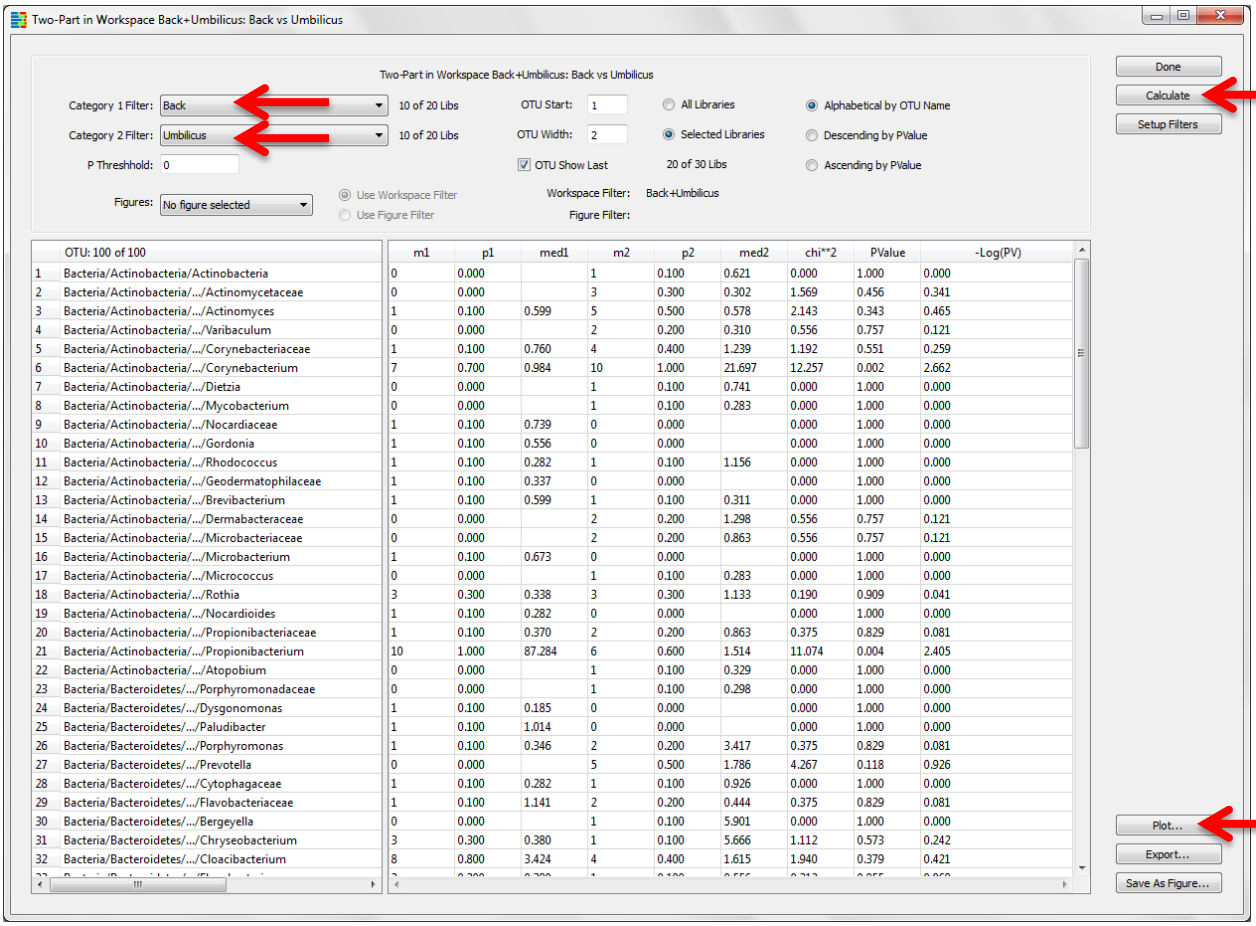
© 2012-2013 Charles E. Robertson and Incubix, Inc.

Click **Done** to return to the Two-Part test setup window

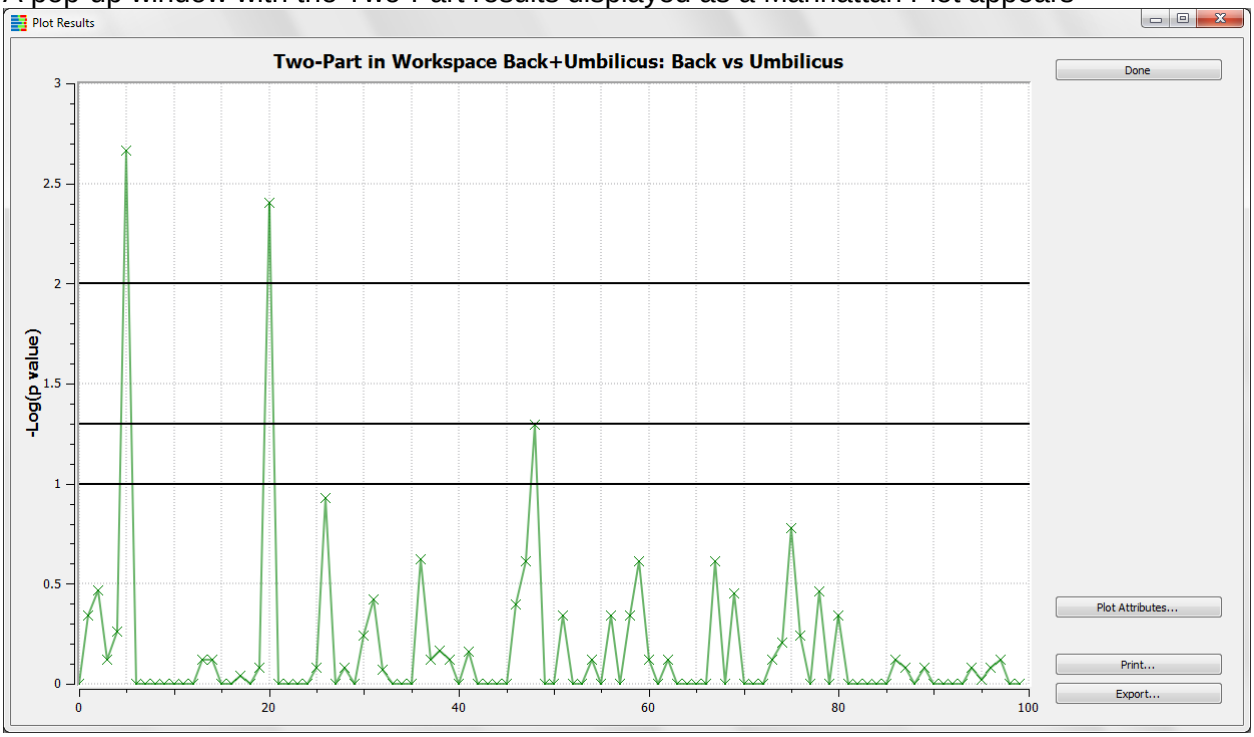


Select “Back” for the **Category 1 Filter**
Select “Umbilicus” for the **Category 2 Filter**
Click **Calculate**
Click **Plot**

© 2012-2013 Charles E. Robertson and Incubix, Inc.



A pop-up window with the Two-Part results displayed as a Manhattan Plot appears



The Manhattan Plot displays logarithmically transformed p-values, with higher peaks representing lower (more significant) p-values. The horizontal lines represent p-values of 0.10, 0.05, and 0.01. Inclusion of the p=0.10

© 2012-2013 Charles E. Robertson and Incubix, Inc.

line is intended to highlight taxa that are approaching significance in an analysis. The x-axis represents the alphabetical position, by number, of each OTU name in the Two-Part setup dialog above.

In the Manhattan Plot, the first significant peak (position 6) corresponds to Corynebacterium, which have a higher proportion and relative abundance in the umbilicus samples. The second peak (position 21) represents Propionibacterium that is present at a higher proportion and relative abundance in the back samples. The third peak that approaches significance (position 49) represents Anaerococcus. This taxon is not seen in many of the libraries generated from back samples, and thus is present at higher proportion and relative abundance in the umbilicus samples.

Data can be exported from the **Plot Results** window as tab delimited text using the export button (available in all graphics windows). The data incorporated for each taxon in the Two-Part statistic are summarized for each category. The number of samples with sequences belonging to an OTU within each category is designated “m”, proportion of positive libraries in a category “p”, and median relative abundance “med”.

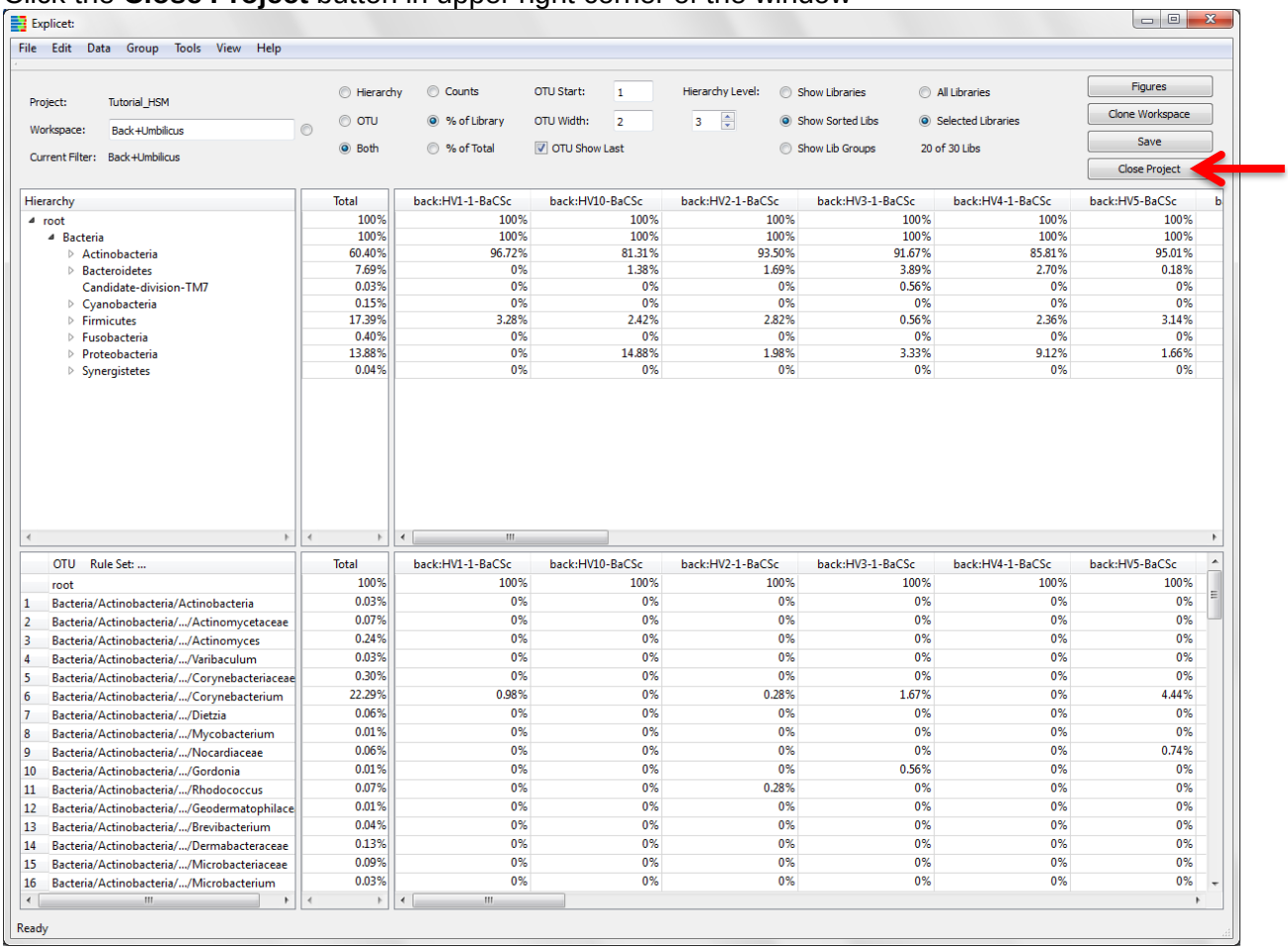
You may choose to save the Two-Part test as a figure. To do so, continue as shown earlier in the stacked bar chart example; close the graphics window, and select **Save As Figure** in the **Two-Part** window.

This tutorial has provided a quick overview of how to use Explicet. For more complete information on Explicet capabilities, please see the Explicet Handbook. We will now save our changes and close the project.

© 2012-2013 Charles E. Robertson and Incubix, Inc.

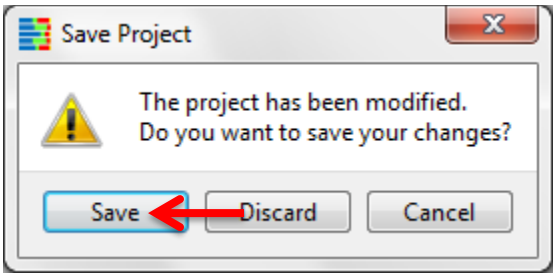
XIII. Close the Project

Click the **Close Project** button in upper right corner of the window



A pop-up window will open

Click **Save**



The Explicet window will close, and all of the OTU data, metadata, and figures are now saved within the project file.

Thus ends a basic overview of some functions contained in Explicet. Please do not hesitate to ask questions or make suggestions via our online Explicet forum. The Explicet forum link can be found on our web site: www.explicet.org.

© 2012-2013 Charles E. Robertson and Incubix, Inc.