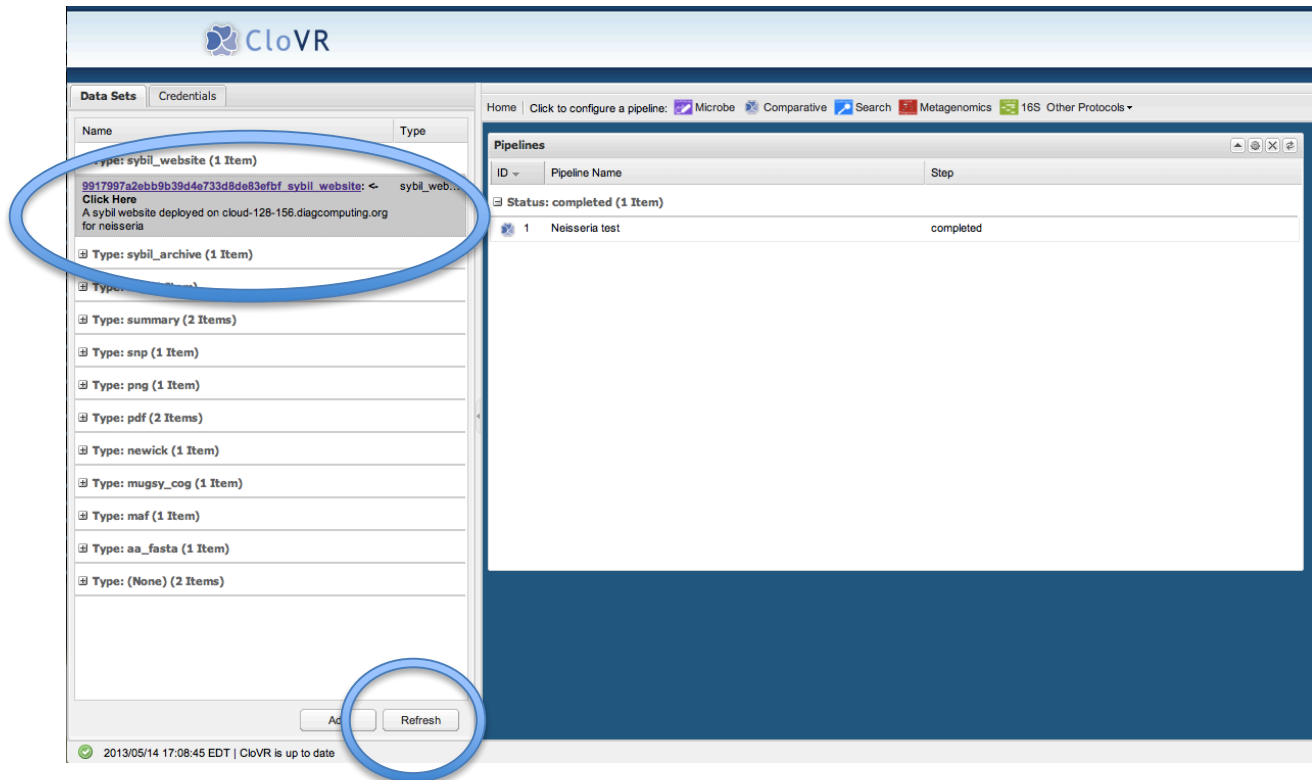# Sybil activity – ASM 2016 CloVR workshop

## Accessing the Sybil website on your VM

Find the Sybil website dataset in your dataset panel.

This could require clicking 'Refresh' at the bottom.

Click the link that should look like 'blablabla_sybil_website' where blablabla is a big identifier.



## Sybil home page

This brings you to the Sybil home page. At the bottom right you will see the genomes included in this particular Sybil instance. We are going to compare five complete gap-free genomes of the Gram-negative bacterium *Neisseria meningitidis*. Each genome has only one chromosome of ~2.2Mb. Please note that CloVR and Sybil also work with genomes containing multiple replicons as well as draft genomes, in which case replicon/contig information would be reported in searches and displays.

# Protein search

1. From the Sybil homepage click on the Protein/Cluster Search button.

This will take you to the protein search page.
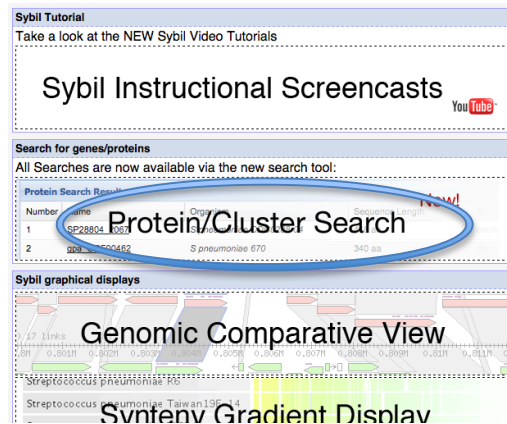


2. First we'll search for a gene by keyword. Enter the word 'protein' in the Search box and click 'Search' at the bottom.
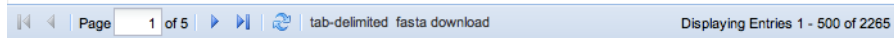


This will bring up genes that contain the word 'protein' in their gene product annotation.

3. Click 'Uncheck All' then select only '*Neisseria meningitidis* MC58' and click 'Search' to restrict the search to only that specific genome/strain.

<u>Search form tips</u>
- Browse the results by clicking the page control buttons at the bottom of the screen.
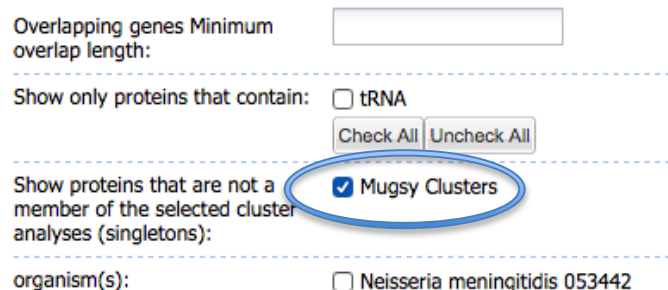


- Hide the search window by clicking the [«] button at the top of the search form.
- Hide columns by clicking the down arrow in the column header and unchecking the box. The arrow will appear when you hover over the header with your mouse. [Organism ▼]
- Reorder columns by clicking and dragging the column header.
- Sort by the 'Name' column by clicking on the column header.

3. Next, let's search for genes that are unique to the '*Neisseria meningitidis* MC58' genome.

The CloVR Comparative pipeline has grouped genes predicted in all genomes into Mugsy-based clusters of orthologs. This method of clustering makes use of the local synteny obtained from the Mugsy whole genome multiple alignment (see this reference for details: Angiuoli SV, Dunning Hotopp JC, Salzberg SL, Tettelin H. (2011) Improving pan-genome annotation using whole genome multiple alignment. BMC Bioinformatics 12: 272). This approach was briefly described in Sonia Agrawal's presentation.

There are many other methods for generating clusters of orthologs including Jaccard-based clusters of orthologs (JOCs)(Crabtree J, Angiuoli SV, Wortman JR, White OR (2007) Sybil: methods and software for multiple genome comparison and visualization. Methods Mol. Biol. 408: 93-108) and Blast Score Ratio-based clusters (see Tracy Hazen's presentation later today). Multiple clustering methods can be loaded into the same database in which case Sybil will allow selection of which one to use for a given search.

- Check the box next to 'Mugsy Clusters'. This will limit our search results to those genes that are singletons in the Mugsy protein clustering. This means that they are not members of a Mugsy protein cluster.
- Click 'Search'.



This will bring up all MC58 genes that contain the word 'protein' somewhere in their annotation but that are not contained in a Mugsy gene cluster. This means that these genes are specific to a particular genome, i.e. singletons (note that some of these genes might still have a BLAST hit to another genome, for example if they are located in repeated regions).

5.  Click on the name of the gene 'MC58:NMB0016' which is annotated as a 'hypothetical protein' to go to the protein report page.



# Protein report

The protein report page provides summary information about a gene in addition to providing graphical representations of the genomic context and the BLAST results.



1.  Click on the '10kb' button just above the image. This will redraw the image with 10kb on either side of the gene.
2.  Click on 'tRNA' just above the 10kb button to show any tRNA present in this region. One appears as a small empty box between NMB0011 and NMB0012.
3.  Click on the tRNA box itself to bring up its popup window with coordinates.
4.  Click on some genes in the display to pull up additional information in popup windows.
5.  Click on the gene labeled in red – MC58:NMB0016 – to bring up its popup window.
6.  Notice that no clusters are listed for this gene since it was selected as a singleton.

7. Click on the '10kb' link just below 'Center on gene in new comparative view'.



Clicking this link will take you to the 'Genomic Comparative View' and will center your view on this gene's coordinates.

# Genomic comparative view

The genomic comparative view provides a graphical representation of the genomic context of multiple genomes.

1. The first step is pulling in some genomes to search. This is done by selecting sequences from the 'Sequences' table and dragging those sequences into the table below – labeled 'Search Mode'. You can select multiple sequences using the shift or command/ctrl key. For this exercise, select the remaining 4 genomes – 053442, FAM18, Z2491 and alpha14 – and drag them into the table.

2. Click 'Draw (Search Mode)' at the bottom of the form.



This will draw the reference region we selected – our gene of interest with 10kb on either side – with matching genomes below.  Once the image has loaded, click the [«] button to hide the search form, leaving just the picture on the screen.
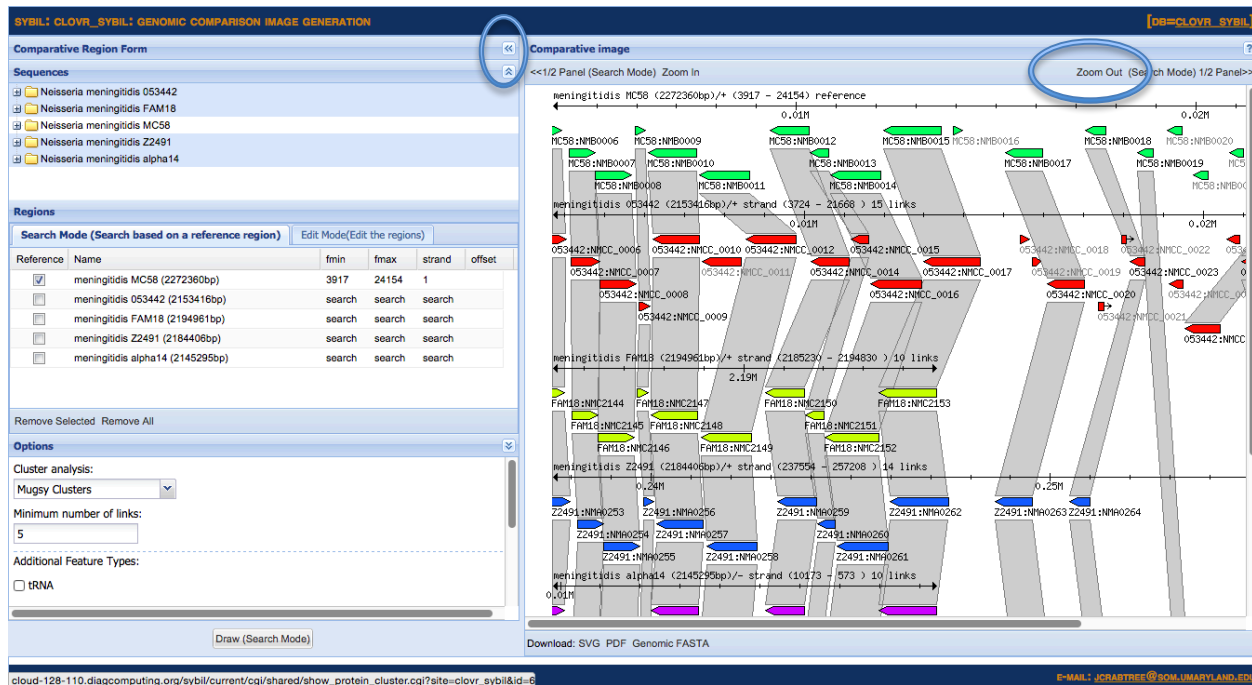
3. Click the box labeled 'Zoom Out' at the top right hand side of the screen to extend the left and right flanks. Notice that now a second segment of the FAM18 genome is displayed on the right side. It only appeared now because enough flanking genes with Mugsy cluster links are available to display it. The default number of links is 5 but it can be changed in the search form.

4. The gene names in grey font are singletons (not members of a Mugsy cluster).

5. Notice the breaks and coordinates in the FAM18 and alpha14 genomes.

Genomic comparative view tips

• Click on genes to see more information about them.

• Click on cluster links (grey polygons) to see more information about them, including a list of all the cluster members whether they are shown on the current display on not.

• Zoom in/out and pan controls are available in the sub-title pane at the top. Note that in 'Search Mode' this will zoom/pan the reference and re-search the other genomes for matching regions.

• Export the image you are seeing to SVG or PDF by clicking the export buttons at the bottom.

# Gradient display

Lastly we'll look at the whole-genome gradient display.

1. Click on the link at the top right hand side of your browser window labeled 'DB=neisseria'. This link is available on all views and will take you back to the Sybil homepage.
2. Click on the button labeled 'Synteny Gradient Display'
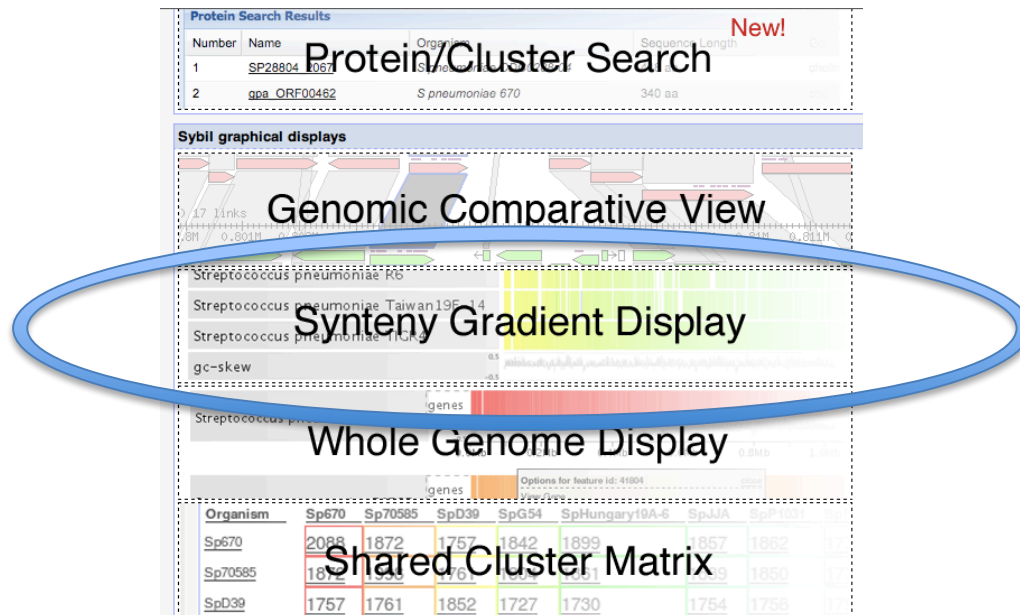


3. Click on the 'Add All Organisms' button at the end of the list of organisms in the 'Sequence Selection' box. The organism in position one will be the default reference.
4. Check the box next to 'Color those genes with multiple copies (paralogs) in the query black'.
5. Check the box next to 'Show %GC graph for reference(s)'.
6. Click 'Draw'.

This will draw a gradient display comparing the reference genome to the other genomes selected.
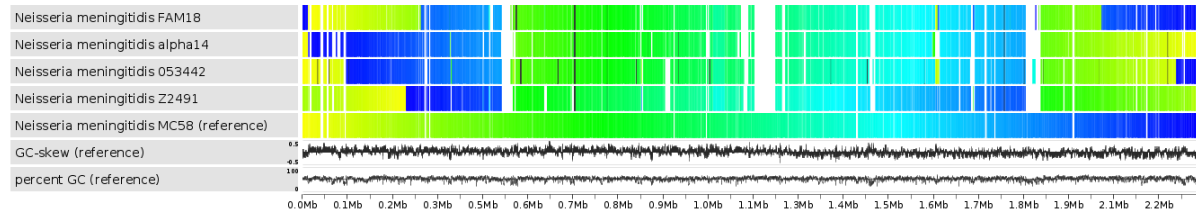


This particular display shows MC58 as the reference genome. Genes from the remaining 4 genomes are drawn above the MC58 gene but are colored based on their position in their native genome.

Gradient display tips
• Clicking on genes in the reference brings up popups about those genes and you can link to protein report/cluster report pages.
• Export the images using the 'PNG', 'SVG', 'PDF' and 'JPEG' buttons.

_____

**View the Sybil screencast tutorials (linked from the front page)**
http://www.youtube.com/user/SybilScreencasts

**Sybil sourceforge website**
http://sybil.sourceforge.net

**Sybil publications**
Riley DR, Angiuoli SV, Crabtree J, Dunning Hotopp JC, Tettelin H (2012) Using Sybil for interactive comparative genomics of microbes on the web. Bioinformatics 28: 160-166.
Crabtree J, Angiuoli SV, Wortman JR, White OR (2007) Sybil: methods and software for multiple genome comparison and visualization. Methods Mol. Biol. 408: 93-108.