

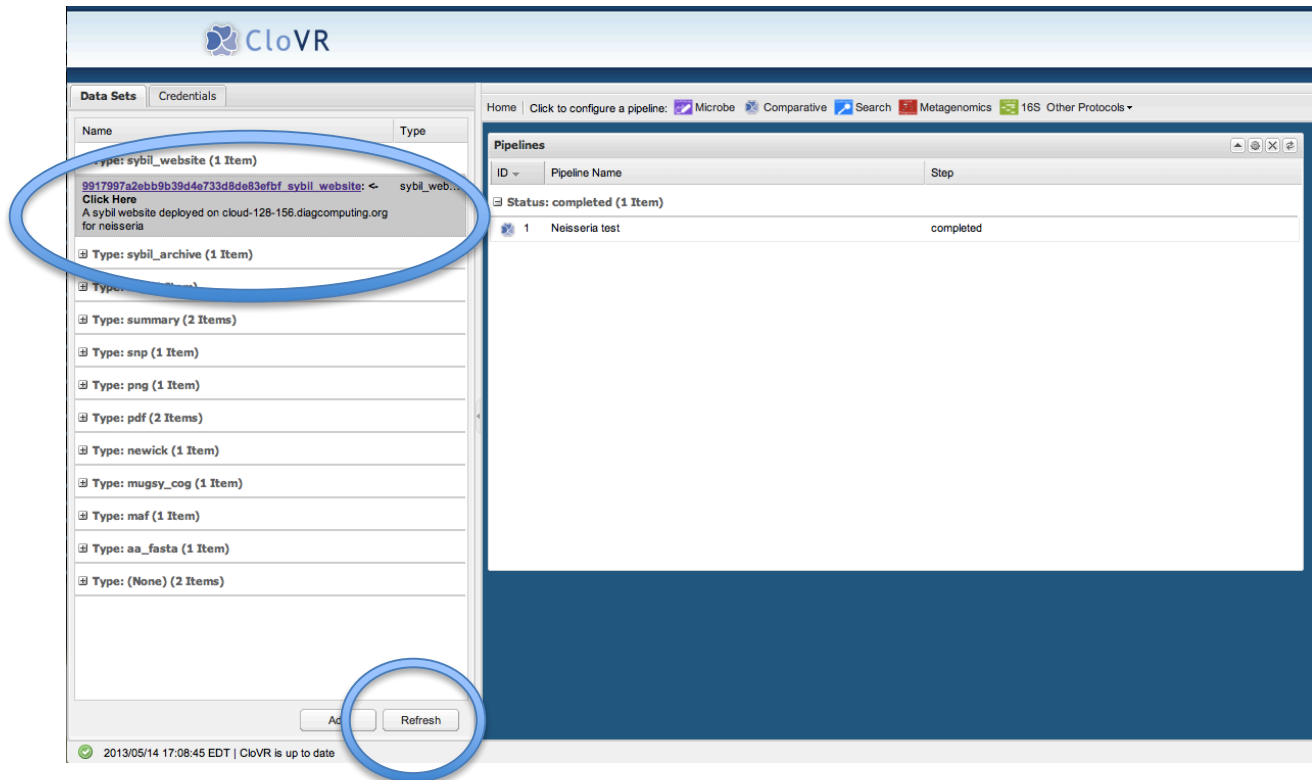
Sybil activity – ASM 2016 CloVR workshop

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.



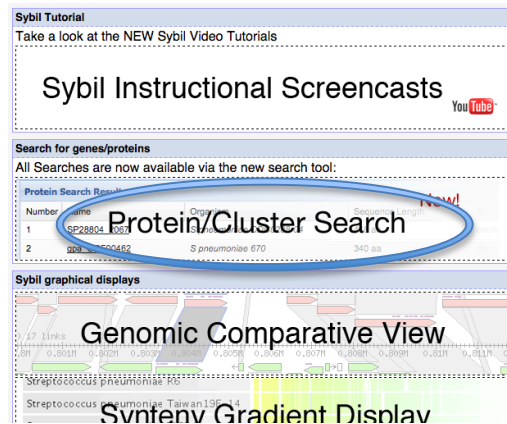
Sybil home page

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.

Number	Name	Organism	Sequence Length	Gene Symbol	Gene Product
1	Z2491.NMA1480	N meningitidis Z2491	397 aa		hypothetical protein
2	MC58.NMB0503	N meningitidis MC58	143 aa		hypothetical protein
3	FAM18.NMC1757	N meningitidis FAM18	405 aa		hypothetical protein
4	053442.NMCC_0420	N meningitidis 053442	371 aa	pilM	type IV pilus assembly protein PilM
5	053442.NMCC_1182	N meningitidis 053442	47 aa		hypothetical protein
6	FAM18.NMC1469	N meningitidis FAM18	725 aa	lbpB	lactoferrin-binding protein
7	FAM18.NMC0869	N meningitidis FAM18	240 aa		phage-like protein
8	FAM18.NMC1254	N meningitidis FAM18	134 aa		hypothetical protein
9	FAM18.NMC2085	N meningitidis FAM18	105 aa		hypothetical protein
10	Z2491.NMA1066	N meningitidis Z2491	218 aa		hypothetical protein
11	Z2491.NMA2058	N meningitidis Z2491	414 aa		cell division protein
12	053442.NMCC_0941	N meningitidis 053442	133 aa		hypothetical protein
13	Z2491.NMA1695	N meningitidis Z2491	262 aa		fimbrial assembly protein
14	FAM18.NMC0332	N meningitidis FAM18	113 aa		hypothetical protein
15	Z2491.NMA1854	N meningitidis Z2491	168 aa		hypothetical protein
16	053442.NMCC_0670	N meningitidis 053442	294 aa		hypothetical protein
17	alpha14.NMO_1966	N meningitidis alpha14	333 aa	tlpA3	thiamine transport system substrate-binding protein
18	MC58.NMB0952	N meningitidis MC58	82 aa		hypothetical protein
19	FAM18.NMC1572	N meningitidis FAM18	197 aa		hypothetical protein
20	053442.NMCC_1053	N meningitidis 053442	253 aa		hypothetical protein
21	alpha14.NMO_1561	N meningitidis alpha14	94 aa		hypothetical protein
22	Z2491.NMA0899	N meningitidis Z2491	124 aa		hypothetical protein
23	053442.NMCC_1922	N meningitidis 053442	157 aa		hypothetical protein
24	Z2491.NMA1862	N meningitidis Z2491	78 aa		hypothetical protein
25	MC58.NMB0861	N meningitidis MC58	181 aa		hypothetical protein
26	FAM18.NMC0125	N meningitidis FAM18	123 aa	rpsL	30S ribosomal protein S12

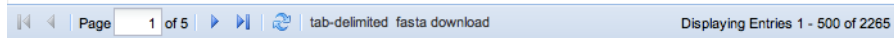
This will bring up genes that contain the word 'protein' in their gene product annotation.


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

☐ Neisseria meningitidis 053442
☐ Neisseria meningitidis FAM18
☒ Neisseria meningitidis MC58
☐ Neisseria meningitidis Z2491
☐ Neisseria meningitidis alpha14

Search form tips

- Browse the results by clicking the page control buttons at the bottom of the screen.



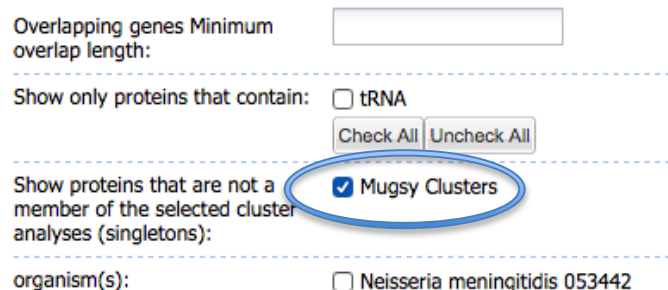
- Hide the search window by clicking the button at the top of the search form.
- Hide columns by clicking the down arrow in the column header and unchecking the box. The arrow will appear when you hover over the header with your mouse. 
- Reorder columns by clicking and dragging the column header.
- 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 syntenicity 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'.



Overlapping genes Minimum overlap length:

Show only proteins that contain: ☐ tRNA

Show proteins that are not a member of the selected cluster analyses (singletons): ☒ Mugsy Clusters

organism(s): ☐ *Neisseria meningitidis* 053442

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).

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

Protein Search Results				
Number	Name ▲	Organism	Sequence Length	Gene Product
123	MC58:NMB0016	<i>N meningitidis MC58</i>	78 aa	hypothetical protein
139	MC58:NMB0046	<i>N meningitidis MC58</i>	74 aa	hypothetical protein
57	MC58:NMB0067	<i>N meningitidis MC58</i>	225 aa	polysialic acid capsule biosynthesis protein SiaD
156	MC58:NMB0093	<i>N meningitidis MC58</i>	28 aa	hypothetical protein
112	MC58:NMB0099	<i>N meningitidis MC58</i>	47 aa	hypothetical protein

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.

SYBIL: CLOVR_SYBIL: PROTEIN MC58:NMB0016

[DB=CLOVR_SYBIL]

PROTEIN PROPERTIES

properties of MC58:NMB0016 ?

property	value
organism	<i>Neisseria meningitidis MC58</i>
product name	hypothetical protein
sequence length	78 aa
created	2015-05-13 16:55:00
last modified	2015-05-13 16:55:00

DATABASE REFERENCES

database refs for MC58:NMB0016 ?

database	accession	version
----------	-----------	---------

PROTEIN CLUSTERS

clusters of which MC58:NMB0016 is a member ?

cluster	program	algorithm	analysis description
---------	---------	-----------	----------------------

GENOMIC CONTEXT

genomic context of the gene show_protein_clusters ?

sequence
meningitidis MC58 (2272360bp) (2.27 Mb)
location
13917-14154
strand
+

additional feature types:
☐ tRNA

display at most: 5 kb 10 kb 15 kb 25 kb 50 kb on either side of MC58:NMB0016:

SEQUENCE

the amino acid sequence in FASTA format ?

Download Protein FASTA
>MC58:NMB0016 polypeptide
MKNSFCGQFACCAAGILLFFTEWLPMSLRTGILNRFERKVCLELVEMVRCRHECLISAYRQSAHPNLCIS
GGKNEICG

Download Nucleotide FASTA
>MC58:NMB0016 nucleotide
ATGAAGAATAGTTTTGTGGACAGTTGCTTGTGTGCAAAATGGCATCCTACTTTTCTTACCGAATGCG
TGCCGATGTCCTTAAGAACCGGAATACGTGGAGGTTTGAGAGGAAAGTGTGTTGGAACTTGTGAAAT
GGTCAGGTGTCGACAGAAATGCTTATTTCTGCATATCGGCAGAGTGCGCATCCGAATTTGTGTATAGT
GGTGGAAAAATGAGATTTGCGGGTAA

SYBIL WEB SITE: SYBIL_SOURCEFORGE.NET

E-MAIL: JCRANTREE@BOM.JMARIYLAND.EDU

- Click on the '10kb' button just above the image. This will redraw the image with 10kb on either side of the gene.
- 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.
- Click on the tRNA box itself to bring up its popup window with coordinates.
- Click on some genes in the display to pull up additional information in popup windows.
- Click on the gene labeled in red – MC58:NMB0016 – to bring up its popup window.
- 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'.

50 kb on either side of MC58:NMB0016:

0.02M

2 MC58:NMB0015 MC58:NMB0017 MC58:NMB0018 MC58:NMB0020 MC58:NMB0023 MC58:NMB0026

NMB0013 MC58:NMB0016 MC58:NMB0019 MC58:NMB0022 MC58:NMB0025

58:NMB0014

Click to drag

MC58:NMB0016 (view gene)

gene product name : hypothetical protein

fmin : 13917

fmax : 14154

Center on gene in new comparative view : Pad: 1kb | 2.5kb | 5kb | 10kb | 15kb | 25kb | 50kb

Center on gene in new comparative view and search all other genomes : Pad: 1kb | 2.5kb | 5kb | 10kb | 15kb | 25kb | 50kb

download polypeptide : [sequence](#)

download nucleotide : [sequence](#)

Clusters

ERKVCLELVEMVRC

TTTGCTT
TACTGTG
TATTCT
TAA

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.

SYBIL: CLOVR_SYBIL: GENOMIC COMPARISON IMAGE GENERATION

Comparative Region Form

Sequences

- Neisseria meningitidis 053442
- Neisseria meningitidis FAM18
- Neisseria meningitidis MC58
- Neisseria meningitidis Z2491
- Neisseria meningitidis alpha14

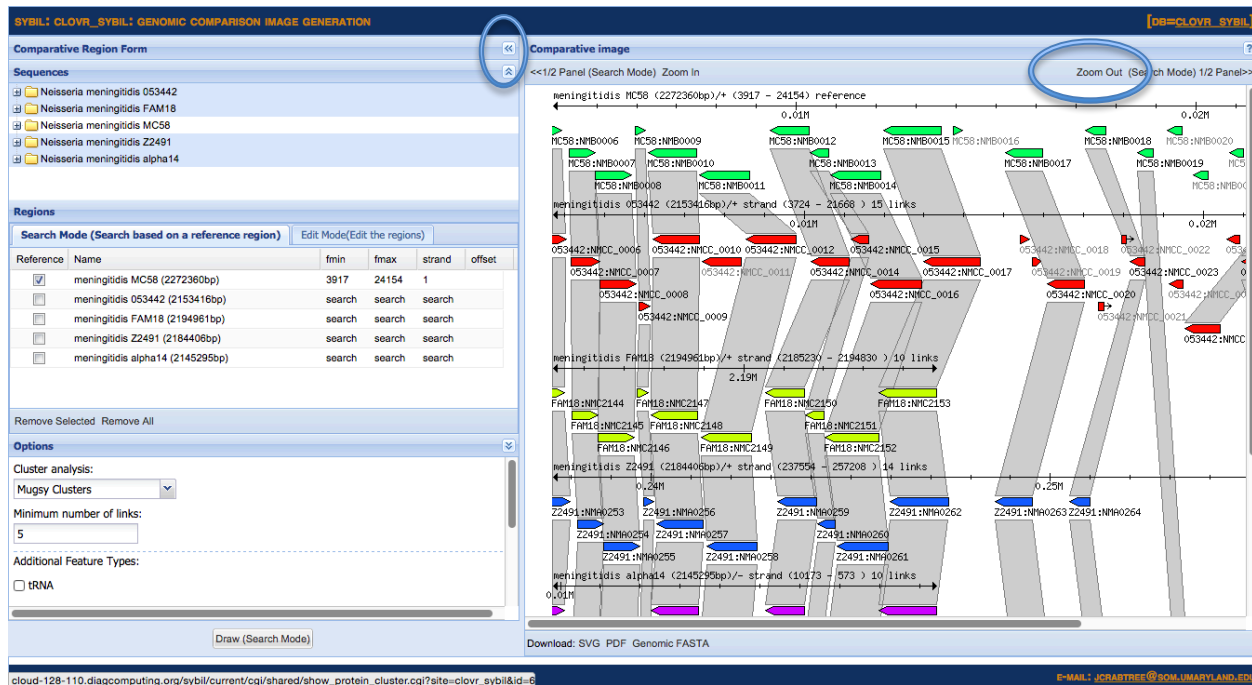
Regions

Search Mode (Search based on a reference region) Edit Mode(Edit the regions)

Reference	Name	fmin	fmax	strand	offset
<input checked="" type="checkbox"/>	meningitidis MC58 (2272360bp)	3917	24154	1	
<input type="checkbox"/>	meningitidis 053442 (2153416bp)	search	search	search	
<input type="checkbox"/>	meningitidis FAM18 (2194961bp)	search	search	search	
<input type="checkbox"/>	meningitidis Z2491 (2184406bp)	search	search	search	
<input type="checkbox"/>	meningitidis alpha14 (2145295bp)	search	search	search	

Remove Selected Remove All

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.



- 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.
- The gene names in grey font are singletons (not members of a Mugsy cluster).
- Notice the breaks and coordinates in the FAM18 and alpha14 genomes.

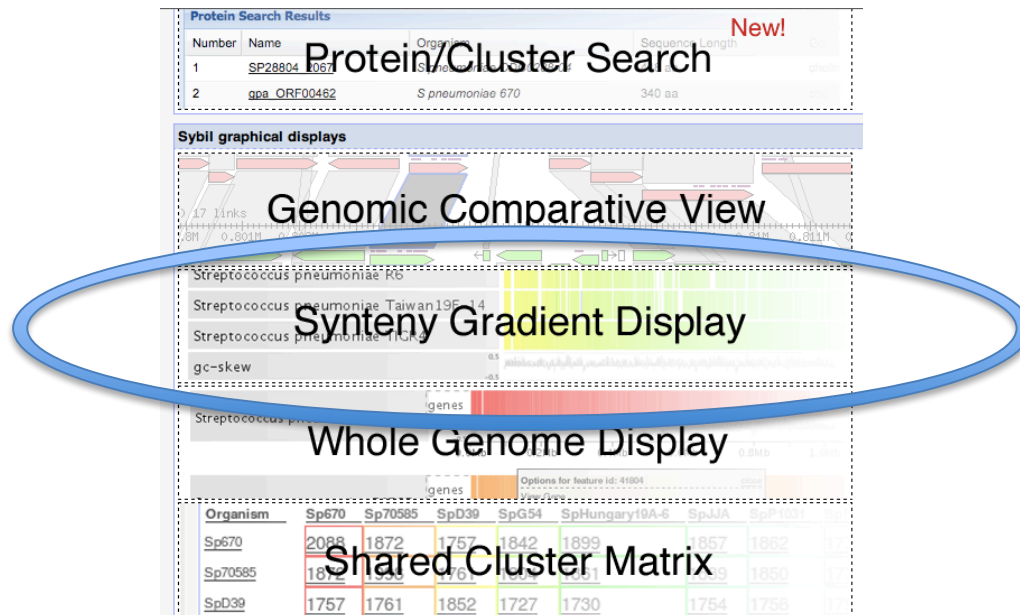
Genomic comparative view tips

- 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 or 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'.

Sequence Selection [hide] Select sequences for display ?

Organism Order	Organism Name	Assemblies Selected
3 add/remove	Neisseria meningitidis 053442	1/1 sequence(s) selected (2153416bp) [change]
5 add/remove	Neisseria meningitidis FAM18	1/1 sequence(s) selected (2194961bp) [change]
1 add/remove	Neisseria meningitidis MC58	1/1 sequence(s) selected (2272360bp) [change]
2 add/remove	Neisseria meningitidis Z2491	1/1 sequence(s) selected (2184406bp) [change]
4 add/remove	Neisseria meningitidis alpha14	1/1 sequence(s) selected (2145295bp) [change]

[Add All Organisms](#) [Remove All Organisms](#)

Options [hide]

Choose the analysis to create links: [Mugy Clusters 2](#)

☐ Use each organism as a reference

☒ Color those genes with multiple copies (paralogs) in the query black (default: no color)

☐ 1 sequence per line (rather than concatenating the sequences together)

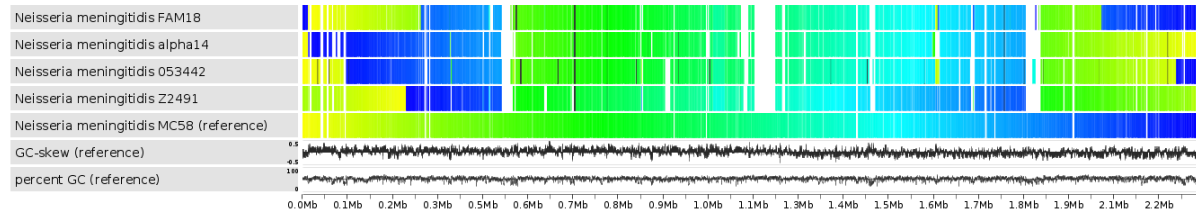
☒ Show %GC graph for reference(s)

☐ Show GC skew graph for reference(s)

☐ Show atypical nucleotide composition graph for reference(s)

image width: 1000px

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.