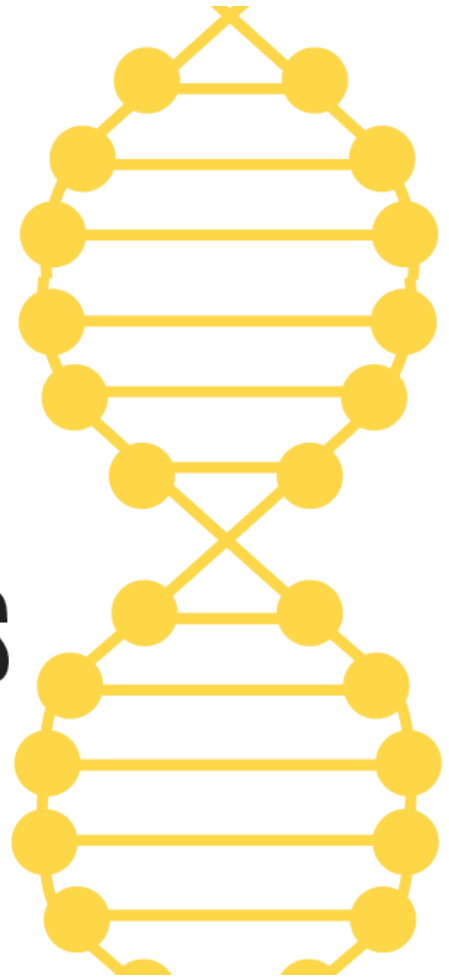




**WS16:**

**Do-it-yourself  
Microbial Genome  
Sequence Analysis**



# **Hands-On Exercises**

# Agenda

8:15	Introduction/Overview	Michelle Giglio
8:45	CloVR background	W. Florian Fricke
9:15	Hands-on: Start CloVR	W. Florian Fricke
9:45	Break	
9:55	Hands-on: Start CloVR-Microbe	W. Florian Fricke
10:15	Hands-on: Start CloVR-Comparative	W. Florian Fricke
10:45	Genome Annotation Pipeline	Michelle Giglio
11:45	Lunch	
12:30	Hands-on: CloVR-Microbe output	W. Florian Fricke
1:00	Hands-on: Manatee	Michelle Giglio
1:30	Comparative Analysis Pipeline	Cesar Arze
2:00	Break	
2:10	Hands-on: CloVR-Comparative output	Hervé Tettelin
3:00	Future Directions	Cesar Arze
3:15	Applications with <i>S. pneumoniae</i> and <i>M. abscessus</i>	Hervé Tettelin
3:45	Applications with <i>E. coli</i>	Tracy Hazen
4:15	End of workshop	

# CloVR Hands-On exercises

We have started Virtual Machines for each of you to use during the workshop. The following steps will allow you to access these VMs and use them to run the analysis pipelines.

## Start CloVR -

1. Go to the DIAG website: <http://diagcomputing.org/>
2. login using the DIAG credentials at the beginning of this packet

**NOTE:** The steps #3 and #4 would be followed if you were starting a VM on your own, however, since we have already started the VMs for you, for the purpose of keeping things flowing on time for the workshop, **DO NOT** do steps 3 and 4 now. They are here for your reference for when you get back home and try this on your own.

3. Click on: 'My Account'. Then choose: 'Start CloVR'
4. This is the CloVR launch page. Click 'Launch CloVR'

5. Access the VMs we have running for you by clicking on 'My Account'. Then choose: 'Monitor VMs'.
  6. There should be one VM displayed on this page with the status 'running'. Click on that row to access the VM.
- 

## Run CloVR Microbe – These steps will start the genome assembly and annotation pipeline

### A. Upload data

1. On the CloVR dashboard, in the datasets panel, click on the 'Add' button, lower left side
2. A box will pop up. This box provides several options for providing CloVR with data including Browsing your computer, directing CloVR to a web location where the information is stored, or using FTP. For today we will use the url option. Paste or type this url into the box (for copying and pasting, an electronic version of this document is on the [clovr.org/workshop](http://clovr.org/workshop) web page): **<http://cb2.igs.umaryland.edu/microbe.tgz>** (Make sure that there are no trailing white spaces at the end of the url which generally happens if you cut and paste.)
3. Choose file type: 'Nucleotide FASTQ'
4. Give the file a name, make sure there are no white spaces, use something short and easy.  
For example, name: 'illumina\_test'

## B. Configure the pipeline

1. Select the desired pipeline from the applications header panel, choose: 'Microbe'
2. Tell the pipeline which data file to use by selecting: 'Sequencing Dataset(s)'  
and then: 'illumina\_test' (or whatever name you used in step 4 above)
3. Select a CloVR Microbe pipeline track. We'll do 'Assembly + Annotation' for this exercise.
4. Fill in the fields:
  - output prefix: the text you provide here will be a prefix on all output files allowing you to easily identify them. For today use 'nmrun'
  - organism: type the genus and species of the organism the data comes from: 'Neisseria meningitidis'
  - DB name: choose a name for the database that will be generated to house the annotation data. For today use: 'nmen'
  - Manatee username: choose a user name for accessing the Manatee database (more on Manatee later) For today use: 'asmguest'
  - Manatee password: choose a password for accessing the Manatee database. For today use: 'stiontio65'
5. Provide a pipeline description, for example: 'illumina\_run' (again no white spaces)
6. Click 'Submit'

---

## Run CloVR Comparative

### A. Upload data

1. On the CloVR dashboard, in the datasets panel, click on the 'Add' button
2. To provide CloVR with the data to use in the pipeline, paste or type this url:  
**<http://cb2.igs.umaryland.edu/compare.tgz>**  
Again remember to make sure there are no trailing white spaces at the end  
One can also use the taxonomy browsing feature to choose genomes to include or enter genome accession numbers.
3. Choose file type: 'GenBank'.
4. Give the file a name, no white spaces. For example, name: 'neisseria\_5genomes\_test'

### B. Configure the pipeline

1. Select the desired pipeline from the applications header panel, choose: 'Comparative'
2. Tell the pipeline which data file to use by selecting Input GenBank Tags:  
'neisseria\_5genomes\_test'
3. Type something in for pipeline description, for example: 'neisseria\_test' (no white spaces)
4. Click 'Run'

# Manatee Demonstration

1. To get started with Manatee go to **<http://manatee.igs.umaryland.edu>**

Due to constraints on our resources of running 50 CloVR VMs at one time, we could not also start up Manatee VMs for this part of the workshop. We will instead use Manatee running on the IGS web server. For reference when you get back home, here is a link to instructions for accessing the Manatee VMs to view the databases resulting from the CloVR Microbe pipeline runs: <http://tinyurl.com/ManateeDocs>

2. Type in these credentials:

username 'asmguest'

password 'stiontio65'

database 'nmen'

This information is also on your credential sheet at the beginning of this packet.

3. You should get a 'Welcome to Manatee' page. (Figure 1) There are numerous search, display, and file download options on this page. Explore this page.

The screenshot displays the Manatee web interface. At the top, there is a section titled 'ACCESS LISTINGS' with a list of links: 'Annotation Tools', 'Genome Viewer', 'Genome Calculations', 'Role Category Breakdown', and 'Overlap Summary'. Below this is a section titled 'ACCESS GENE CURATION PAGE' with a 'gene\_id' input field. Next is a section titled 'SEARCH GENES BY PROTEIN NAME' with a 'protein name' input field. This is followed by a section titled 'CHANGE ORGANISM DATABASE' with a 'database' input field. On the left side of the interface, there are 'submit' and 'reset' buttons. Below the input fields, there are three radio button options: 'BLASTN' (blast nucleotide sequence against nucleotide sequence of predicted genes in this genome), 'BLASTP' (blast protein sequence against amino acid sequence of predicted genes in the genome), and 'TBLASTN' (blast protein sequence against the entire genome sequence). Below these is a text area labeled 'Paste nucleotide or protein sequence below:'. At the bottom, there is a section titled 'Data file downloads (potentially long download times)' with a list of links: 'GO Dumper' (Tab delimited file of GO annotation), 'Nucleotide Sequence Dumper' (Multifasta File), 'Protein Sequence Dumper' (Multifasta File), 'Coordinate Dumper' (Tab delimited file of gene coordinates), 'Whole Genome Dumper' (Nucleotide Fasta File), 'Annotation Dumper' (Tab delimited file of annotation), 'Genbank Dumper' (For use in Artemis, BioPerl, etc.), 'GFF3 Dumper' (For use in GBrowse, JBrowse, etc.), and 'TBL Dumper' (For submission to NCBI, along with the nucleotide FASTA).

Figure 1

4. Let's click on 'Annotation Tools' at the top of the Welcome page. This gives you a page with more search options (Figure 2). Let's leave the default selection: 'all genes, ordered by role category'. Click 'Submit'.

submit

reset

ACCESS GENE LISTS

molecule: All molecules

all genes, ordered by role category

main role category Unclassified

single role category role\_id

select coordinate range:

end5:end3:

SEARCH GENES BY:

gene\_id / locus:

protein name:

gene symbol:

EC number:

Comment:

Figure 2

5. We then get a Gene List page (Fig 3). Click on a gene\_id to get to a Gene Curation Page. (I will demo **nmen\_3**)

(snippet of the list)

Biosynthesis and degradation of surface polysaccharides and lipopolysaccharides										Role id: 90	
C	seq id	gene_id	locus	end5	end3	gene name	gene symbol	ec	other roles	start_edit	
	MC58_region_for_assembly	nmen_9		10245	9322	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	lpxC	3.5.1.-			
	MC58_region_for_assembly	nmen_7		6212	4941	glycosyl transferases group 1 family protein					

Figure 3

6. The Gene Curation Page – let's look at **nmen\_3** (from the 'Biosynthesis of murein sacculus and peptidoglycan' section).

This page displays all of the automatically assigned annotations as well as the supporting evidence that was used to make them. (This is a long page – below I have showed a sampling of segments.) Users can change the annotations by entering information into the fields and clicking 'submit'. (Although for this workshop you have only read access to the database.) Explore the information on this page.

**GENE IDENTIFICATION** submit

gene name:

gene\_sym:

EC number(s):  [EC name](#)

EC GO suggestions: [GO:0008760](#)  UDP-N-acetylglucosamine 1-carboxyvinyltransferase activity

private comment:

public comment:

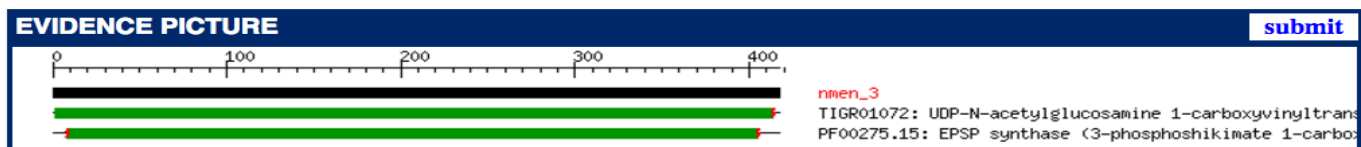
Assigned By:  Date:

Protein name and other annotations.

**GENE ONTOLOGY** submit

delete	go id		assigned	date	evidence
<input type="checkbox"/>	<a href="#">GO:0008760</a> [ <a href="#">add</a> ] [ <a href="#">edit</a> ]	(F) UDP-N-acetylglucosamine 1-carboxyvinyltransferase activity			IEA: TIGR01072 [ <a href="#">make ISS</a> ]
<input type="checkbox"/>	<a href="#">GO:0009252</a> [ <a href="#">add</a> ] [ <a href="#">edit</a> ]	(P) peptidoglycan biosynthetic process			IEA: TIGR01072 [ <a href="#">make ISS</a> ]

Gene Ontology annotations



Evidence Summary graphical display

**HMM** submit | all hmms

[TIGR01072: UDP-N-acetylglucosamine 1-carboxyvinyltransferase](#) gene\_sym: murA ec#: 2.5.1.7 role\_id: 89 tc\_num: none [ [Add To Annotation](#) ]

Isology: **equivalog**  
Total score: 607.5 Trusted cutoff: 392.70 Gathering cutoff: 392.70 Noise cutoff: 234.80 Total expect: 1e-182  
Trusted cutoff2: 392.70 Gathering cutoff2: 392.70 Noise cutoff2: 234.80

Coords	HMM Coords	Score	Expect	[ <a href="#">Add To GO Evidence</a> ]
1-414	1-414 / 416	607.4	1.1e-182	<input type="checkbox"/>

[GO:0008760](#)  UDP-N-acetylglucosamine 1-carboxyvinyltransferase activity (F)

[GO:0009252](#)  peptidoglycan biosynthetic process (P)

HMM evidence

**BER SKIM** Show all submit

View BER Searches search date: Fri Jun 3 16:58:48 2016

accession	%ID	length	description	taxon	p-value
UniRef100_D3A2J8	91.1	416	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	Neisseria	0
UniRef100_E3D4Z4	99.3	416	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	Neisseria meningitidis	0
UniRef100_R0U805	99.0	416	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	Neisseria meningitidis NM82	0
UniRef100_H8DZU0	87.8	419	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	Kingella kingae PYKK081	0
UniRef100_C6SIQ2	98.5	397	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	Neisseria meningitidis alpha275	0
UniRef100_R0U571	99.0	416	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	Neisseria meningitidis 73696	0
UniRef100_F9EY7M	92.1	416	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	Neisseria macacae ATCC 33926	0

BER evidence

**BER SKIM** Show all submit

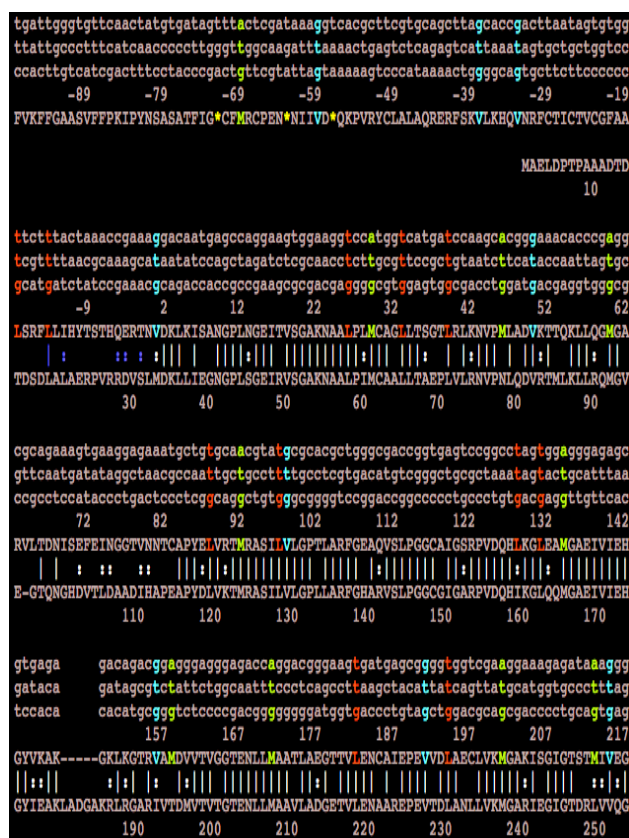
View BER Searches search date: Wed Apr 27 11:14:39 2011

accession	%ID	length	description	taxon	p-value
UniRef100_Q9HKE3	49.5	371	FAD-dependent oxidoreductase	Pseudomonas aeruginosa	7e-98
UniRef100_D3P9W8	56.6	376	FAD-dependent oxidoreductase	Acetivibrio sp. 5810	1e-115
UniRef100_D1LDB1	55.7	369	FAD-dependent oxidoreductase	Cupriavidus metallidurans DSM	1e-114
UniRef100_E5V5L8	55.5	374	FAD-dependent oxidoreductase	Neisseria meningitidis EPS	1e-113

BER evidence with characterized protein

## 7. Explore the links from this page

Page links include BER alignment files (by clicking on match protein names in the BER skim section), sequence pages (by clicking on the 'View Sequences' button near the top left of the page), TIGRFAM and Pfam links (by clicking on the HMM accession numbers), etc.



Example alignment.

### CDS

```
>nmen_3
GTGGACAAACTGAAATCTCCGCAACGGCCCGCTCAACGGGGAAATAACGGTCTCGGGC
CGCAAAACCGCGCATTGCCGCTGATGTGCGCGGGTTTGCTGACATCGGTACGTTGCCG
CTGAAAAACGTCCTTATGCTGGCAGATGTGAAACACCGCAAAAGCTGCTTCAGGGGATG
GGCGCGCGCTCCTGACCGACAATATCAGCGAATTTGAAATCAACGGCGGTACGGTAAC
AATACCTGCGCCCTTACGAGTTGGTCCGACGATGCGCCCTTCGATTTTGGTGCCTGGGC
CCGACGCTGGCGCGTTTCGGCGAGGCGCAAGTCAGCCTGCCGGCGCGCTGCGCCATCGGT
TCGCGCCCGCTCGATCAGCATTGAAAGGCTTGAAGCGATGGGTGCTGAGATTGTTATC
GAACACGGTTACGTCAAAGCCAAAGGCAAACTCAAAGGTACGCGCGTGGCGATGGATGTC
GTTACCGCTCGCGCGCGCAAGAACTGCTGATGGCGCGCACGCTGGCGGAAGGTACGACG
GTTTTGAAAACTGCGCCATTGAGCCTGAAGTGGTTCGATTGGCGGAATGCTGCTGCTAAA
ATGGCGCGCAAAATCAGCGCGCATCGGTACGTCACCAATGATTGTGAAGGGGTGGACGAG
CTGCAAGGCTGCGAACACAGCGTCGTCGCCGACCGGATCGAGGCGGGGACGCTTCCTGTC
GCGGTGGCGGATAACCGGTGGCAGGGTGGTTTTCGCGGAATCCCGCGCGCAAAACGATGGA
GTGCTGTTGGAACAACTGGTTGAGCGAGGTGCGGTGATTGAGCGCGCGCACGATTGGATC
GCCATCGATATGCGCGCAGCGTCCGAAGGCGGTGGACATCCGCGACGGTTCCTCCACCGCGC
TTCGCCACGGATATGCGCGCGCAGTTTATGGCATTTGAATCCCGTGGCGAGGGGAAGCTGC
CGCGTGGTGAACAGATTTTGAACCGCTTTATGACAGTCCCGAGTTGAACCGGATG
GGGGCGAATACACACCGAGGCGCAATACGCGATTTGTCAGGGTGTGGAACAGCTTTCC
GGCGCATGCTCAAGGCGACGATTTGCGTGGCTCCCGCGCGCTCGTTATCGCGCGTTTG
GCGCGCGAGGCGAAACCTGGTCCGACAGATTACCACTTGGATCGCGGTATGAAAAAT
ATTGAAAAAACTCGCGCAGCTCGCGCGCAAAATCAGCGCGCTATCCGCGCTGA
```

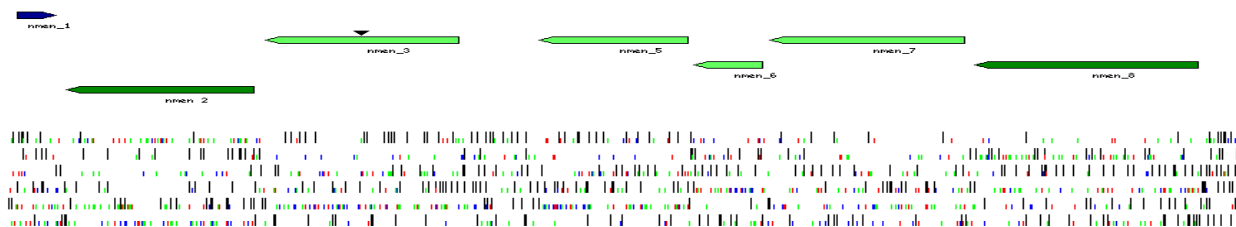
### Protein

```
>nmen_3
VDKLKISANGPLNGEITVSGAKNAALPLMCAGLLTSGTLRLKNVPMADVKTQKLLQGM
GARVLTDNISEFEINGGTVNNTCAPYELVTRMRASILVLGPTLARFGEAQVSLPGGCAIG
SRPVDQHLKLEAMGAEIVIEHGYVKAQKGLKTRVAMDVTVVGTENLLMAATLAEGTT
VLENCAIEPEVVDLAECVLKMGAKISIGTSTHIVEGVDELQCEHSVVPDRIEAGTFLC
AVAITGGRVRLRNAAPKTMEEVVDLKLVEAGAVIEAGDDWIAIDMRQPKAVDIRTVVHPG
FPTDMQAFMALNAVAEGSCRVVETIFENRFMHVPELNRMGANITTEGNTAFVQGVQELS
GAVVKATDLRASALVIAGLAARGETVVEQIYHLDRGYENIEKLGSLVGAKIERVSG
```

Example Sequences

8. You can get back to the Welcome page at any time by clicking 'Home' in the upper right section of all Manatee pages.

9. Let's explore Genome Viewer. Click on the Genome Viewer link at the top of the Gene Curation Page, or go back to the Home page and click on Genome Viewer there.



10. A standing Manatee demo is always available at [manatee.igs.umaryland.edu](http://manatee.igs.umaryland.edu). The database is 'cgspu', the username is 'training1', the password is 'training1'. This user account has read, but not write, privileges to the database.



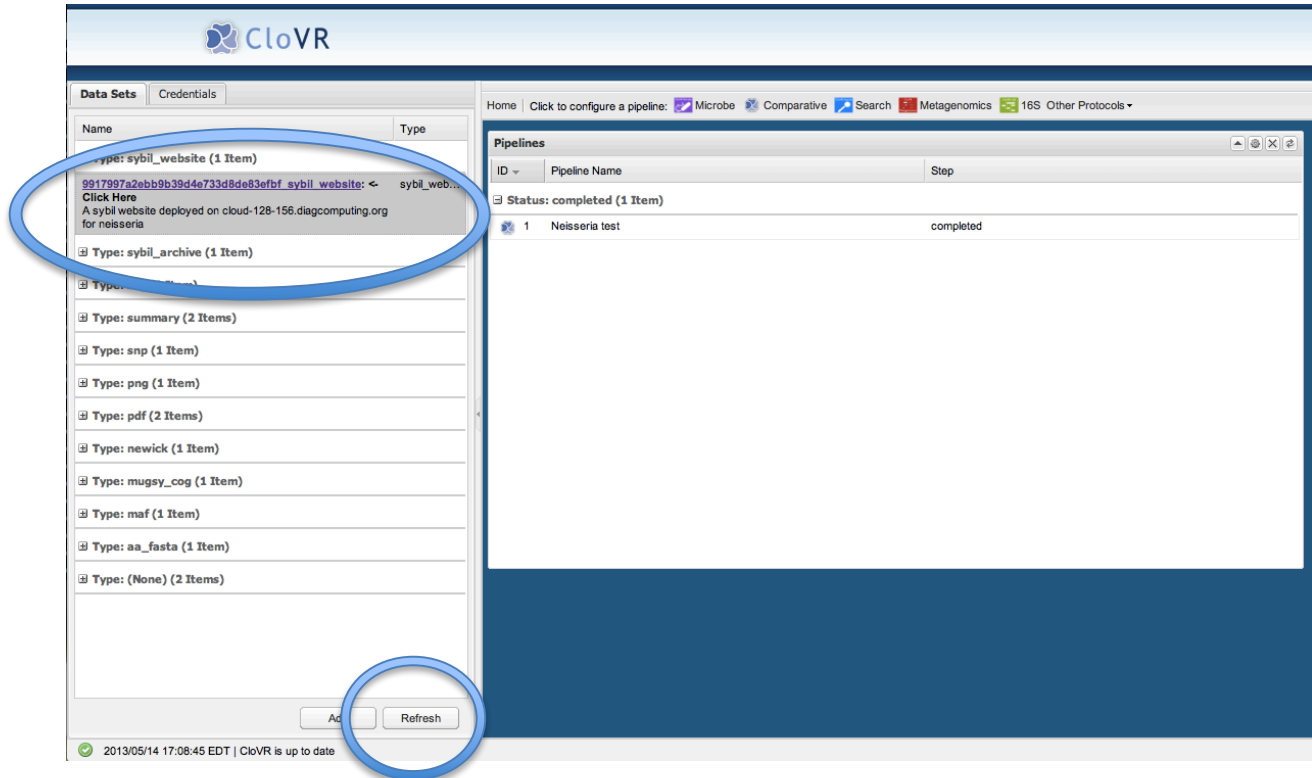
# Sybil activity – ASM 2016 CloVR workshop

## 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.



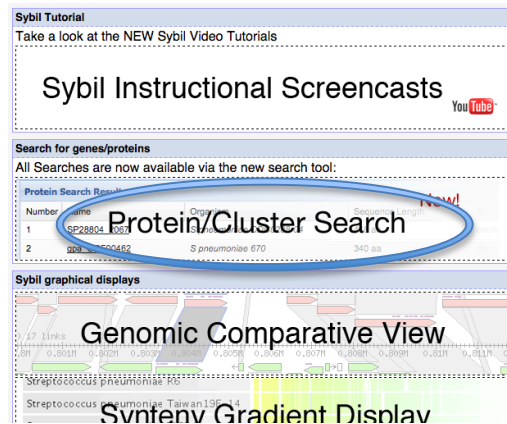
## Sybil home page

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	tbpA3	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

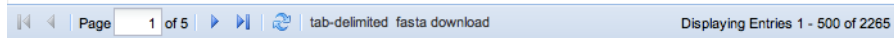
This will bring up genes that contain the word 'protein' in their gene product annotation.


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.



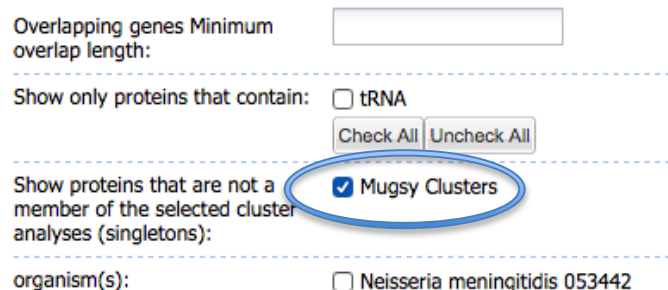
- Hide the search window by clicking the button at the top of the search form.
- Hide columns by clicking the down arrow in the column header and unchecking the box. The arrow will appear when you hover over the header with your mouse. 
- Reorder columns by clicking and dragging the column header.
-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

- 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	<u>MC58:NMB0016</u>	<i>N meningitidis</i> MC58	78 aa	hypothetical protein
139	<u>MC58:NMB0046</u>	<i>N meningitidis</i> MC58	74 aa	hypothetical protein
57	<u>MC58:NMB0067</u>	<i>N meningitidis</i> MC58	225 aa	polysialic acid capsule biosynthesis protein SiaD
156	<u>MC58:NMB0093</u>	<i>N meningitidis</i> MC58	28 aa	hypothetical protein
112	<u>MC58:NMB0099</u>	<i>N meningitidis</i> MC58	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: CLOVRN\_SYBIL: PROTEIN MC58:NMB0016

[DB=CLOVRN\_SYBIL]

PROTEIN PROPERTIES

properties of MC58:NMB0016 ?

property	value
organism	<i>Neisseria meningitidis</i> MC58
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  
MKNSFCGQFACCANGILLFFTEWLPMSLRTGILNRFERKVCLELVEMVRCRHECLISAYRQSAHPNLCIS  
GGKNEICG

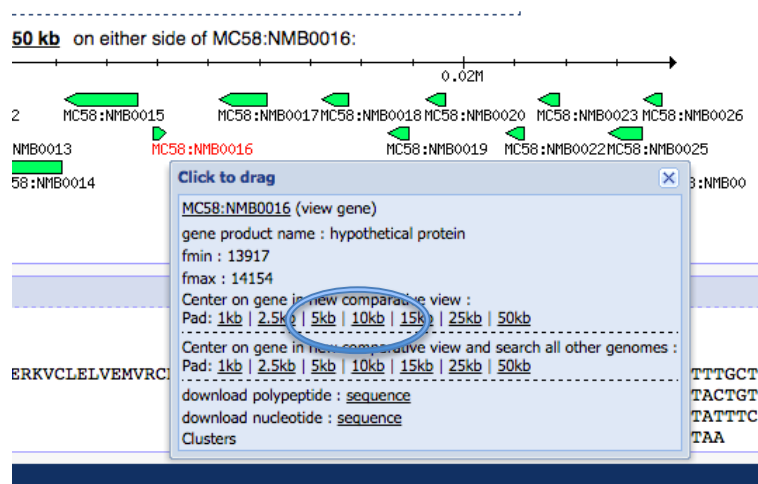
Download Nucleotide FASTA  
>MC58:NMB0016 nucleotide  
ATGAAGAATAGTTTTGTGGACAGTTGCTTGTGTGCAAAATGGCATCCTACTTTTCTTACCGAATGCG  
TGCCGATGTCCTTAAGAACCGGAATACGTGGAGGTTTGAGAGGAAAGTGTGTTGGAACTTGTGGAAT  
GGTCAGGTGTCGACAGAAATGCTTATTTCTGCATATCGGCAGAGTGGCATCCGAATTTGTGTATAGT  
GGTGGAAAAATGAGATTTGCGGGTAA

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'.



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.

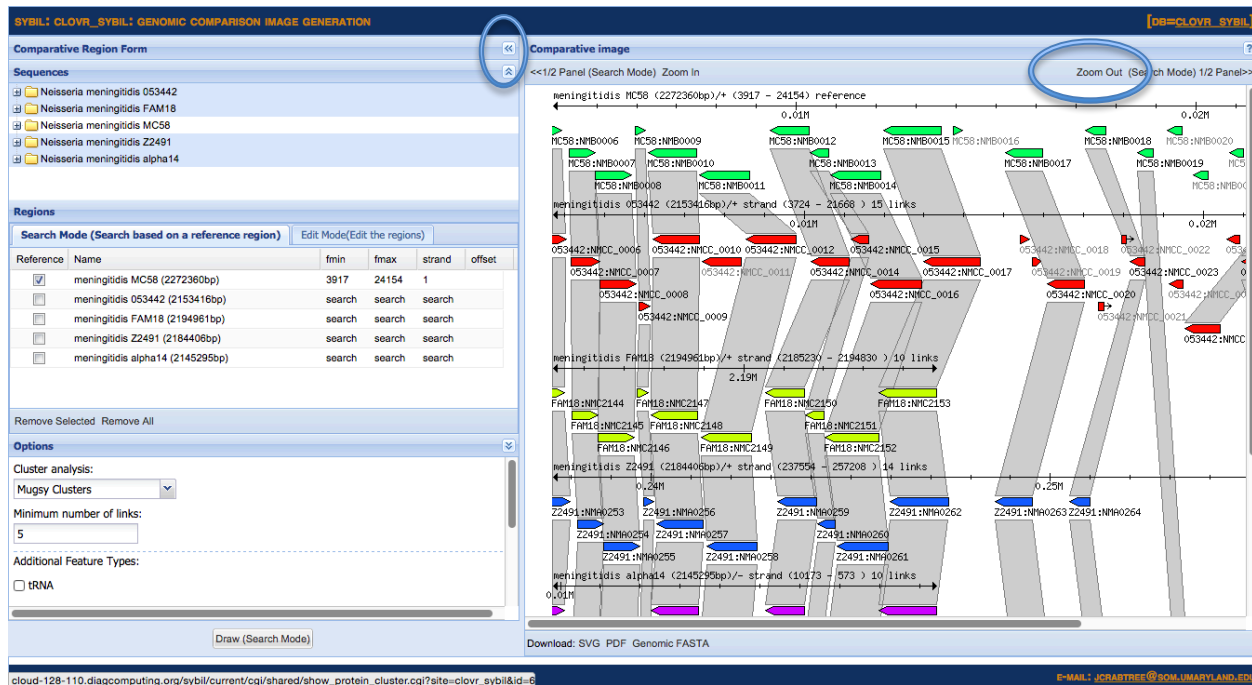
2. Click 'Draw (Search Mode)' at the bottom of the form.

The screenshot shows the 'SYBIL! CLOVR\_SYBIL! GENOMIC COMPARISON IMAGE GENERATION' form. It has two main sections: 'Sequences' and 'Regions'. The 'Sequences' section lists five Neisseria meningitidis sequences: 053442, FAM18, MC58, Z2491, and alpha14. The 'Regions' section has a 'Search Mode (Search based on a reference region)' button and an 'Edit Mode (Edit the regions)' button. Below these is a table with columns: Reference, Name, fmin, fmax, strand, and offset. The table shows the selected reference region for MC58 (2272360bp) and search regions for the other four sequences.

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.

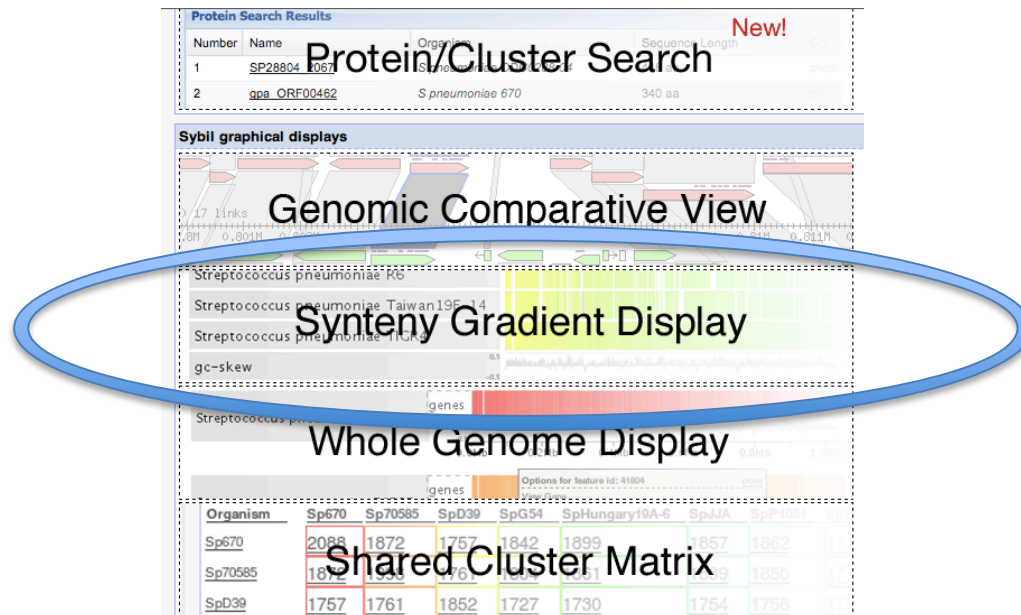
### Genomic comparative view tips

- 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'.

The screenshot shows the 'Sequence Selection' and 'Options' sections of the Sybil interface. The 'Sequence Selection' table lists four Neisseria meningitidis strains. The 'Options' section has several checkboxes, with 'Color those genes with multiple copies (paralogs) in the query black' and 'Show %GC graph for reference(s)' checked. The 'Add All Organisms' button is circled in blue.

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

[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)

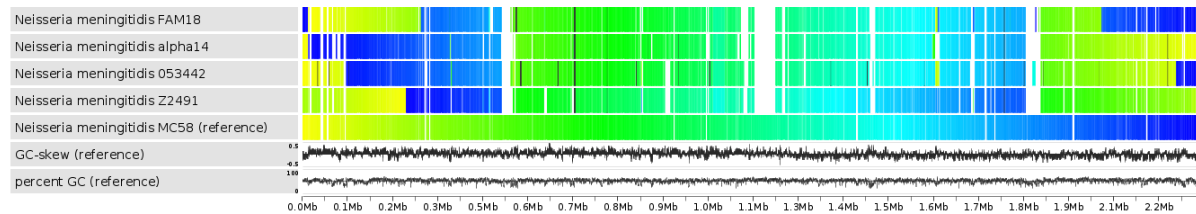
☐ Show GC skew graph for reference(s)

☐ 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'

---

### **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.