

## MetaVir- Analysis of viromes

## Simon Roux

## Abstract

METAVIR is a web server designed to annotate viral metagenomic sequences (raw reads or assembled contigs). A set of published viromes, identified as "public projects", is already available, and your own data sets can be processed in a private environment.

**Citation:** Simon Roux MetaVir- Analysis of viromes. **protocols.io**

[dx.doi.org/10.17504/protocols.io.d5i84d](https://doi.org/10.17504/protocols.io.d5i84d)

**Published:** 14 Jan 2016

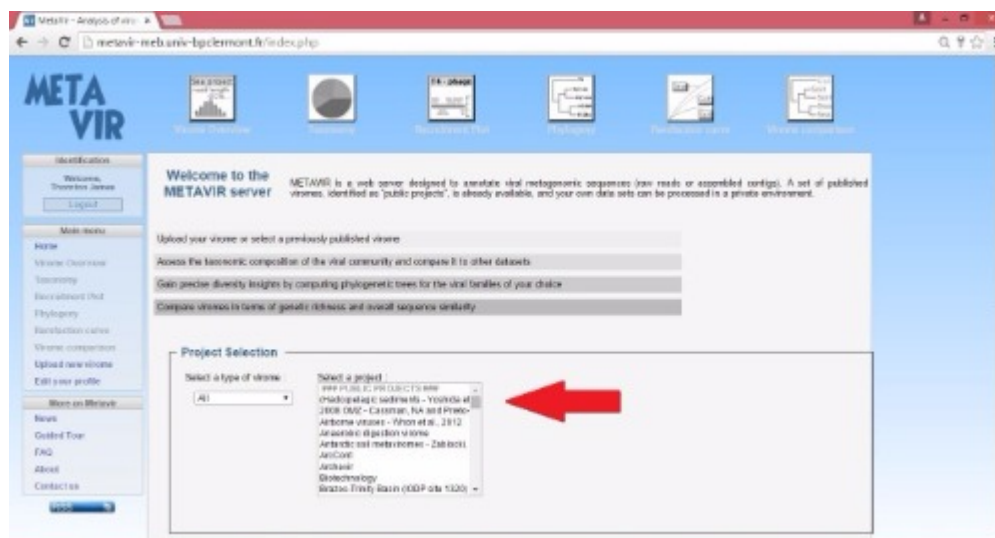
## Guidelines

## Video tutorial

## Protocol

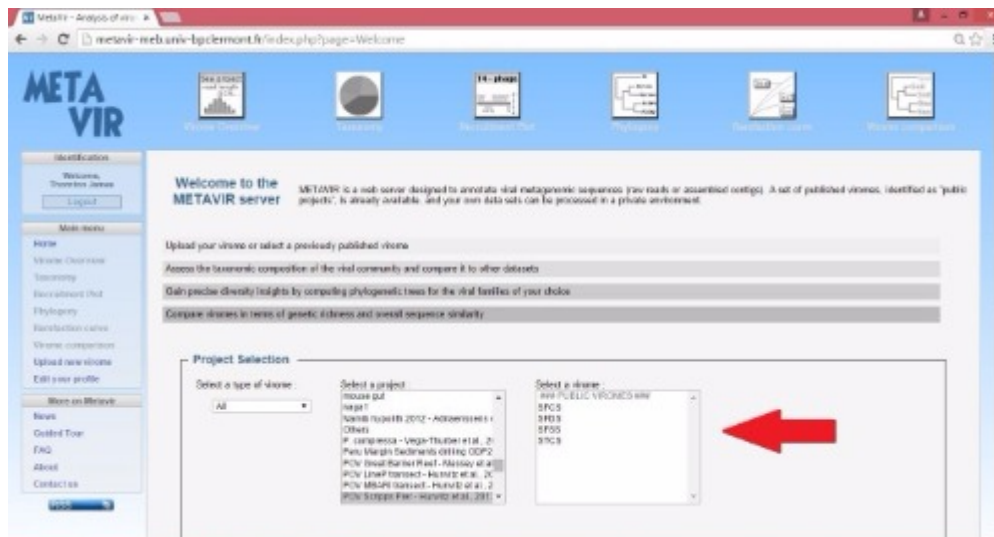
### Step 1.

First select a project.



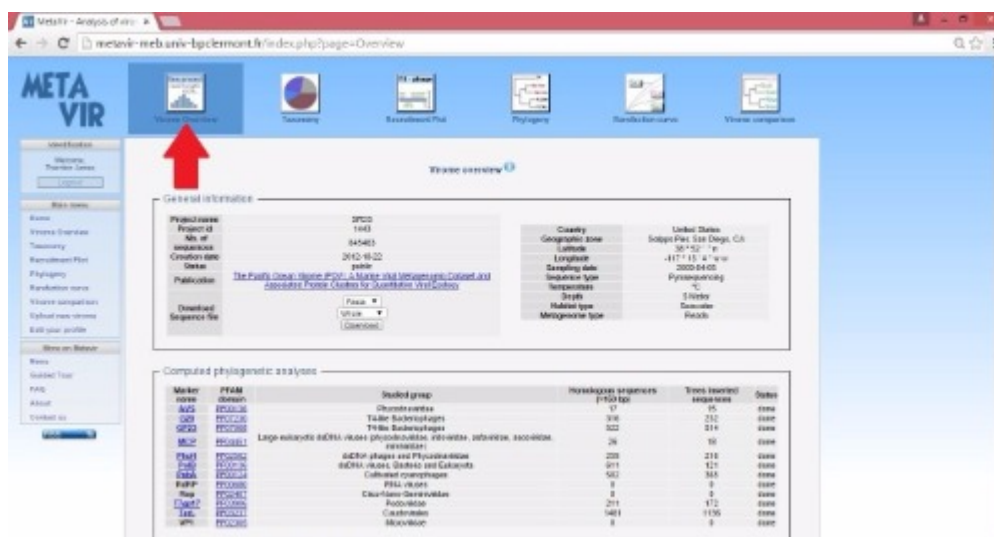
### Step 2.

Then select a virome.



### Step 3.

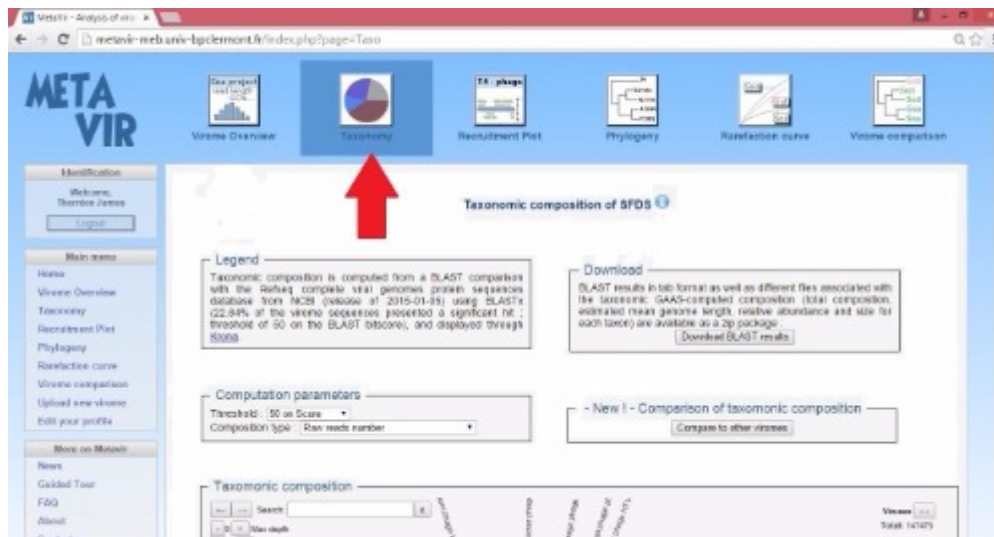
First tab links to the virome overview page.



### Taxonomic composition

### Step 4.

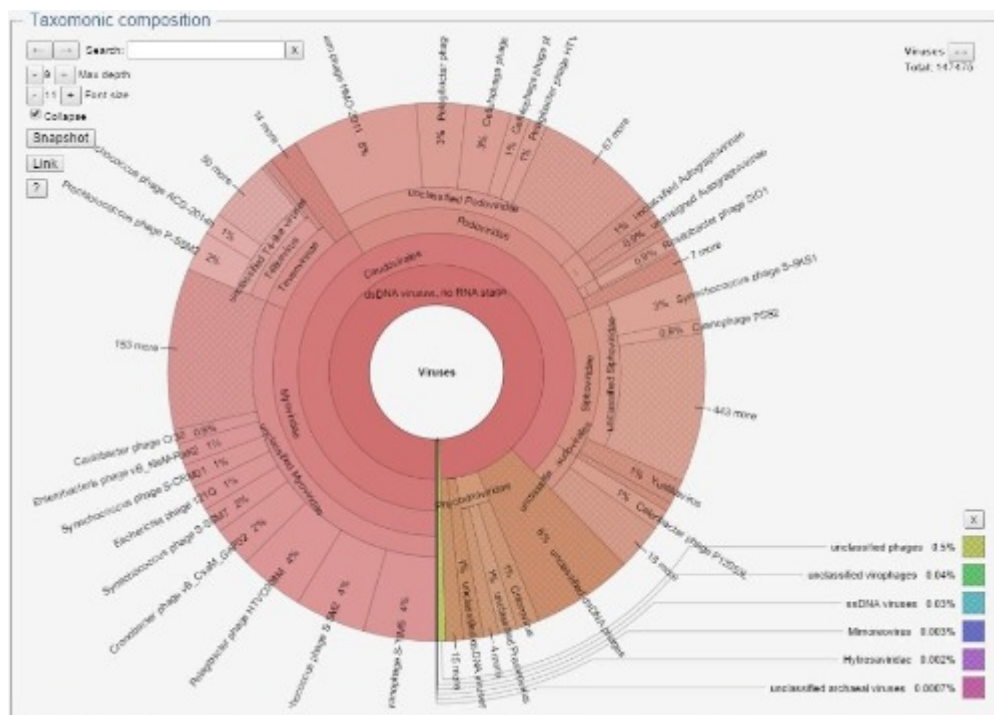
The next tab links to the taxonomic composition page. The different taxonomic compositions computed are all displayed through this page.



## Taxonomic composition

### Step 5.

Krona makes it possible to navigate through the piechart diagram.

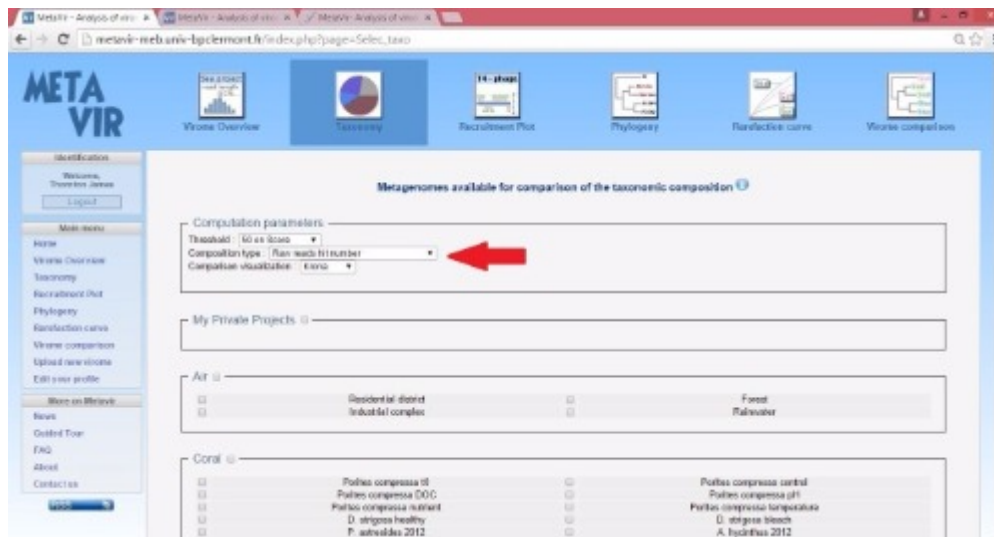


## Taxonomic composition

### Step 6.

All different ratios are displayed on the right.

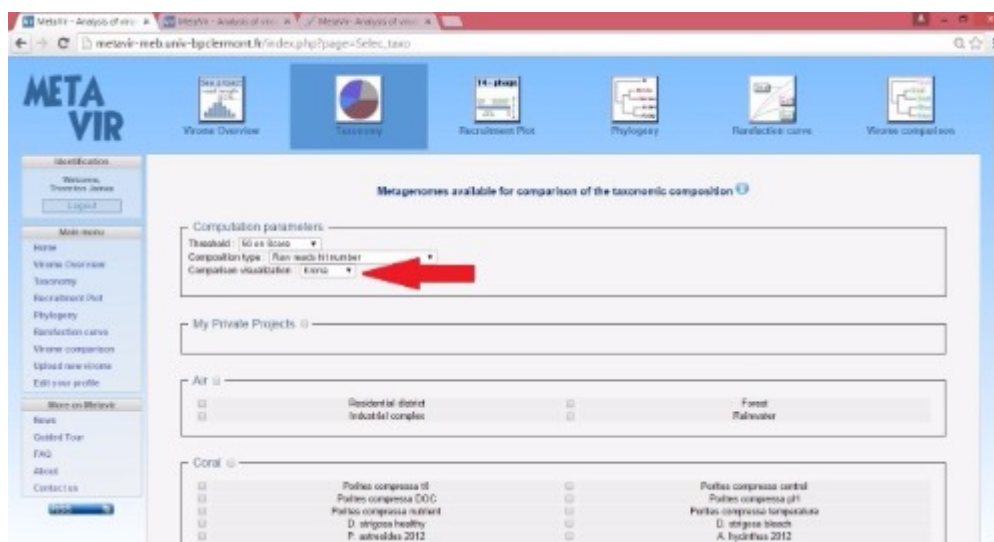




## Taxonomic composition

### Step 10.

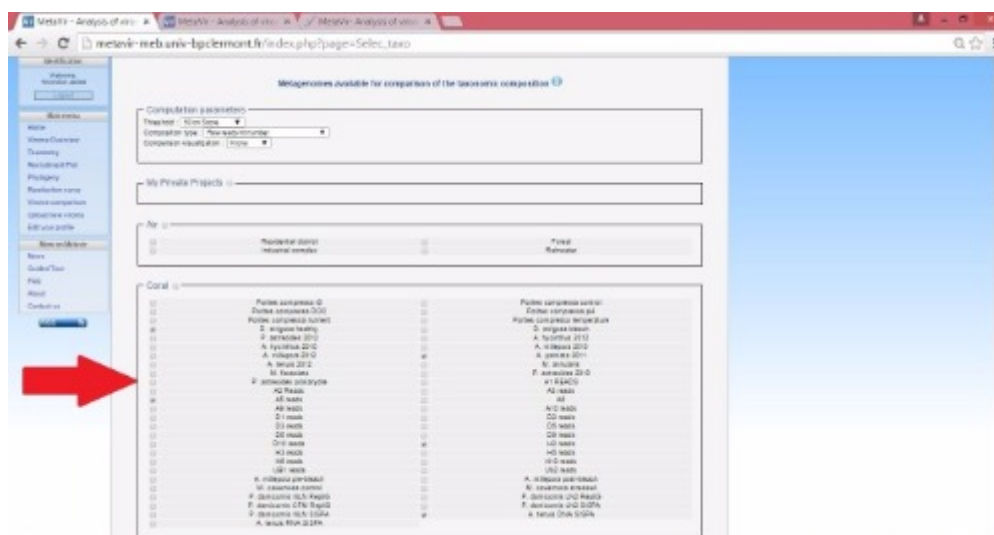
Then choose a type of visualization.



## Taxonomic composition

### Step 11.

Then select all the viromes you want to compare.

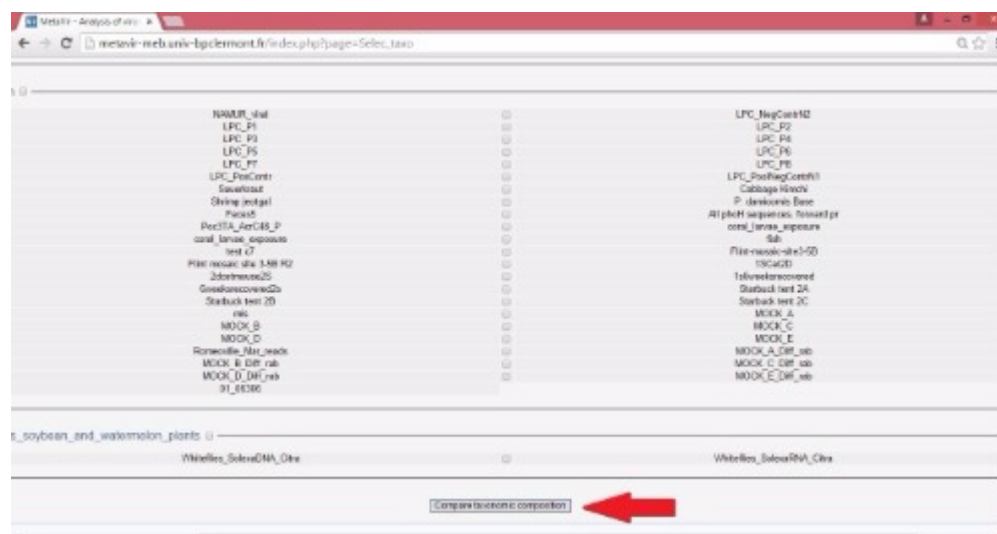




## Taxonomic composition

### Step 12.

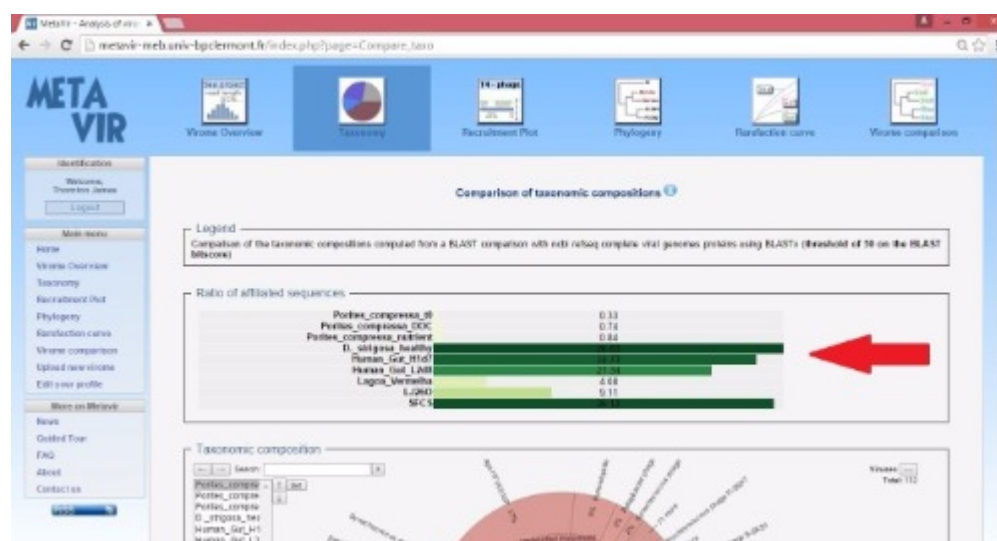
Click compare



## Taxonomic composition

### Step 13.

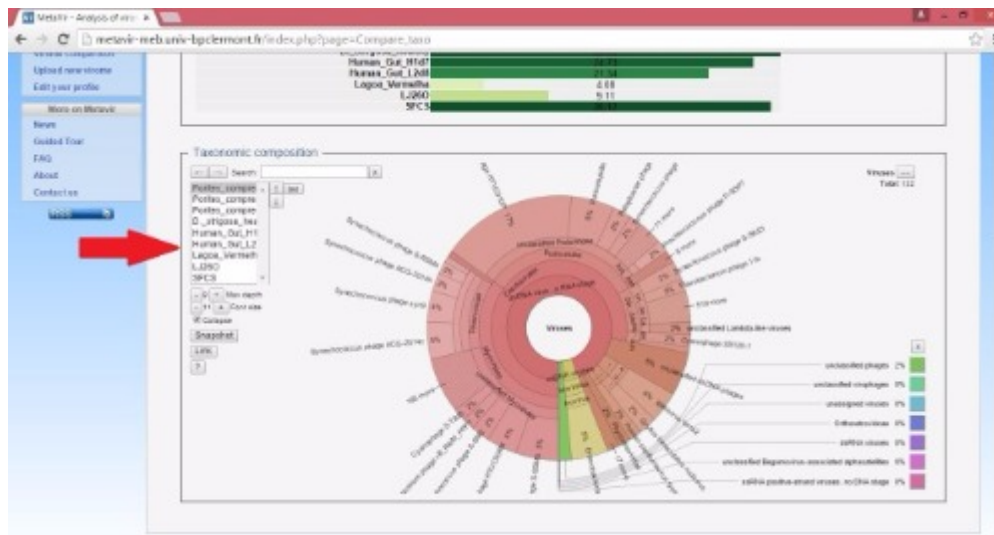
Affiliation rates are reminded here.



## Taxonomic composition

### Step 14.

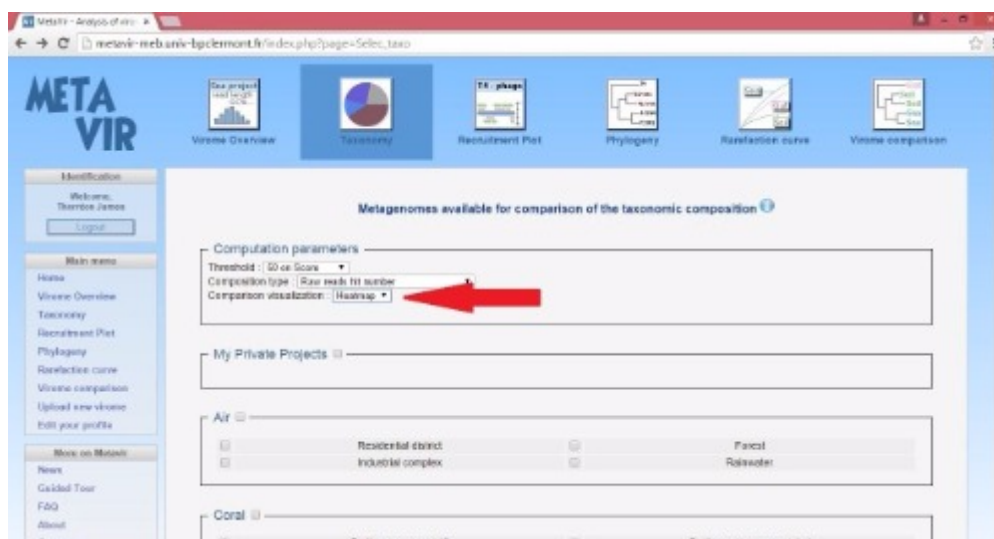
Several compositions can be displayed on the same interactive Krona chart.



## Taxonomic composition

### Step 15.

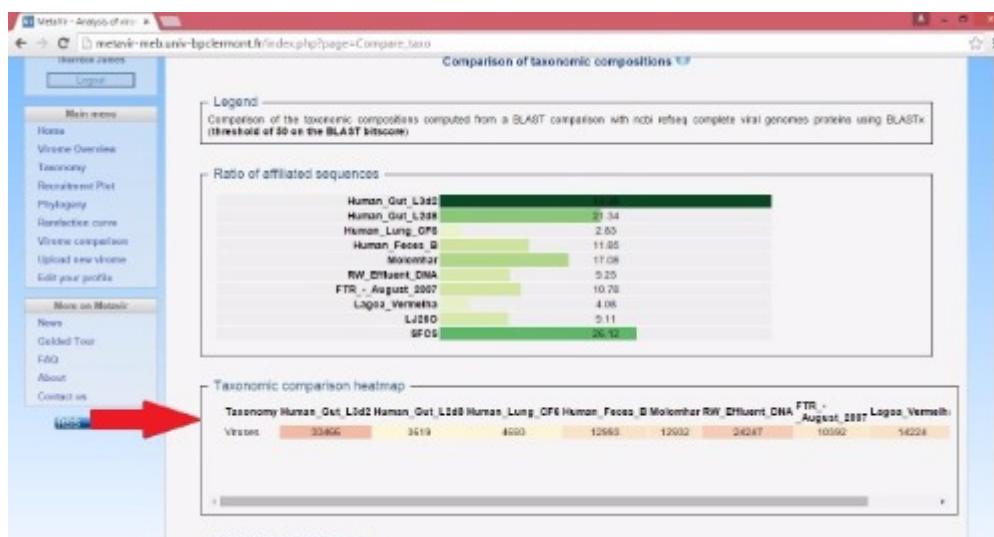
The same comparison can be displayed through a heatmap



## Taxonomic composition

### Step 16.

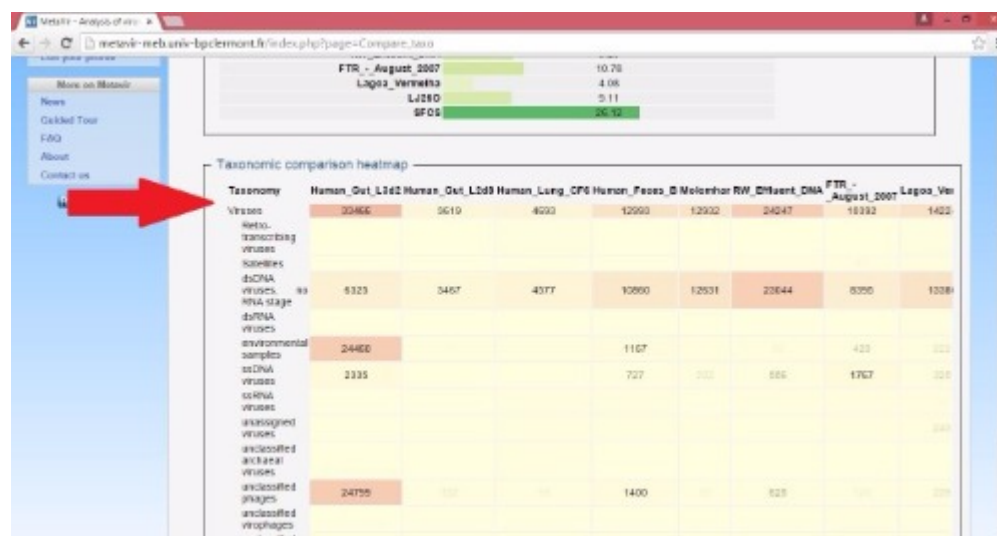
This is an interactive heatmap.



## Taxonomic composition

### Step 17.

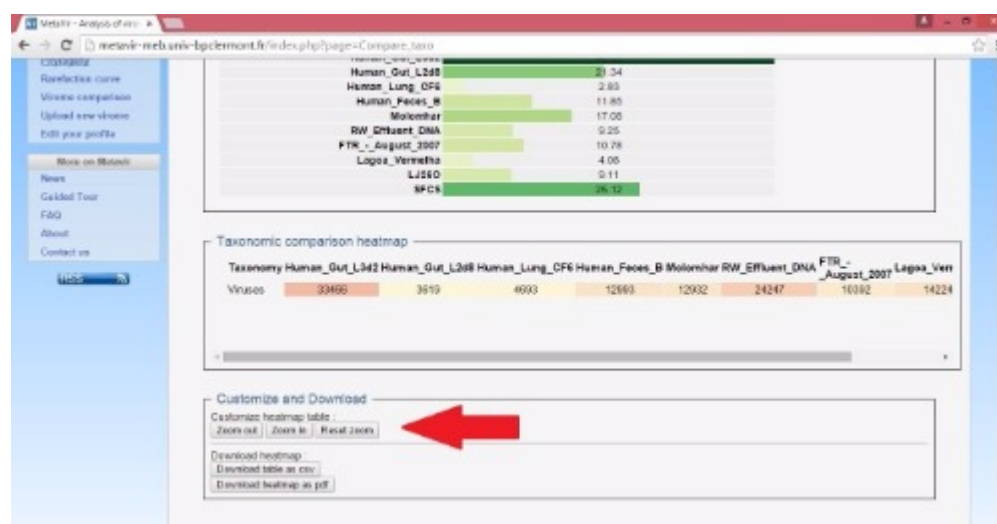
Fold and unfold taxonomic groups with a mouse click.



## Taxonomic composition

### Step 18.

Font size can be modified to fit your screen.

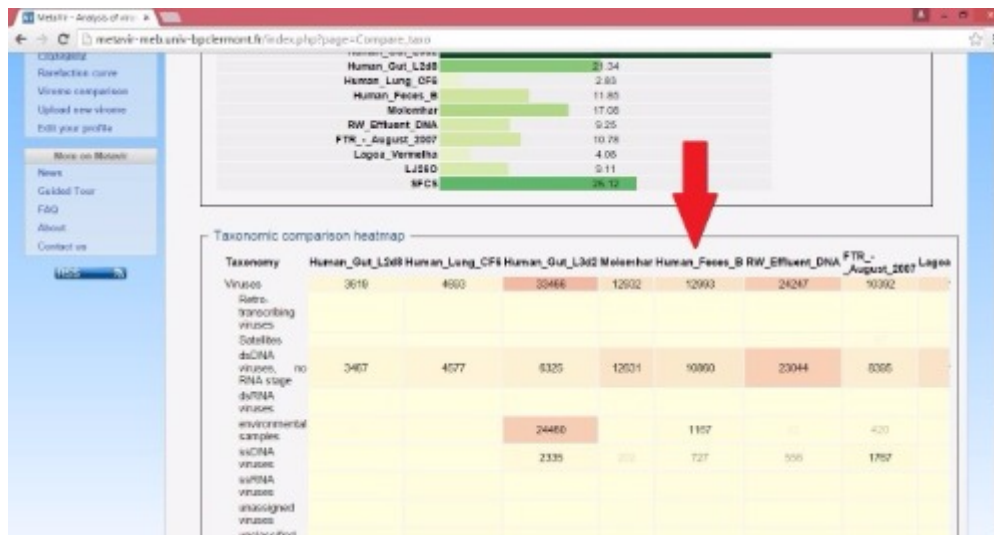


## Taxonomic composition

### Step 19.

Drag and drop columns to modify their order.

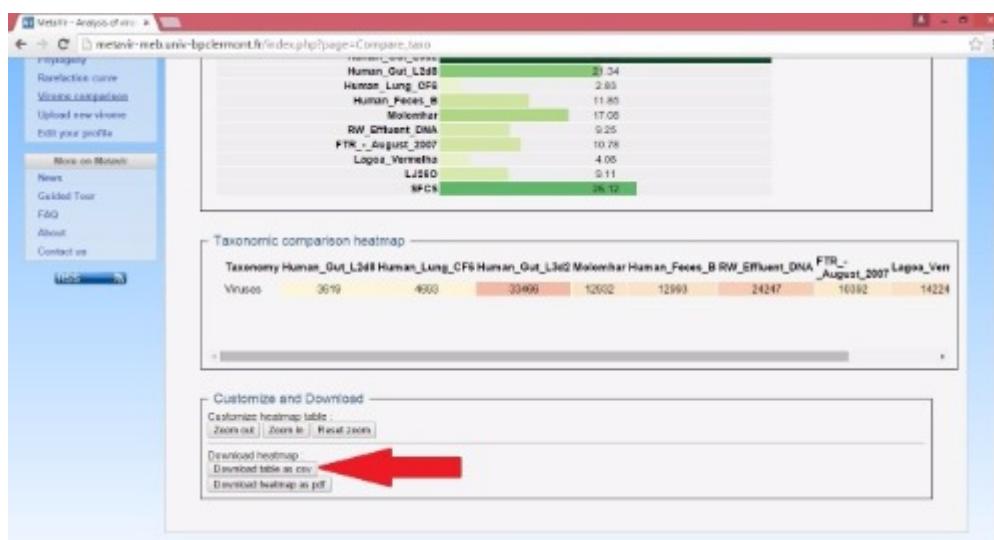




## Taxonomic composition

### Step 20.

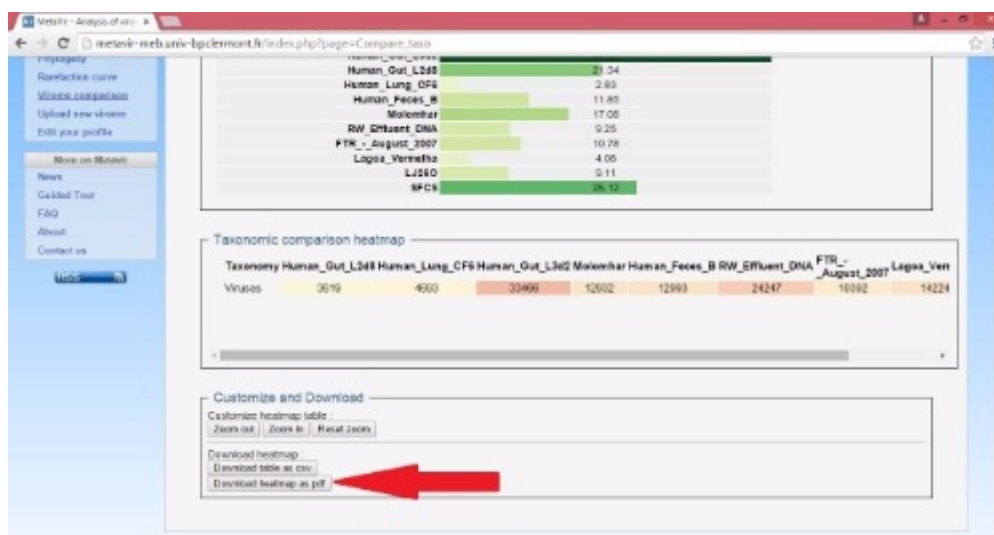
The heatmap can be downloaded as a csv file.



## Taxonomic composition

### Step 21.

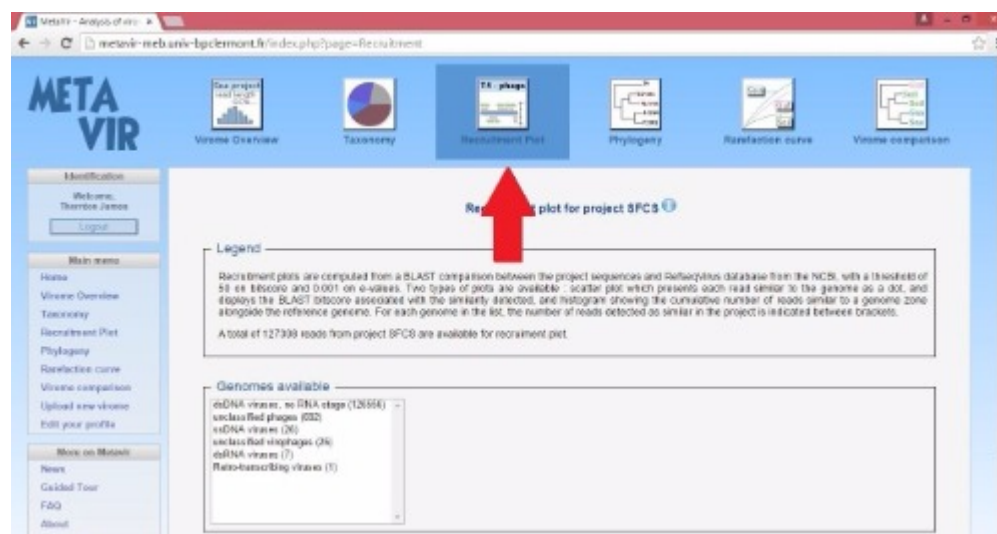
The heatmap can also be downloaded as a pdf file.



## Recruitment plots

### Step 22.

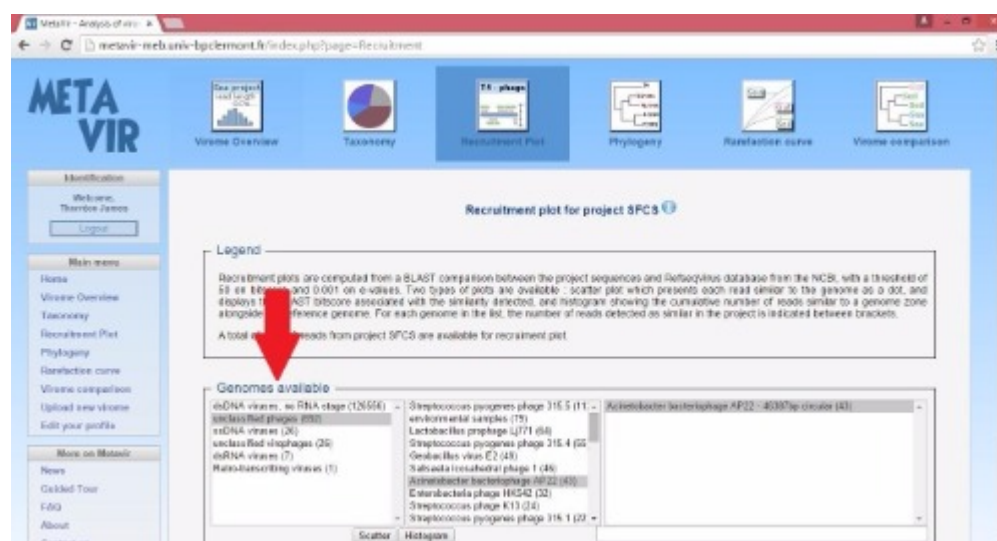
This button links to the recruitment plot.



## Recruitment plots

### Step 23.

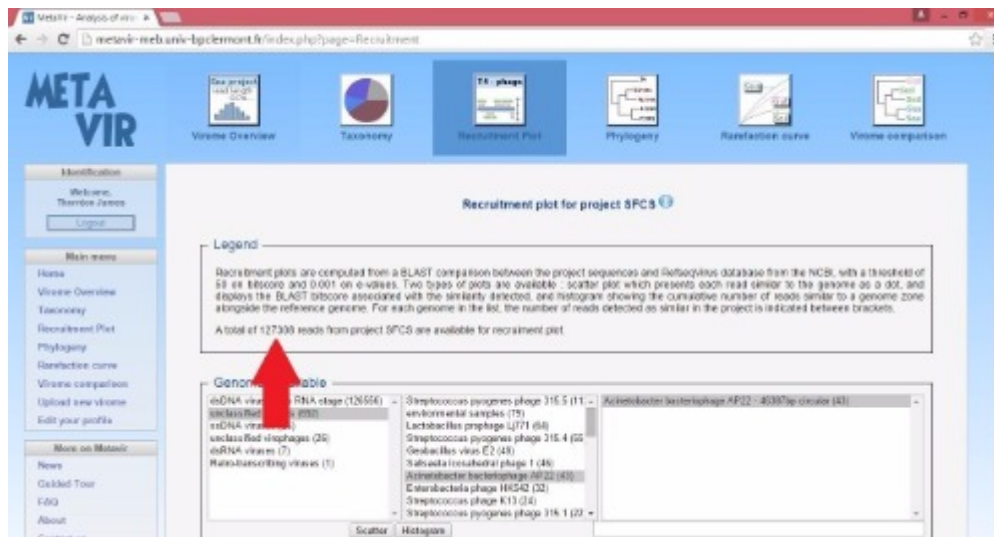
Available genomes are classified by taxonomic group.



## Recruitment plots

### Step 24.

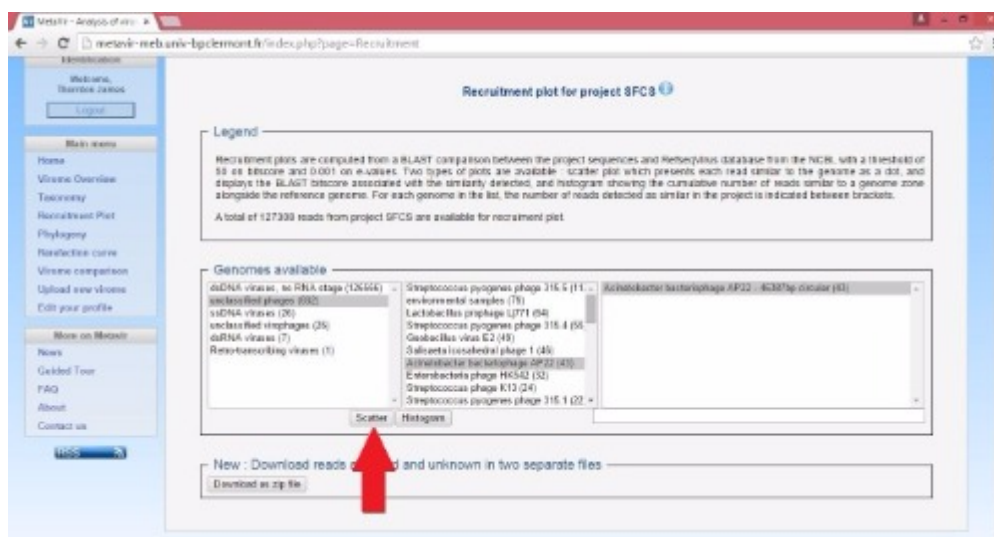
Total number of reads available is indicated here and the number of reads for each group/genome is displayed here.



## Recruitment plots

### Step 25.

On this first type of recruitment plot each read is displayed as a dot (scatter plot).



## Recruitment plots

### Step 26.

A genome map is displayed at the bottom with genes name when possible.



## Recruitment plots

### Step 27.

This is a list of potential list of viromes to add to the recruitment plot, the number of reads is between brackets.



Projects available for this genome (number of reads recruited)

<input type="checkbox"/> P. denitrona S100 (1)	<input type="checkbox"/> Shrimp (ectoparasite) (22)
<input type="checkbox"/> Colicoidae (16)	<input type="checkbox"/> Symbiote (3)
<input type="checkbox"/> NIMAR_vial (1)	<input type="checkbox"/> Puccinia_Grass_P (11)
<input type="checkbox"/> Boreomite_Mar_reads (5)	<input type="checkbox"/> D. strigosus healthy (1)
<input type="checkbox"/> C1 reads (1)	<input type="checkbox"/> M. canescens control (1)
<input type="checkbox"/> P. denitrona 142-555PA (262)	<input type="checkbox"/> P. denitrona HJH-555PA (184)
<input type="checkbox"/> P. denitrona HJH-555PA (2)	<input type="checkbox"/> P. denitrona CFM-555PA (17)
<input type="checkbox"/> A. tenuis DNA 555PA (21)	<input type="checkbox"/> Atlantic - Extra (1)
<input type="checkbox"/> Arctic - Vial (2)	<input type="checkbox"/> Ica-Opocnessa Trench (2)
<input type="checkbox"/> Mosquito DNA - 1 (5)	<input type="checkbox"/> Human_Gut_Diet_control (1)
<input type="checkbox"/> Human_Gut_L141 (1)	<input type="checkbox"/> Human_Gut_L141 (82)
<input type="checkbox"/> Human_Gut_H141 (1)	<input type="checkbox"/> Human_Gut_L141 (16)
<input type="checkbox"/> Human_Gut_L141 (34)	<input type="checkbox"/> Human_Gut_L141 (42)
<input type="checkbox"/> Human_Gut_L141 (38)	<input type="checkbox"/> Human_Gut_L141 (1)
<input type="checkbox"/> Human_Gut_H141 (4)	<input type="checkbox"/> Human_Gut_H141 (3)
<input type="checkbox"/> Human_Gut_H141 (4)	<input type="checkbox"/> Human_Gut_L141 (2)
<input type="checkbox"/> Human_Gut_L141 (1)	<input type="checkbox"/> Human_Gut_H141 (18)
<input type="checkbox"/> Human_Gut_H141 (18)	<input type="checkbox"/> Human_Gut_H141 (1)
<input type="checkbox"/> Human_Lung_C11 (12)	<input type="checkbox"/> Human_Feces (2)
<input type="checkbox"/> Human_Feces A (2)	<input type="checkbox"/> Human_Feces (2)
<input type="checkbox"/> Human_Feces C (74)	<input type="checkbox"/> Human_Feces (2)
<input type="checkbox"/> Human_Feces E (5)	<input type="checkbox"/> Human_salivary_Sub1_D1 (2)
<input type="checkbox"/> Human_salivary_Sub1_D1 (1)	<input type="checkbox"/> Human_salivary_Sub1_D1 (2)
<input type="checkbox"/> Human_salivary_Sub1_D1 (2)	<input type="checkbox"/> Human_salivary_Sub1_D1 (1821)
<input type="checkbox"/> Antarctic Lake Summer (1)	<input type="checkbox"/> Lake Bourget (586)
<input type="checkbox"/> Antarctic Lake Spring (1)	<input type="checkbox"/> Tlapala Channel - 0406 (1)
<input type="checkbox"/> Tlapala Channel - 0406 (4)	<input type="checkbox"/> Tlapala Channel - 0406 (1)
<input type="checkbox"/> Tlapala Channel - 0406 (2)	<input type="checkbox"/> Prebait pond - 0406 (1)
<input type="checkbox"/> E. coli (2)	<input type="checkbox"/> Handson (22)
<input type="checkbox"/> Lake Mendocino PNA June (1)	<input type="checkbox"/> Molokai (8)
<input type="checkbox"/> RVI Effluent DNA (71)	<input type="checkbox"/> Lake Mendocino PNA November (1)
<input type="checkbox"/> RVI Effluent DNA (2)	<input type="checkbox"/> RVI Effluent DNA (2)
<input type="checkbox"/> RVI Effluent DNA (24)	<input type="checkbox"/> RVI Effluent DNA (2)
<input type="checkbox"/> RVI Effluent DNA (2)	<input type="checkbox"/> Poulaka DNA (1123)
<input type="checkbox"/> PTR - January 2006 (18)	<input type="checkbox"/> Poulaka Mouth (13)
<input type="checkbox"/> Cern Cell Mouth (5)	<input type="checkbox"/> Metabala Open (14)

## Recruitment plots

### Step 28.

Each virome is identified by a specific color.



## Recruitment plots

### Step 29.

Plots of type histogram display the number of hits for 500-bp wide regions along the genome.



## Recruitment plots

### Step 30.

Plots can be downloaded in different file formats.

## Recruitment plots

### Step 31.

A table of number of hits for each gene is also available.



## Recruitment plots

### Step 32.

Finally, sequences of reads associated with the recruitment plot can be downloaded in fasta format.

**Legend**  
On this histogram, the genome is divided in 500 bp-long parts, and the number of reads with a similarity against each part is represented. If several viruses were chosen, each is associated with a color, and the count are cumulated. The genomic map is indicated on the x-axis with each gene colored according to the legend on the left, and the number of reads on the y-axis.  
[\[Back to genome list\]](#)

**Genome NC\_017094**

Name	Accession	Length (bp)	Taxonomy	Type
Adenovirus bacteriophage AP22	AP22	38,387	unclassified phages - Adenovirus bacteriophage AP22	circular

**Downloads**

Matrix file as csv

Gene table as csv

Reads as fasta file (in zip package)

Chart as pdf

**Projects available for this genome (number of reads recruited)**

<input type="checkbox"/>	P. damicornis Bore (1)	<input type="checkbox"/>	Shrimp (Jedgal) (55)
<input type="checkbox"/>	Cabbage Kirsch (64)	<input type="checkbox"/>	Sauerkraut (53)
<input type="checkbox"/>	NOAMR - virus (1)	<input type="checkbox"/>	Pocilla_Anc48_P (11)
<input type="checkbox"/>	Roseville - May, week (5)	<input type="checkbox"/>	D. singora healthy (1)
<input type="checkbox"/>	Q1 reads (1)	<input type="checkbox"/>	M. lowensis control (1)
<input type="checkbox"/>	P. damicornis LN2 SPSA (242)	<input type="checkbox"/>	P. damicornis LN2 SPSA (184)
<input type="checkbox"/>	P. damicornis NLN RepAG (2)	<input type="checkbox"/>	P. damicornis CFM RepAG (17)
<input type="checkbox"/>	A. trawls GNA SPSA (21)	<input type="checkbox"/>	Alfalfa - Eche (1)
<input type="checkbox"/>	Arctic - Virus (2)	<input type="checkbox"/>	Uta-Capensis Trench (2)
<input type="checkbox"/>	Mosquito GNA - 1 (5)	<input type="checkbox"/>	Human_Gut_Elav_control (1)
<input type="checkbox"/>	Human_Gut_L141 (1)	<input type="checkbox"/>	Human_Gut_L148 (83)
<input type="checkbox"/>	Human_Gut_H141 (1)	<input type="checkbox"/>	Human_Gut_L142 (15)
<input type="checkbox"/>	Human_Gut_L143 (34)	<input type="checkbox"/>	Human_Gut_L145 (42)
<input type="checkbox"/>	Human_Gut_L141 (24)	<input type="checkbox"/>	Human_Gut_L147 (1)
<input type="checkbox"/>	Human_Gut_H142 (4)	<input type="checkbox"/>	Human_Gut_H147 (2)
<input type="checkbox"/>	Human_Gut_H148 (4)	<input type="checkbox"/>	Human_Gut_L148 (2)
<input type="checkbox"/>	Human_Gut_L147 (1)	<input type="checkbox"/>	Human_Gut_H141 (18)

## Recruitment plots

### Step 33.

Generating the file package can take up to a few minutes.

## Phylogenetic trees

### Step 34.

This tab links to the phylogeny page.

**Phylogenetic tree computation from SPCs**

Marker	PRM	Studied group	Phylogenetic sequences (nt)	Traces inserted sequences	Analysis state
AP22	AP22	Phycodnaviridae	7	7	done
Q1	Q1	T4-like bacteriophages	248	175	done
Q2	Q2	T4-like bacteriophages	173	98	done
Q3	Q3	Large eukaryotic dsDNA viruses (phycodnaviridae, herpesviridae, astroviridae, adenoviridae, mimivirus)	2	2	done
Q4	Q4	dsDNA viruses, Herpes and Eukaryota	251	173	done
Q5	Q5	dsDNA viruses, Herpes and Eukaryota	438	212	done
Q6	Q6	Adenoviridae: Subgroup, various Coxsackieviruses and Anguillaviridae	0	0	done
Q7	Q7	Cultivated eukaryotic viruses	279	249	done
Q8	Q8	RNA viruses	0	0	done
Q9	Q9	Circoviridae: Circoviridae	0	0	done
Q10	Q10	Parvoviridae	175	16	done
Q11	Q11	Caulimoviridae	325	168	done
Q12	Q12	Mimoviridae	0	0	done

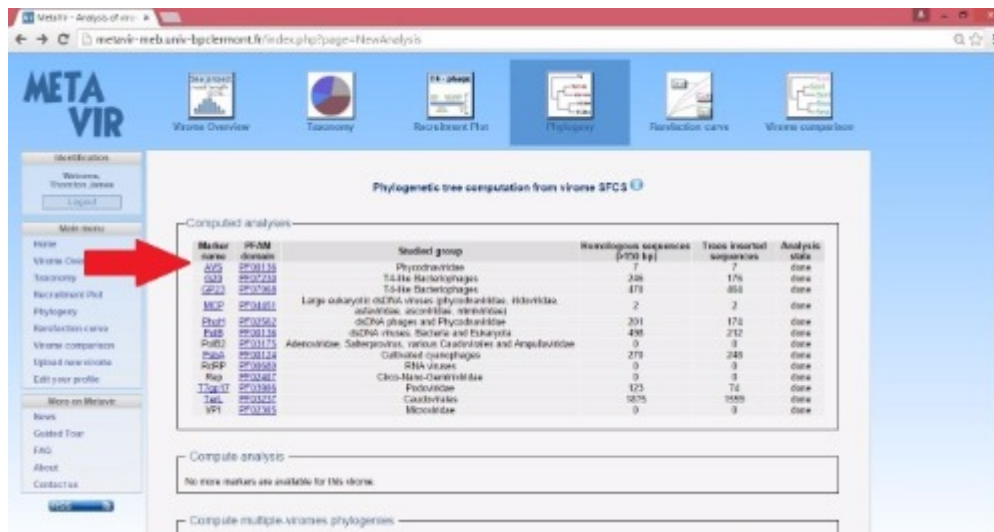
**Compute analysis**  
No more markers are available for this genome.

**Compute multiple viruses phylogenies**

## Phylogenetic trees

### Step 35.

Different marker genes are available.



## Phylogenetic trees

### Step 36.

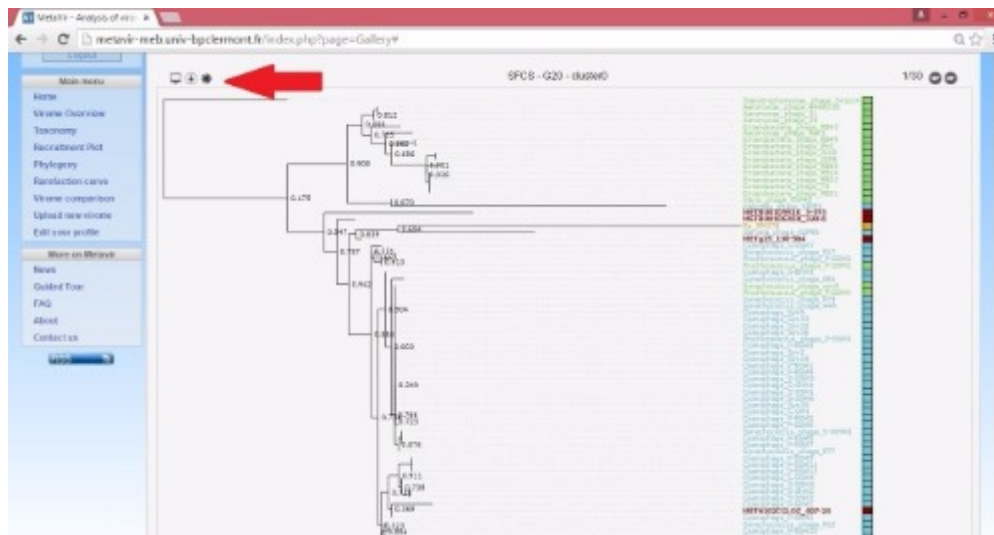
For each marker and each virome, a tree gallery is displayed. Leaves are colored according to the taxonomy, except for metagenomic sequences highlighted in red and bold.



## Phylogenetic trees

### Step 37.

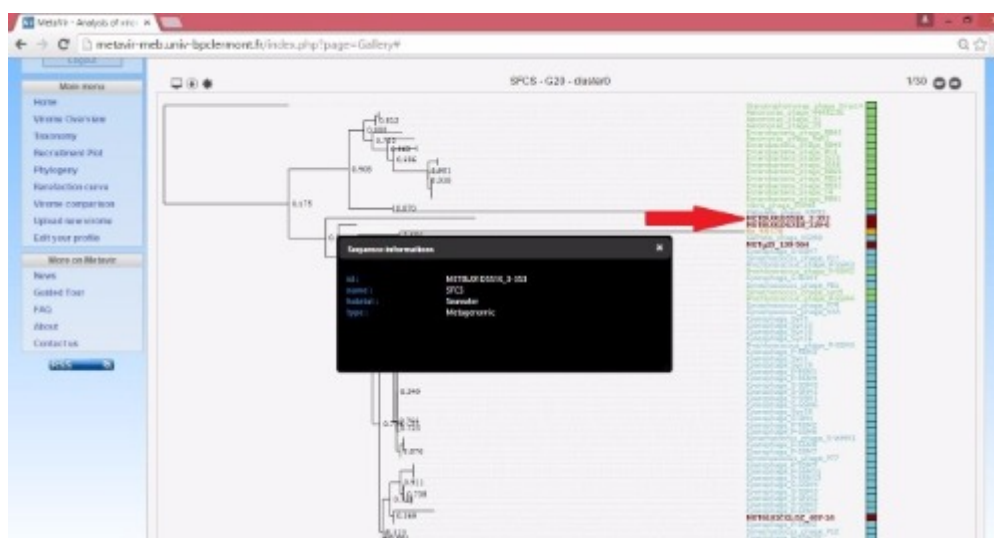
Trees can be displayed in circular or linear form.



## Phylogenetic trees

### Step 38.

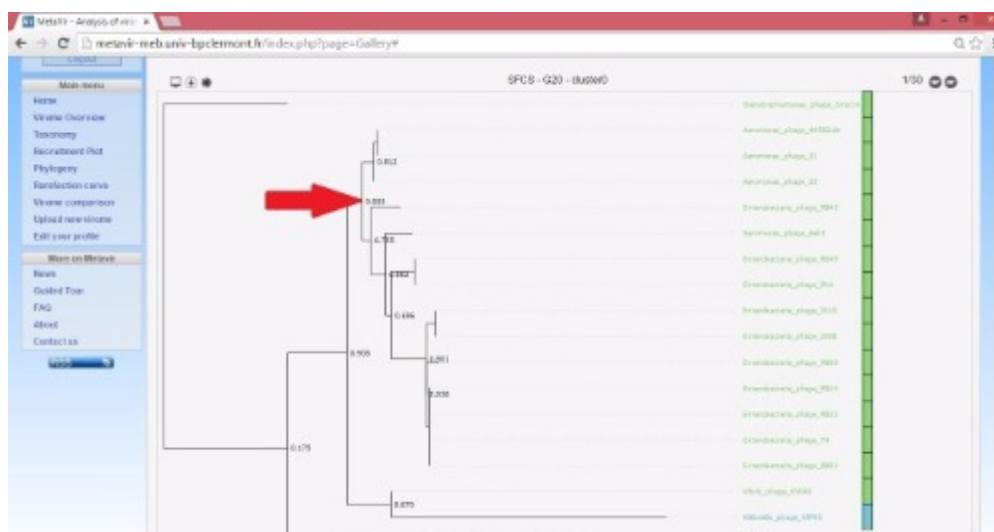
Information is displayed on each leaf when clicked.



## Phylogenetic trees

### Step 39.

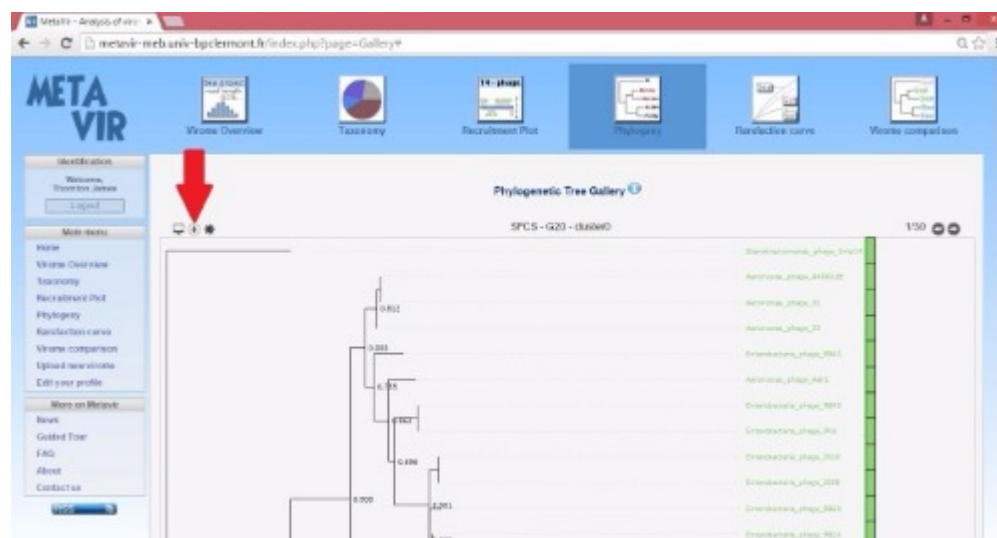
Internal nodes can be gathered and unfolded with a mouse click.



## Phylogenetic trees

### Step 40.

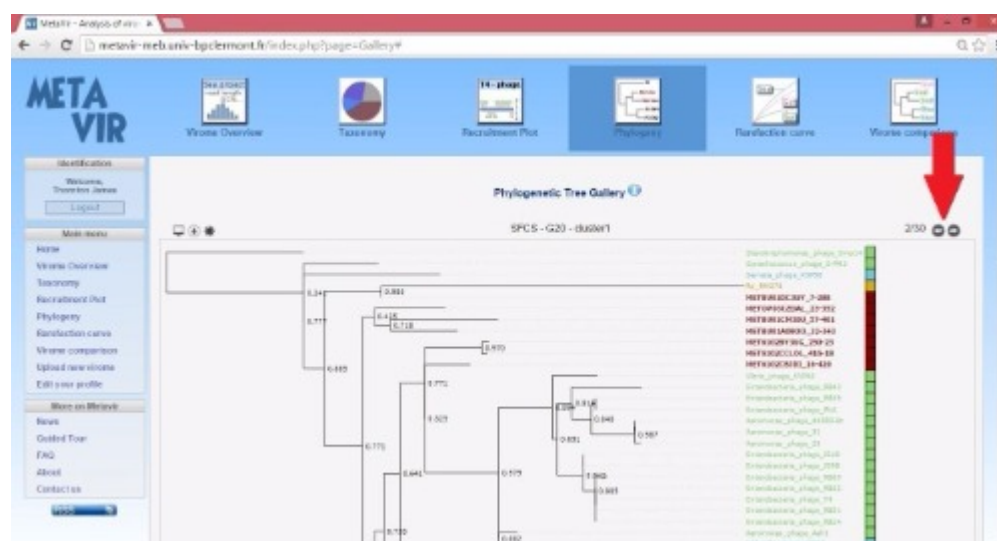
Picture tree can be downloaded as a svg file.



## Phylogenetic trees

### Step 41.

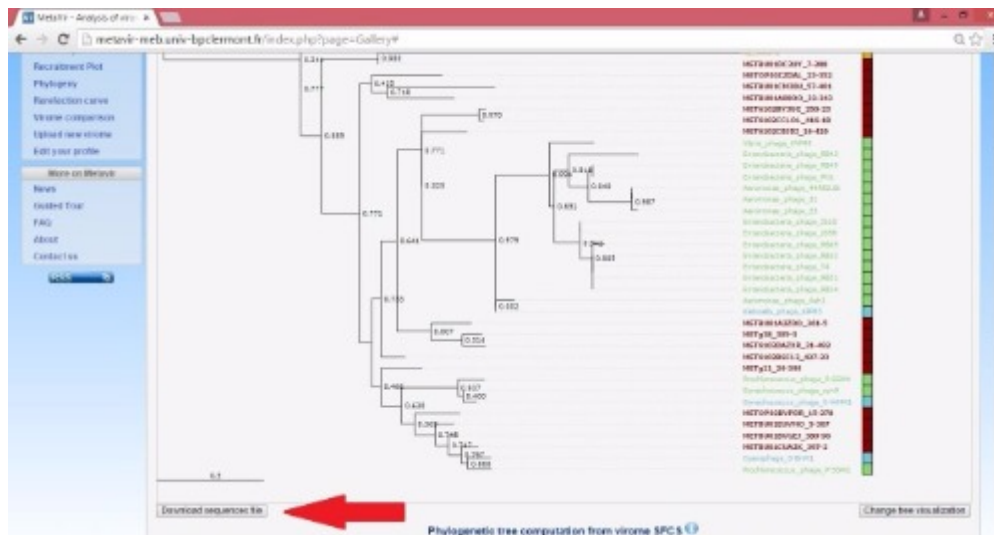
The tree gallery is browsed through these two buttons.



## Phylogenetic trees

### Step 42.

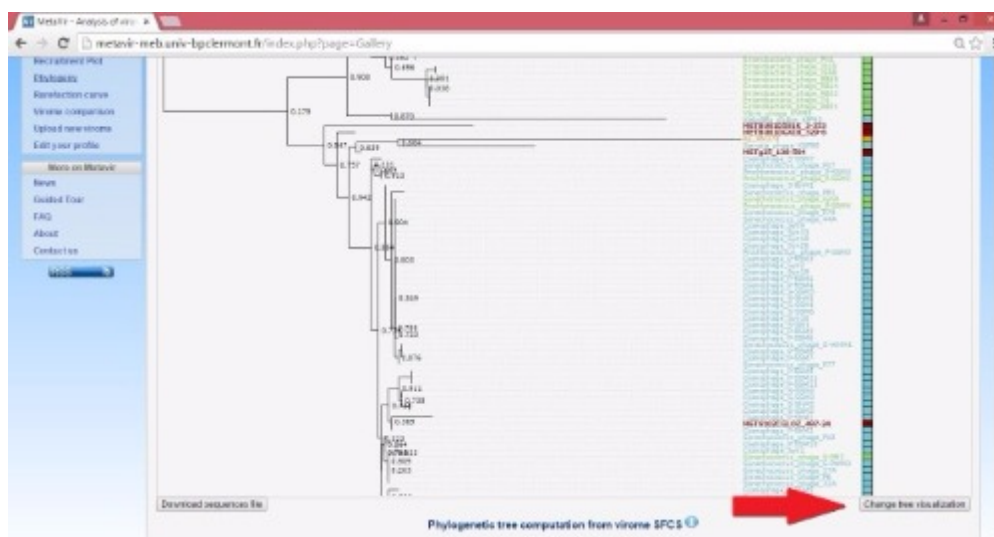
A package containing all sequences used in the trees as well as the alignments and trees as text and picture is available to download.



## Phylogenetic trees

### Step 43.

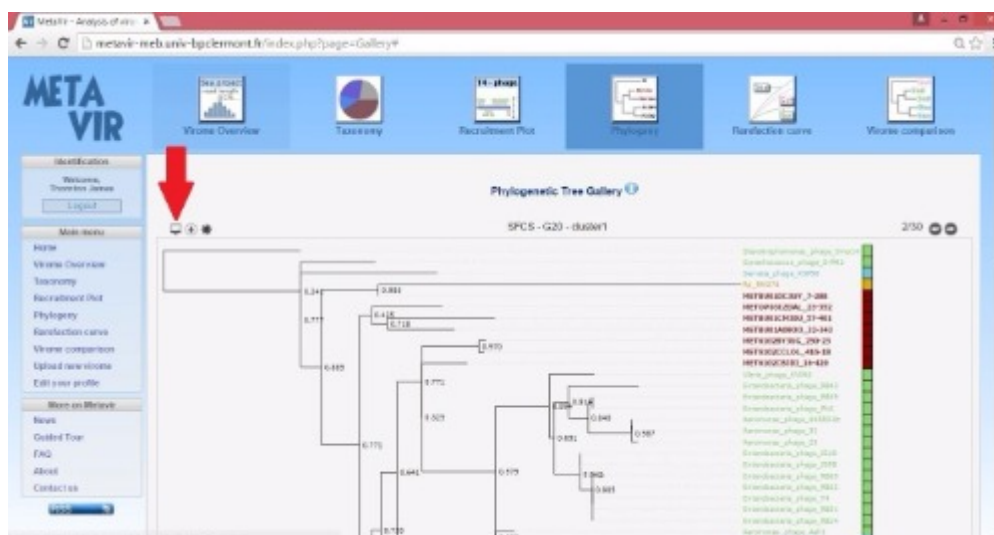
A static tree gallery is also available.



## Phylogenetic trees

### Step 44.

Trees can be displayed on fullscreen.





## Phylogenetic trees

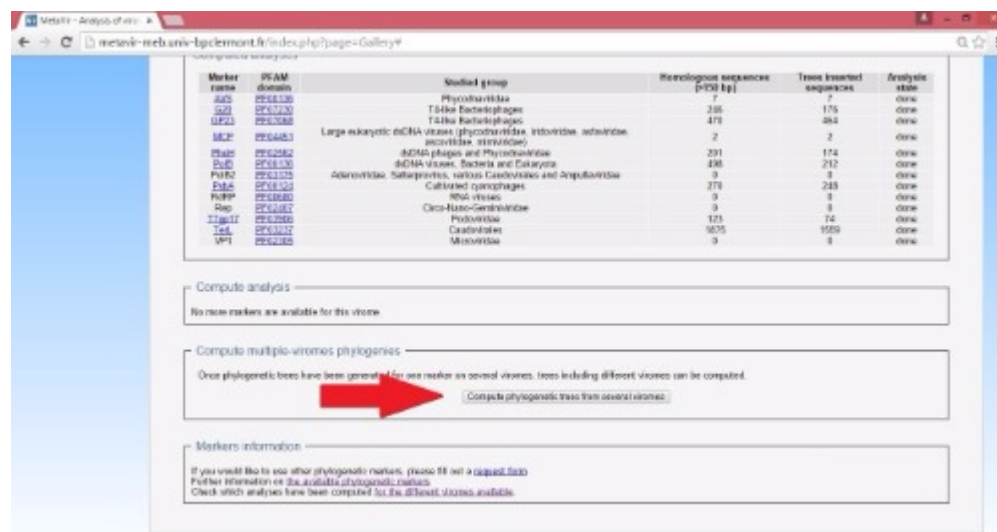
### Step 45.

The fullscreen view is especially useful when dealing with large trees.

## Phylogenetic trees

### Step 46.

Trees can be computed from multiple viromes if the trees with the same marker were computed for each dataset.



## Phylogenetic trees

### Step 47.

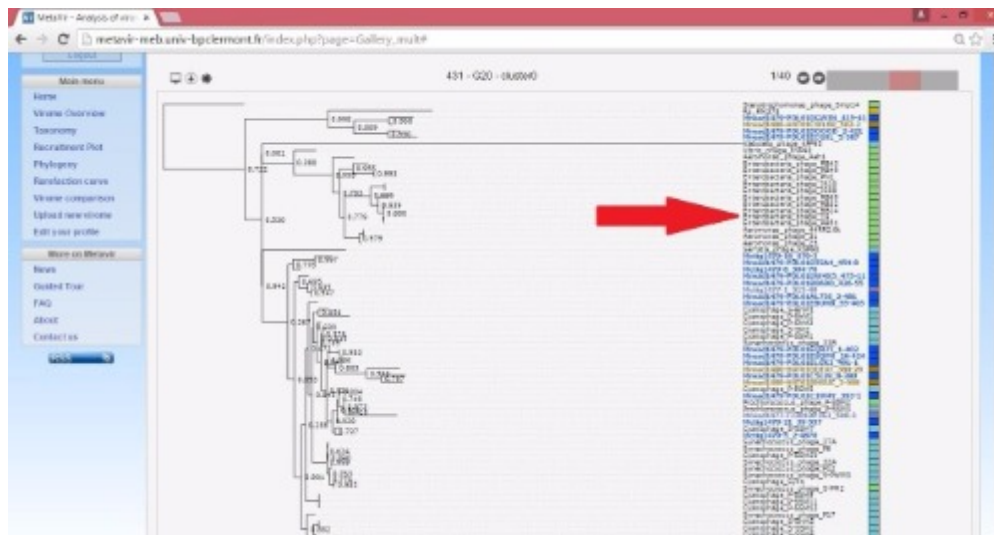
Leaves are then colored according to the virome, reference sequences being left gray.



## Phylogenetic trees

### Step 48.

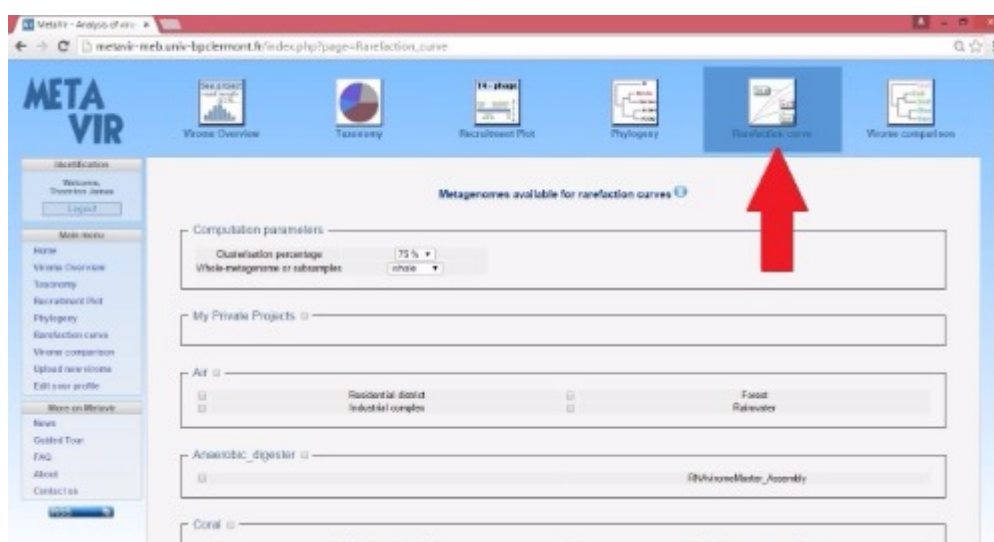
Taxonomy of reference sequences is still indicated next to the leaf name.



## Rarefaction curves

### Step 49.

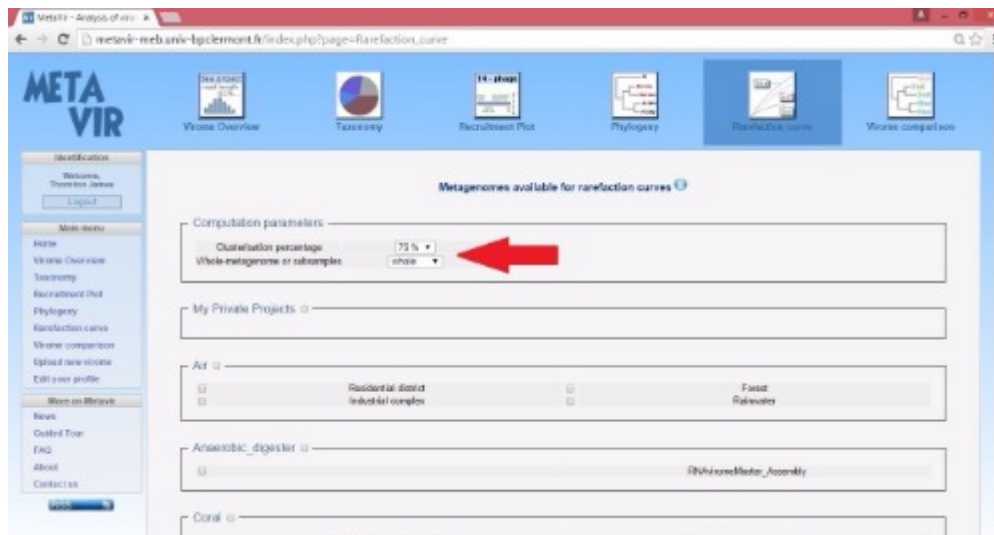
This tab links to rarefaction curves.



## Rarefaction curves

### Step 50.

Three clustering thresholds are available that can be used with whole virome or normalized subsamples (50,000 sequences, 100bp)



## Rarefaction curves

### Step 51.

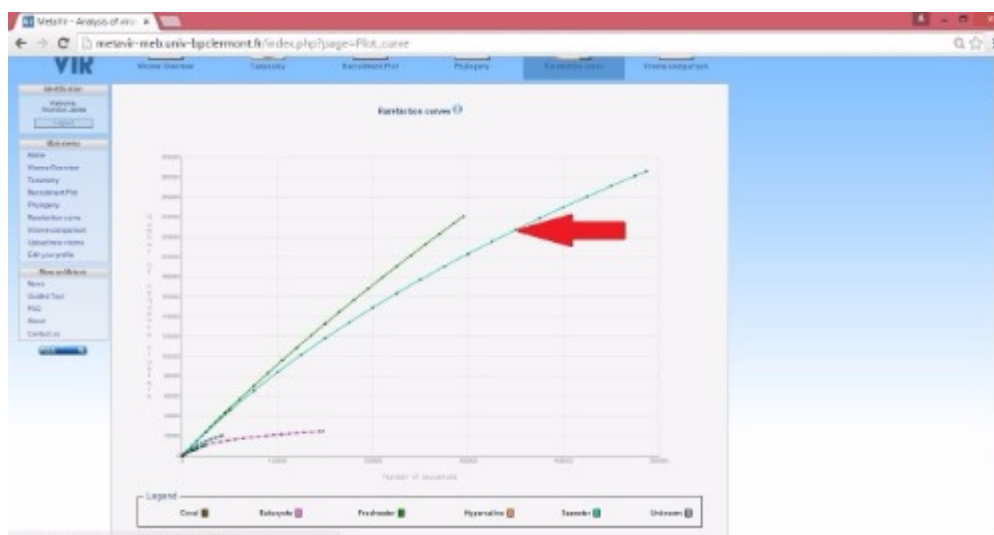
Curve colors are associate with the virome biome.



## Rarefaction curves

### Step 52.

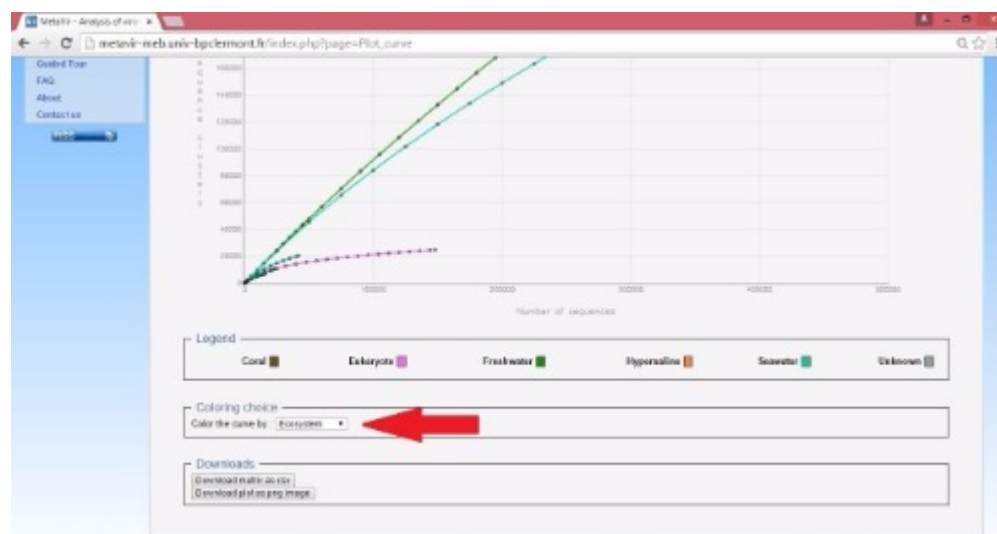
A range of dots on each curve makes it possible to display the virome name when hovered.



## Rarefaction curves

### Step 53.

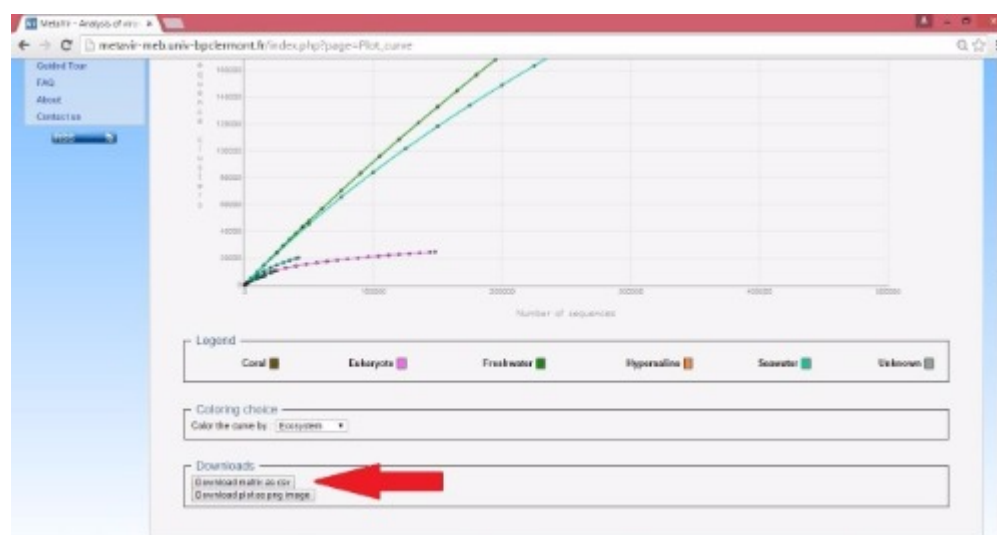
Curves can be colored by virome.



## Rarefaction curves

### Step 54.

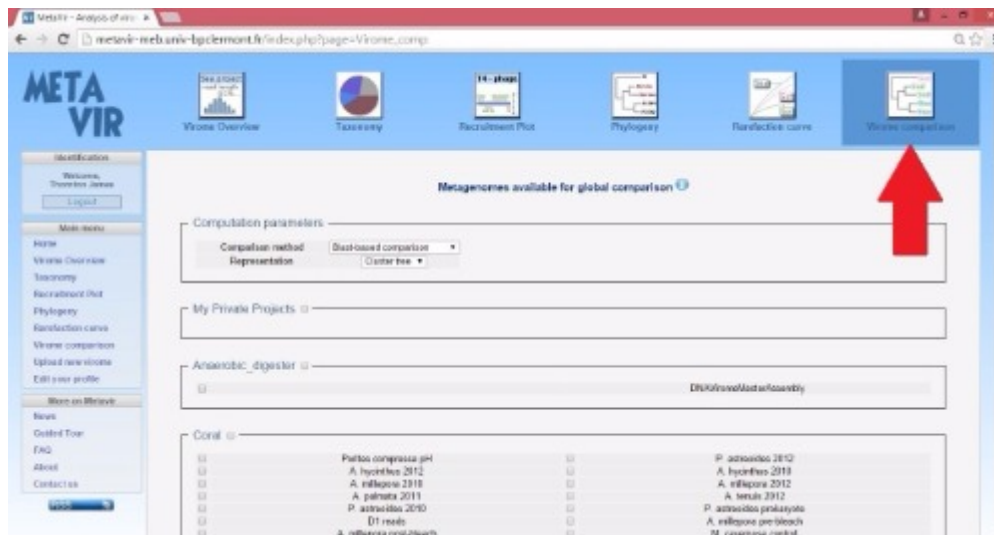
The matrix used to plot the curve is available to download via this link.



## Virome comparison

### Step 55.

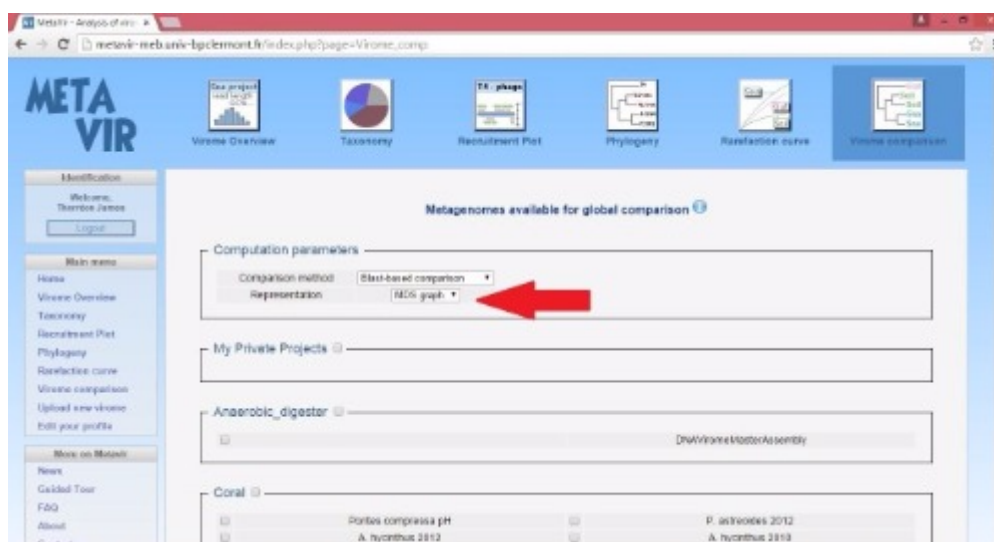
This button links to virome comparison.



## Virome comparison

### Step 56.

Comparisons can be based on k-mer frequencies bias or tBLASTx reads similarities.



## Virome comparison

### Step 57.

An MDS can be computed to display the different viromes.

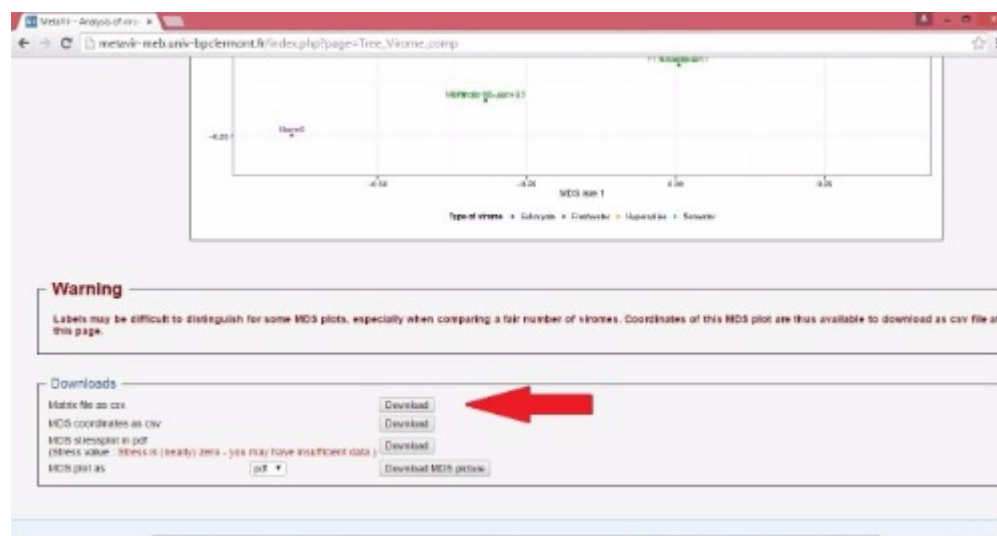




## Virome comparison

### Step 58.

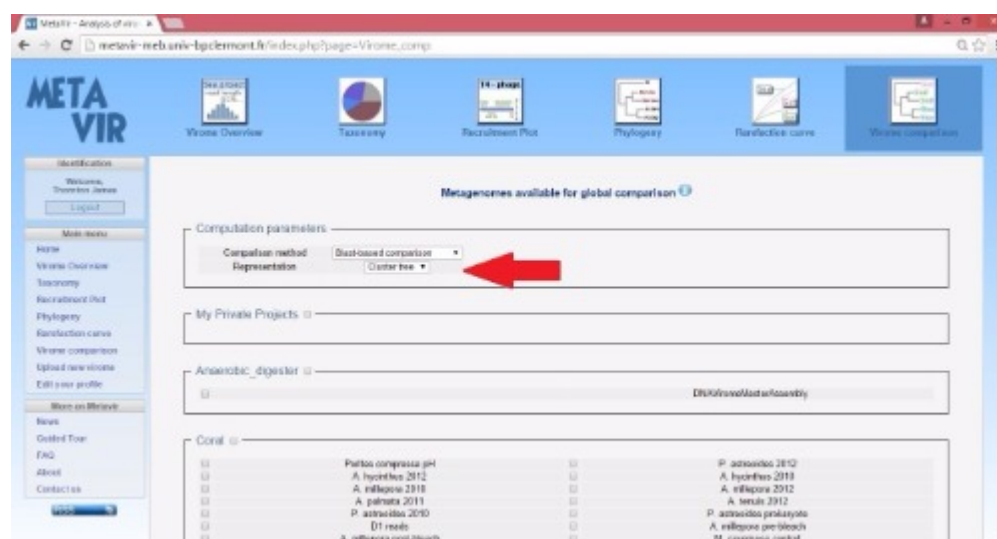
Matrix and plot can be downloaded.



## Virome comparison

### Step 59.

Same comparison can be displayed through a clustering tree.



## Contigs: Taxonomic composition

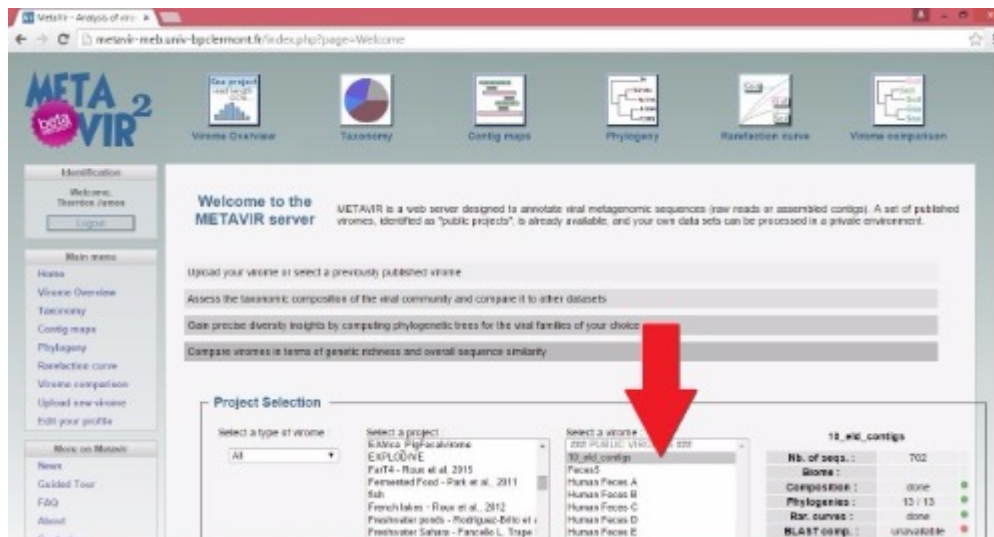
### Step 60.

New tools were developed for contigs.

## Contigs: Taxonomic composition

### Step 61.

They are automatically loaded when a project of type "contigs" is selected.



Contigs: Taxonomic composition

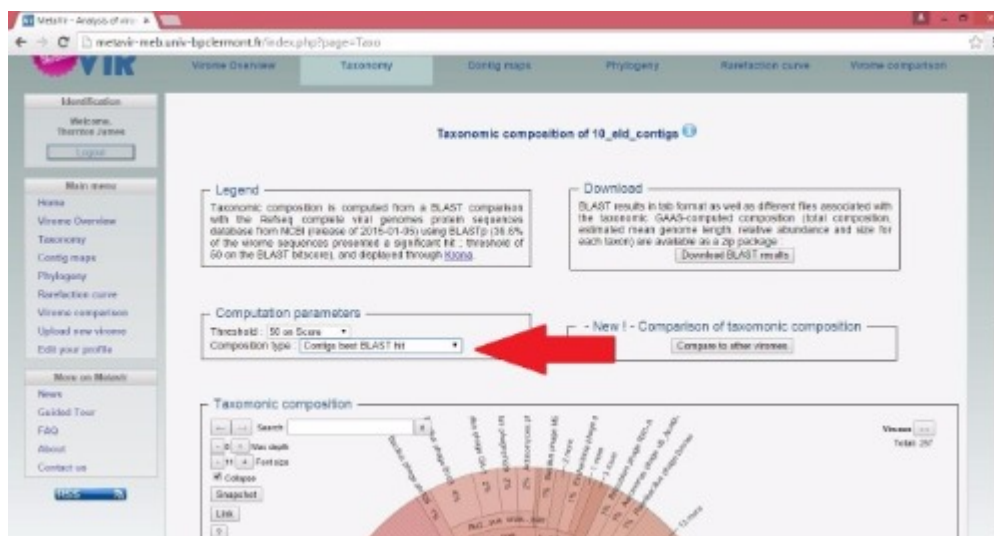
## Step 62.

Three types of taxonomic composition are computed for contigs:

Predicted gene best Blast hit

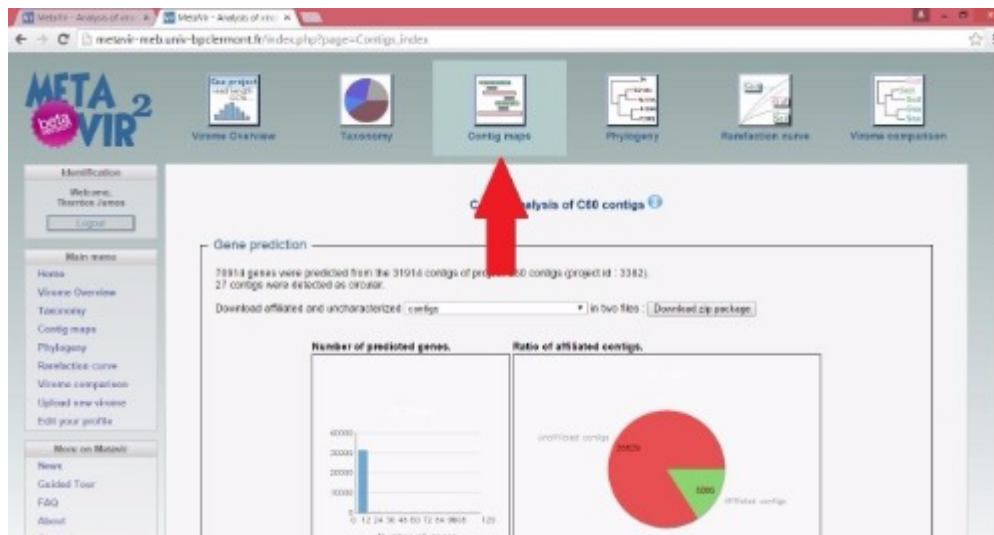
Contig best BLAST hit

Contig lowest common ancestor



## Step 63.

This tab links to the new contigs index.



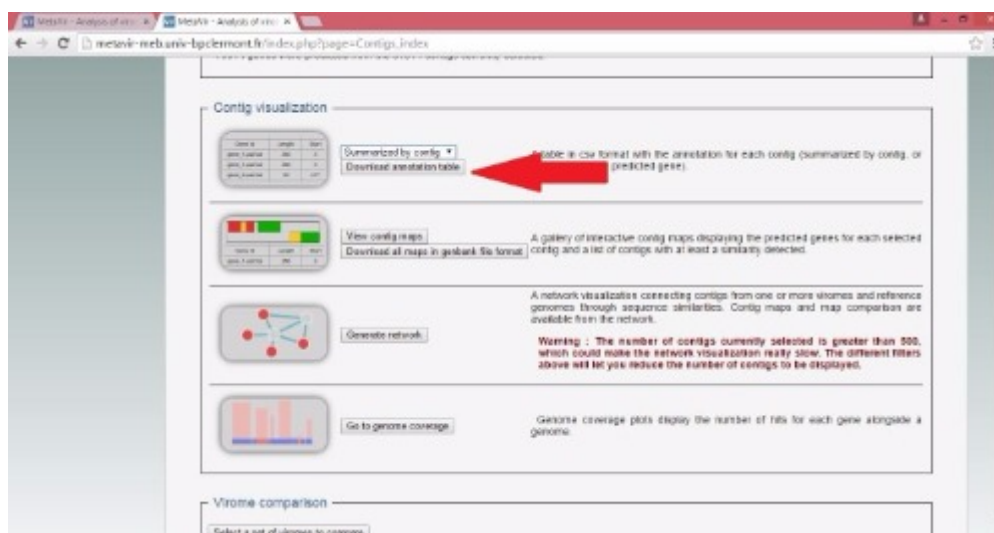
## + NOTES

Bonnie Hurwitz 09 Nov 2015

Loading contigs dataset can take some time...

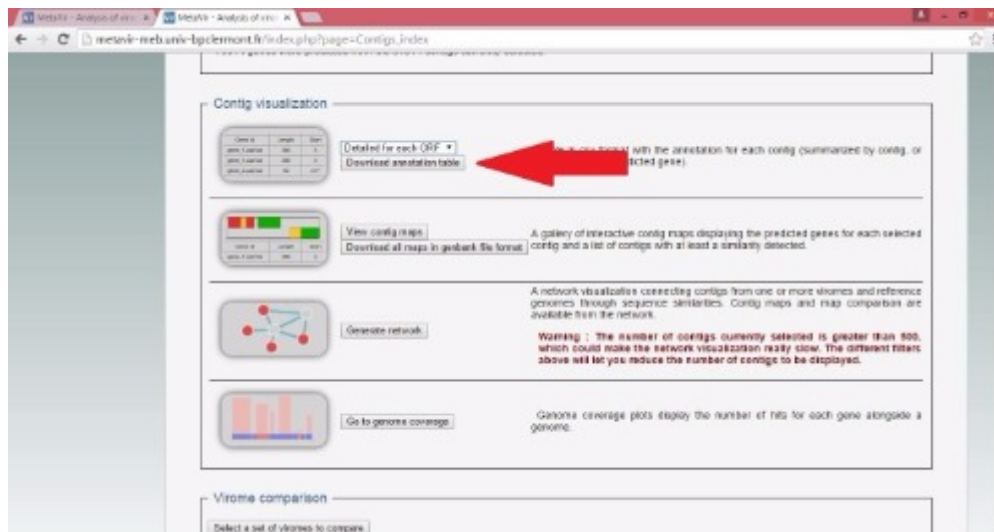
## Step 64.

Contigs annotation can be downloaded as a table.



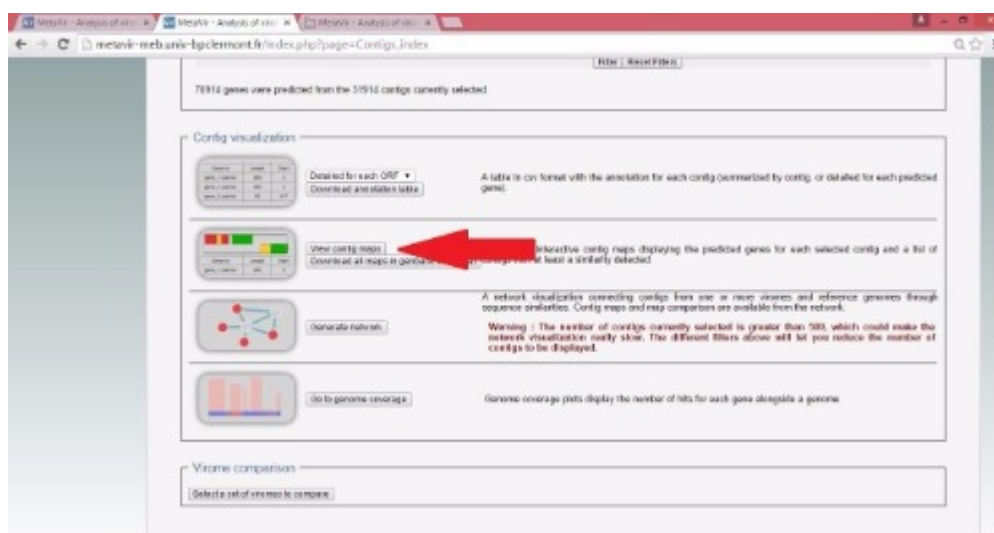
## Step 65.

Annotation of each gene can be downloaded in the same format.



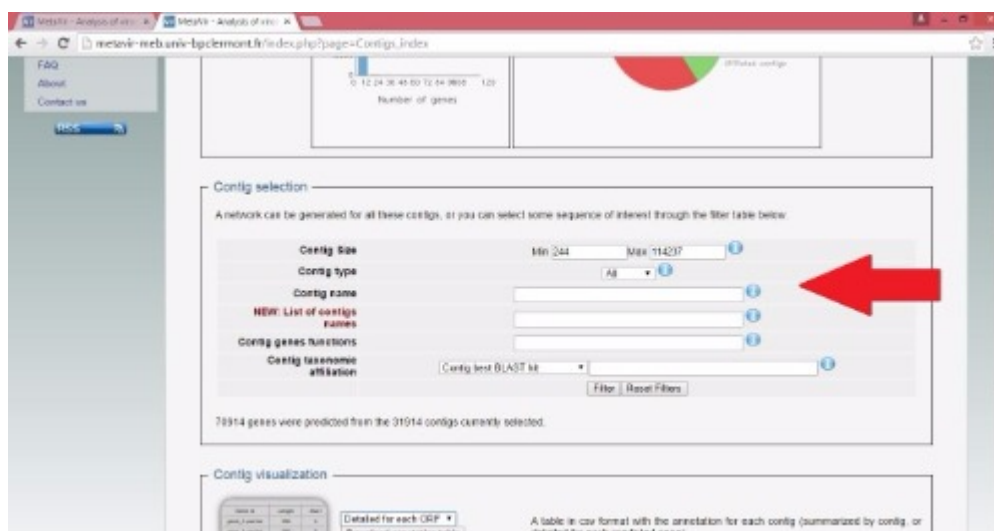
## Step 66.

Annotation can be displayed through contigs map.



## Step 67.

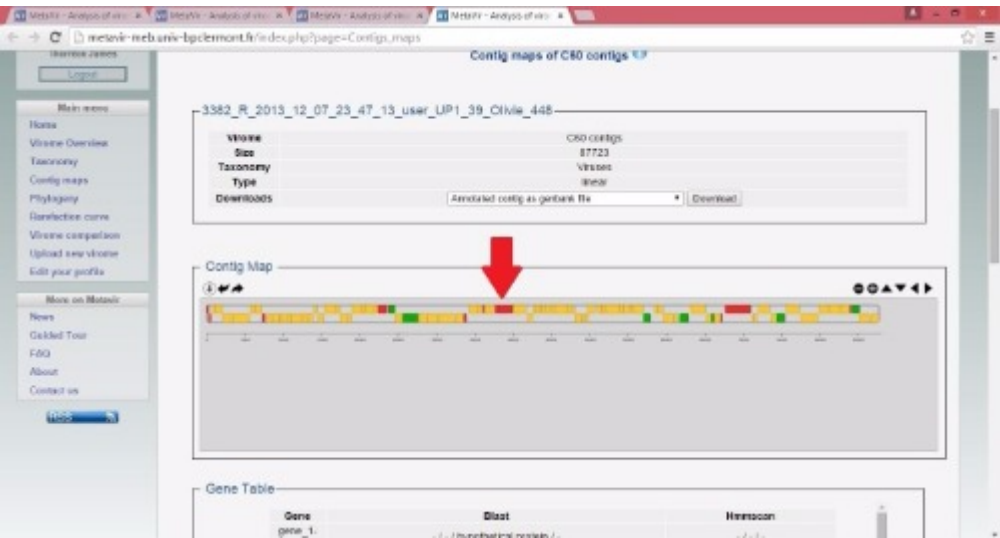
Contigs to be displayed can be filtered based on their name, affiliation or size.



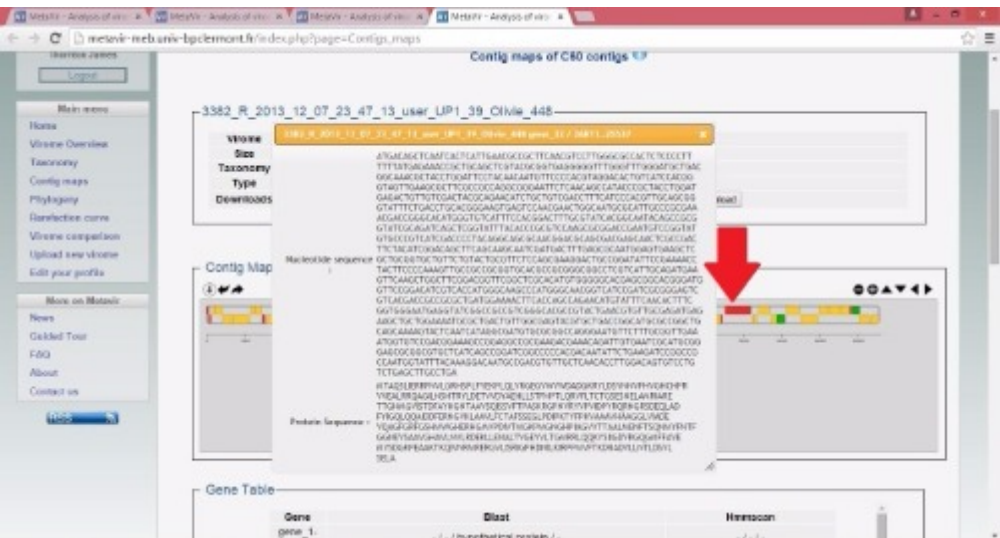
## Step 68.

Predicted genes are indicated along the contig in red (unaffiliated), green (affiliated) and orange

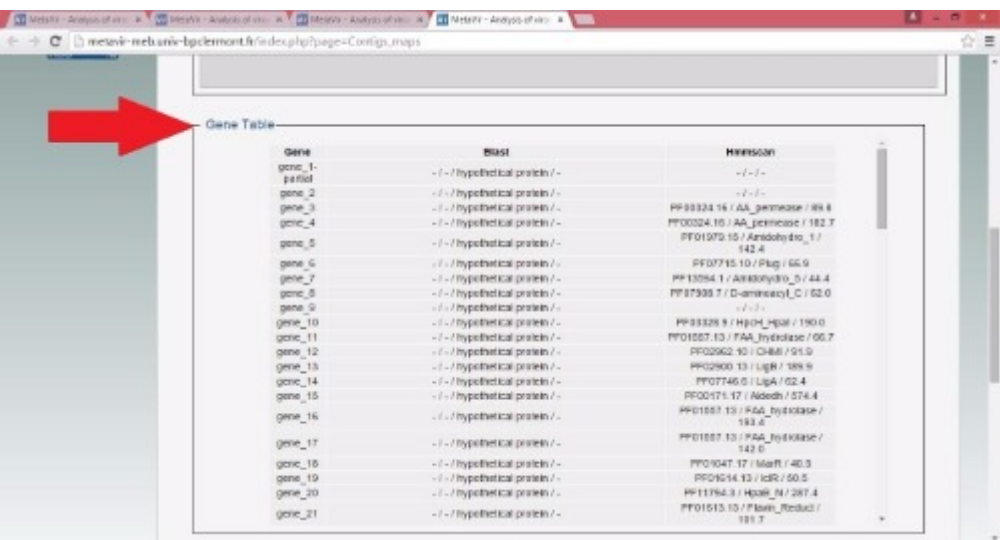
(similarity to a function domain only).



**Step 69.**  
Sequences of each OFR are displayed when clicked.



**Step 70.**  
This table sums up the affiliation of the different predicted genes.





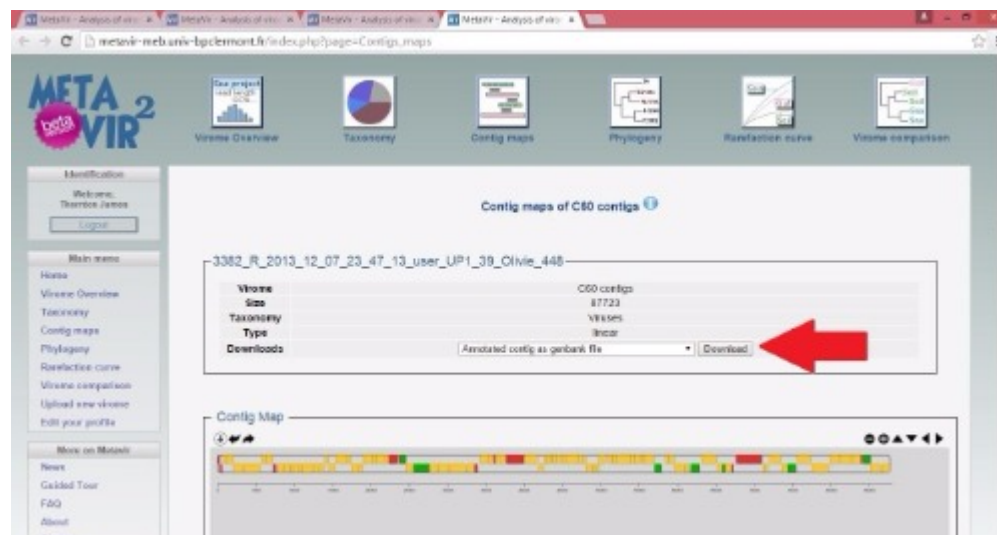
## NOTES

Bonnie Hurwitz 13 Nov 2015

Genes are highlighted on the map when hovered on the table.

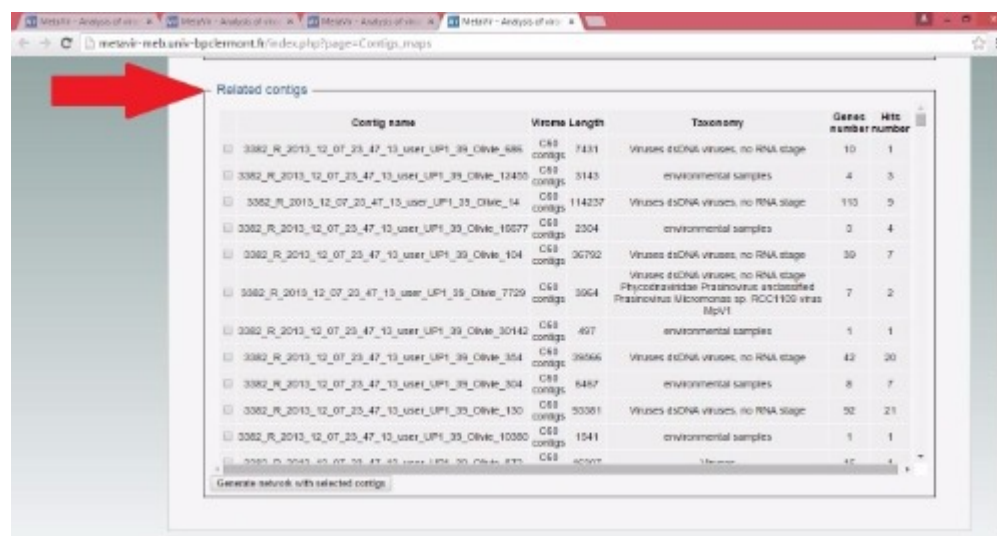
### Step 71.

The corresponding genbank file, contig sequence, and predicted ORFs sequences are available to download.



### Step 72.

All contigs with (at least) an ORF similar to the current contig are listed in this 'Related contigs' table.



### Step 73.


These contigs can then be selected to generate a network.

Metasvi - Analysis of virus...Metasvi - Analysis of virus...Metasvi - Analysis of virus...Metasvi - Analysis of virus...

metasvi-web.univ-bpclermont.fr/index.php?page=Config\_maps

Related contigs

Contig name	Virus Length	Taxonomy	Genes number	Hits number	
<input checked="" type="checkbox"/> 3362_R_2013_12_07_23_47_13_user_UP1_39_Olive_586	C68 contigs	7431	Viruses dsDNA viruses, no RNA stage	10	1
<input checked="" type="checkbox"/> 3362_R_2013_12_07_23_47_13_user_UP1_39_Olive_12450	C68 contigs	3143	environmental samples	4	3
<input checked="" type="checkbox"/> 3362_R_2013_12_07_23_47_13_user_UP1_39_Olive_14	C68 contigs	114237	Viruses dsDNA viruses, no RNA stage	110	5
<input checked="" type="checkbox"/> 3362_R_2013_12_07_23_47_13_user_UP1_39_Olive_10077	C68 contigs	2304	environmental samples	3	4
<input checked="" type="checkbox"/> 3362_R_2013_12_07_23_47_13_user_UP1_39_Olive_104	C68 contigs	30792	Viruses dsDNA viruses, no RNA stage	39	7
<input type="checkbox"/> 3362_R_2013_12_07_23_47_13_user_UP1_39_Olive_7729	C68 contigs	3064	Viruses dsDNA viruses, no RNA stage Phycobirnaviridae Birnavirus unclassified Prasinovirus Micromonas sp. RCC1190 virus Mpv1	7	2
<input type="checkbox"/> 3362_R_2013_12_07_23_47_13_user_UP1_39_Olive_30142	C68 contigs	497	environmental samples	1	1
<input checked="" type="checkbox"/> 3362_R_2013_12_07_23_47_13_user_UP1_39_Olive_354	C68 contigs	39066	Viruses dsDNA viruses, no RNA stage	42	20
<input type="checkbox"/> 3362_R_2013_12_07_23_47_13_user_UP1_39_Olive_304	C68 contigs	6487	environmental samples	8	7
<input type="checkbox"/> 3362_R_2013_12_07_23_47_13_user_UP1_39_Olive_130	C68 contigs	50381	Viruses dsDNA viruses, no RNA stage	52	21
<input type="checkbox"/> 3362_R_2013_12_07_23_47_13_user_UP1_39_Olive_10380	C68 contigs	1041	environmental samples	1	1
<input type="checkbox"/> 3362_R_2013_12_07_23_47_13_user_UP1_39_Olive_875	C68 contigs	3007	environmental samples	17	1

Generate network with selected contigs

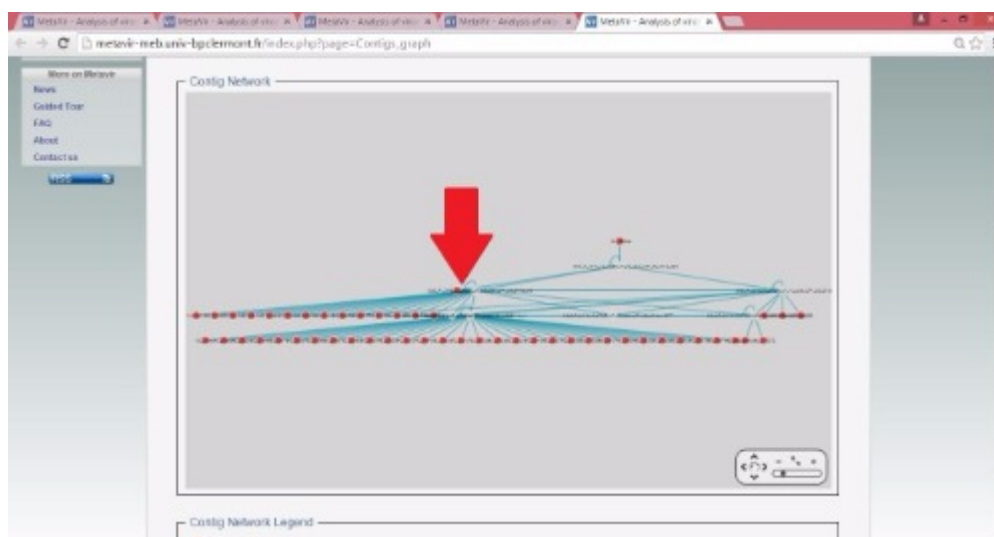
## ⊕ NOTES

**Bonnie Hurwitz** 09 Nov 2015

Network generation can take some time.

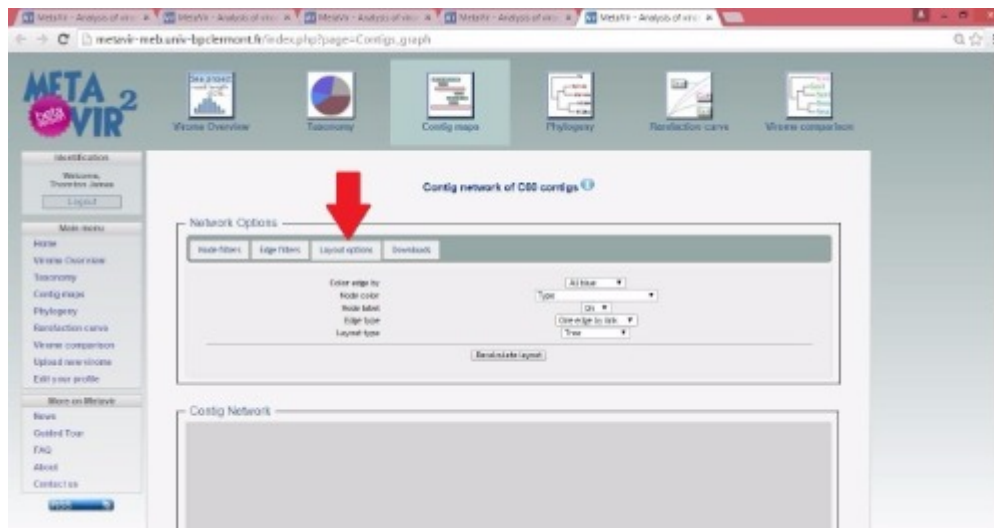
## Step 74.

The default network displays contigs as gray nodes, and reference genomes as red nodes.



## Step 75.

Color and type of layout are fully customizable.

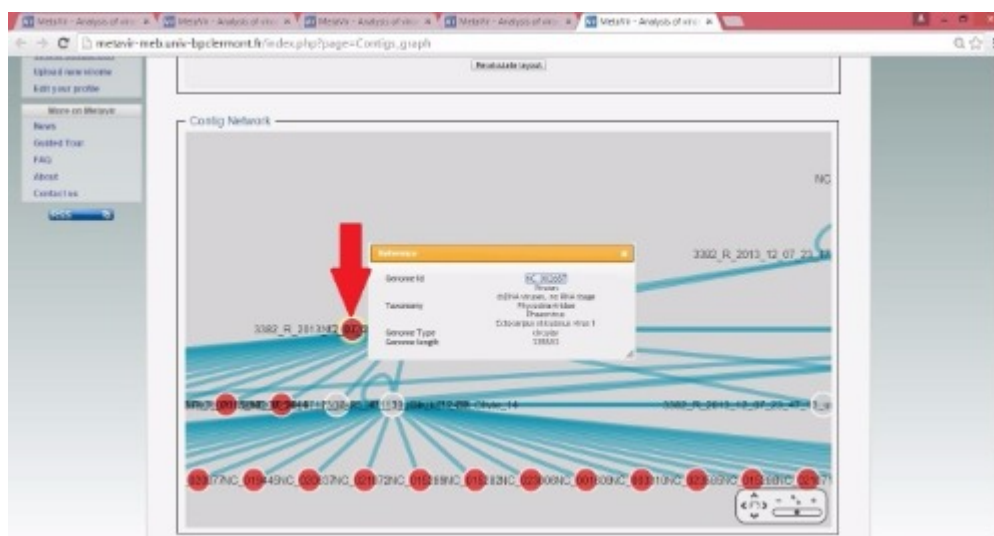


### Step 76.

This network is made with Cytoscape web, and downloadable as a plot or a network file ready to be imported in Cytoscape.

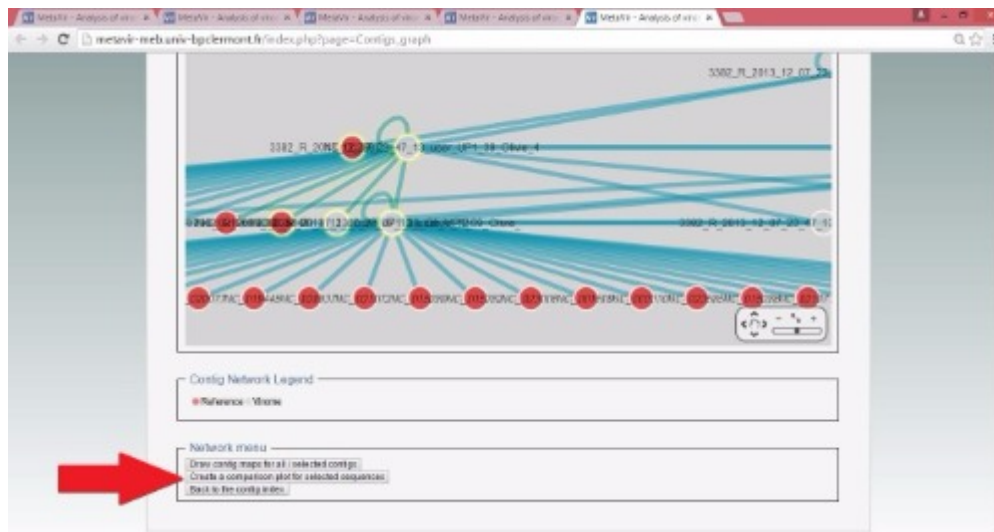
### Step 77.

A click on a network element will display information about the contig, the reference genome, or the BLAST hit.



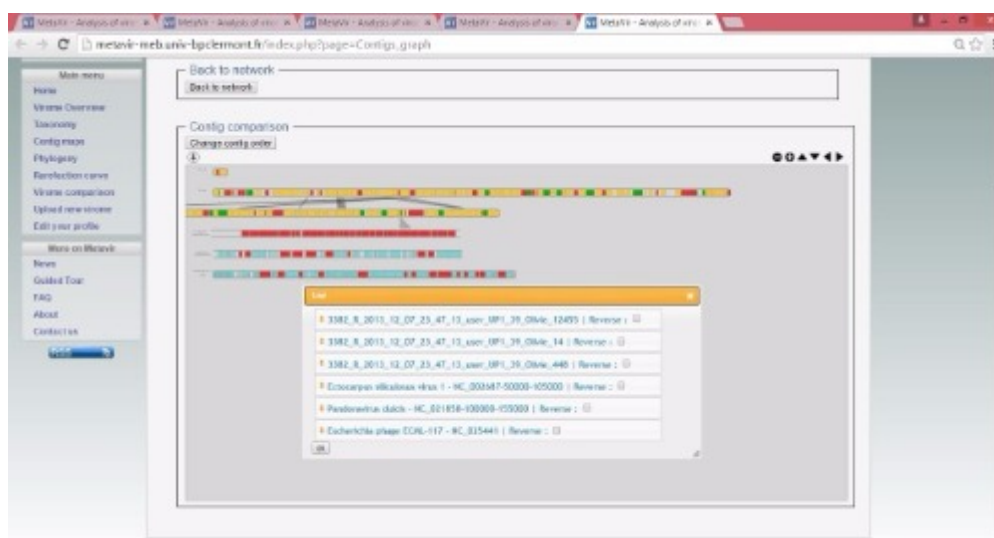
### Step 78.

Finally, a multiple map viewer is available for nodes selected on network.



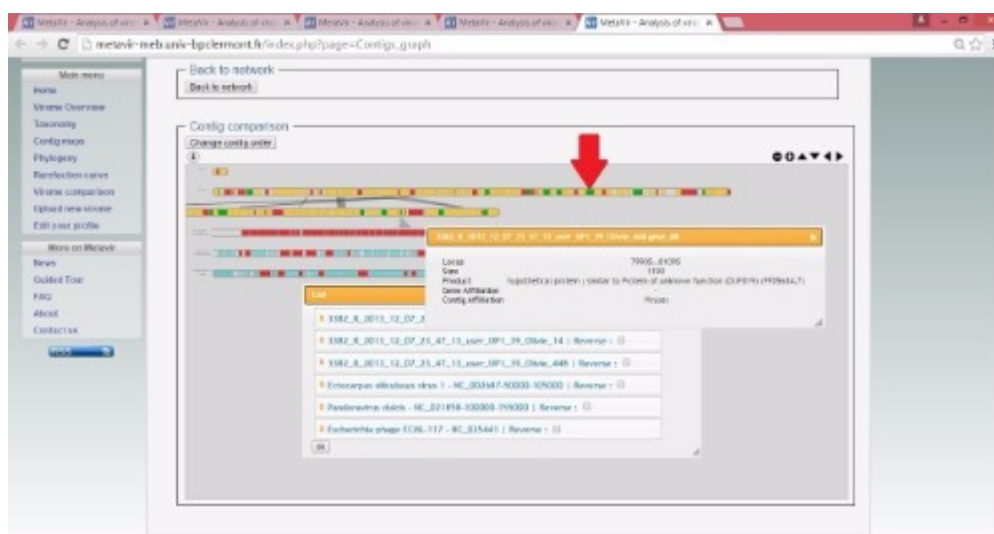
## Step 79.

The order of sequences can be modified with a drag-and-drop. Each sequence can also be reverse-complemented.



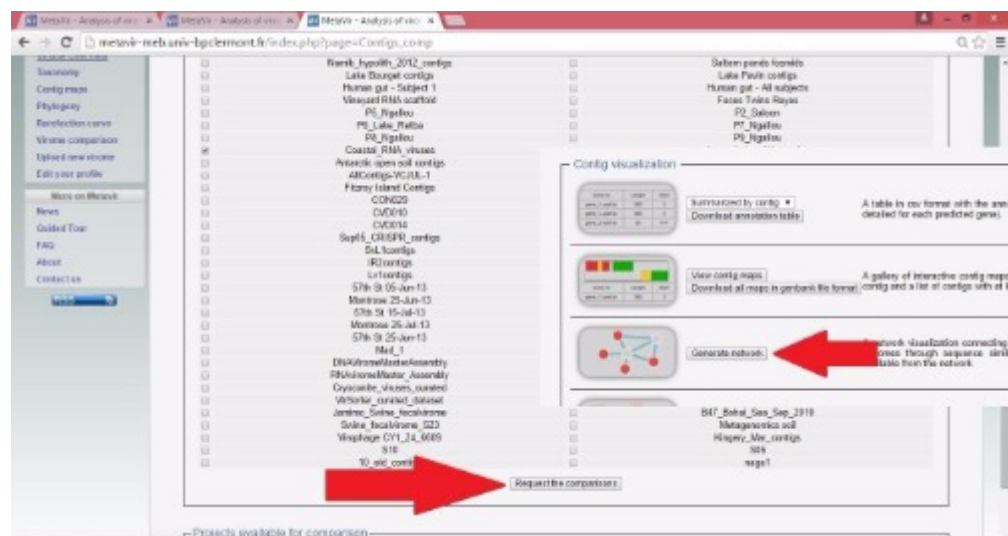
## Step 80.

A click on a gene will display information about it.



## Step 81.

Multiple viromes can be added to the same network / map gallery.



### NOTES

Bonnie Hurwitz 09 Nov 2015

Again, the first loading step can be a bit long.

## Step 82.

Virome nodes can be colored according to the virome.

