

VEME 2022: BV-BRC Tutorial

August 23, 2022

Use Case 1: Finding a specific mutation reported in the literature


A recent paper has shown that "[A single mutation in Crimean-Congo hemorrhagic fever virus discovered in ticks impairs infectivity in human cells](#)". In this manuscript, the authors show that "R1116G", a point mutation in the glycoprotein precursor complex (GPC) protein contributes to host tropism of CCHFV (human versus tick). For this use case, we will compare CCHFV GPC proteins from human and tick hosts using the following methods:


Search and Assemble dataset for exploration (Documentation: [Genome/Protein Search](#))

- Navigate to <https://www.bv-brc.org> and click on "Viruses"

The screenshot shows the BV-BRC website homepage. At the top, there is a navigation bar with the BV-BRC logo and menu items: ORGANISMS, SEARCHES, TOOLS & SERVICES, WORKSPACES, HELP, and ABOUT. Below the navigation bar is the main heading "BACTERIAL AND VIRAL BIOINFORMATICS RESOURCE CENTER" followed by a welcome message and a search bar. The search bar contains the text "All Data Types" and "Find a gene, genome, microarray, etc". Below the search bar is a "BROWSE" section with four buttons: BACTERIA, ARCHAEA, VIRUSES, and EUKARYOTIC HOSTS. The "VIRUSES" button is highlighted with a red box. Below the "BROWSE" section is the "ANALYZE DATA IN BV-BRC" section, which includes a sub-heading "Upload and analyze your data in the private workspace." and four columns of links: SEARCH, ANALYZE, MANAGE DATA, and BATCH ACCESS. At the bottom of the page are three buttons: QUICK START, REFERENCE GUIDES, and TUTORIALS.





























- This will take you to the Virus Overview Homepage.


 ORGANISMS SEARCHES TOOLS & SERVICES
 All Data Types Find a gene, genome, microa


 Virus View
Viruses (8394767 Genomes)

Overview **Taxonomy** Genomes Proteins Protein Structures Domains and Motifs Epitopes Experiments

Virus Data Summary **Virus Families**

Virus Data Summary		Virus Families			
Families	74	Single-Stranded Positive-Sense RNA	Single-Stranded Negative-Sense RNA	Double-Stranded RNA	Featured Viruses
Genera	1685	 <i>Caliciviridae</i>	 <i>Bunyavirales</i>	 <i>Reoviridae</i>	 <i>Bacteriophages</i>
Species	21485	 <i>Coronaviridae</i>	 <i>Filoviridae</i>	Single-Stranded DNA	 <i>Dengue virus</i>
Genomes / Segments	8394761	 <i>Flaviviridae</i>	 <i>Paramyxoviridae</i>	 <i>Parvoviridae</i>	 <i>Ebolavirus</i>
Protein Coding Genes (CDS)	154352767	 <i>Hepeviridae</i>	 <i>Orthomyxoviridae</i>	Partially Double-Stranded DNA	 <i>Enterovirus</i>
Mature Peptides	148816269	 <i>Picornaviridae</i>	 <i>Pneumoviridae</i>	 <i>Hepadnaviridae</i>	 <i>Hepatitis C virus</i>
3D Protein Structures (PDB)	11502	 <i>Togaviridae</i>	 <i>Rhabdoviridae</i>	Double-Stranded DNA	 <i>Influenza A virus</i>
				 <i>Adenoviridae</i>	 <i>Lassa Virus</i>
				 <i>Asfarviridae</i>	 <i>Monkeypox Virus</i>
				 <i>Herpesviridae</i>	 <i>SARS-CoV-2</i>
				 <i>Polyomaviridae</i>	 <i>Zika Virus</i>
				 <i>Poxviridae</i>	

- Note the different Virus Families and Features Viruses available on the right. Also note the Virus Data Summary on the left.
- Now click on "**Bunyavirales**"
- Select the Taxonomy tab and migrate through the Nairoviridae => Orthonairovirus => to select Crimean-Congo hemorrhagic fever orthonairovirus
- Select the "**Genomes**" view option in the green vertical bar

Taxon View
 Viruses » Negamaviricota » Ellioviricetes » Bunyvirales » Nairoviridae » Orthonaviruses » Crimean-Congo hemorrhagic fever orthonaviruses » GENOME STATUS IS

HOST COMMON NAME IS "Human" OR HOST COMMON NAME IS "Tick" (143 Genomes)

Overview | Taxonomy | Strains | Genomes | **Proteins** | Protein Structures | Domains and Motifs | Epitopes | Experiments

DOWNLOAD | KEYWORDS | ADV Search | CDS | PATRIC | FILTERS

Genome Name	Accession	Feature Type	BRC ID	RefSeq Locus Tag	Start	End	Strand	Length (NA)	Gene Symbol	Product
<input checked="" type="checkbox"/>	Crimean-Congo hemorrhagic fe	HM452306	CDS	fig 1980519.65f	93	5147	+	5055	GPC	Pre-glycoprotein polyprotein GP compl
<input checked="" type="checkbox"/>	Crimean-Congo hemorrhagic fe	DQ211628	CDS	fig 1980519.59f	16	5165	+	5094	GPC	Pre-glycoprotein polyprotein GP compl
<input checked="" type="checkbox"/>	Crimean-Congo hemorrhagic fe	DQ211626	CDS	fig 1980519.59f	36	5180	+	5103	GPC	Pre-glycoprotein polyprotein GP compl
<input checked="" type="checkbox"/>	Crimean-Congo hemorrhagic fe	KX056060	CDS	fig 1980519.61f	78	5144	+	5067	GPC	Pre-glycoprotein polyprotein GP compl
<input checked="" type="checkbox"/>	Crimean-Congo hemorrhagic fe	KX056057	CDS	fig 1980519.60f	58	5109	+	5052	GPC	Pre-glycoprotein polyprotein GP compl
<input checked="" type="checkbox"/>	Crimean-Congo hemorrhagic fe	KX056054	CDS	fig 1980519.60f	52	5103	+	5052	GPC	Pre-glycoprotein polyprotein GP compl
<input checked="" type="checkbox"/>	Crimean-Congo hemorrhagic fe	KX056051	CDS	fig 1980519.60f	93	5144	+	5052	GPC	Pre-glycoprotein polyprotein GP compl
<input type="checkbox"/>	Crimean-Congo hemorrhagic fe	KX238957	CDS	fig 1980519.61f	92	763	+	672	GPC	putative Pre-glycoprotein polyprotein G
<input type="checkbox"/>	Crimean-Congo hemorrhagic fe	KX238957	CDS	fig 1980519.61f	885	1683	+	799	GPC	putative Pre-glycoprotein polyprotein G
<input type="checkbox"/>	Crimean-Congo hemorrhagic fe	KX238957	CDS	fig 1980519.61f	1738	5146	+	3409	GPC	putative Pre-glycoprotein polyprotein G
<input checked="" type="checkbox"/>	Crimean-Congo hemorrhagic fe	MF547416	CDS	fig 1980519.64f	93	5147	+	5055	GPC	Pre-glycoprotein polyprotein GP compl
<input checked="" type="checkbox"/>	Crimean-Congo hemorrhagic fe	KY484026	CDS	fig 1980519.61f	36	5180	+	5103	GPC	Pre-glycoprotein polyprotein GP compl
<input checked="" type="checkbox"/>	Crimean-Congo hemorrhagic fe	DQ211630	CDS	fig 1980519.59f	93	5144	+	5052	GPC	Pre-glycoprotein polyprotein GP compl
<input checked="" type="checkbox"/>	Crimean-Congo hemorrhagic fe	KY484029	CDS	fig 1980519.61f	93	5144	+	5052	GPC	Pre-glycoprotein polyprotein GP compl
<input checked="" type="checkbox"/>	Crimean-Congo hemorrhagic fe	MG659726	CDS	fig 1980519.65f	92	5146	+	5055	GPC	Pre-glycoprotein polyprotein GP compl
<input checked="" type="checkbox"/>	Crimean-Congo hemorrhagic fe	KX013445	CDS	fig 1980519.60f	83	5134	+	5052	GPC	Pre-glycoprotein polyprotein GP compl
<input checked="" type="checkbox"/>	Crimean-Congo hemorrhagic fe	MN832722	CDS	fig 1980519.79f	93	5147	+	5055	GPC	Pre-glycoprotein polyprotein GP compl
<input checked="" type="checkbox"/>	Crimean-Congo hemorrhagic fe	KY484032	CDS	fig 1980519.61f	93	5147	+	5055	GPC	Pre-glycoprotein polyprotein GP compl

GUIDE | DWNLD | COPY | FEATURES | GENOMES | MSA | ID MAP | GROUP

- Change "ID Type" in green bar to "Strain"
- **Note sequences causing alignment gaps for potential removal**
- Scroll to the appropriate area "LVSGRSES" consensus region (~1138) to view the region of interest.
- Search for the "Malko Tarnovo-BG2012-T1303" name to find the genome reported in the publication referenced above.

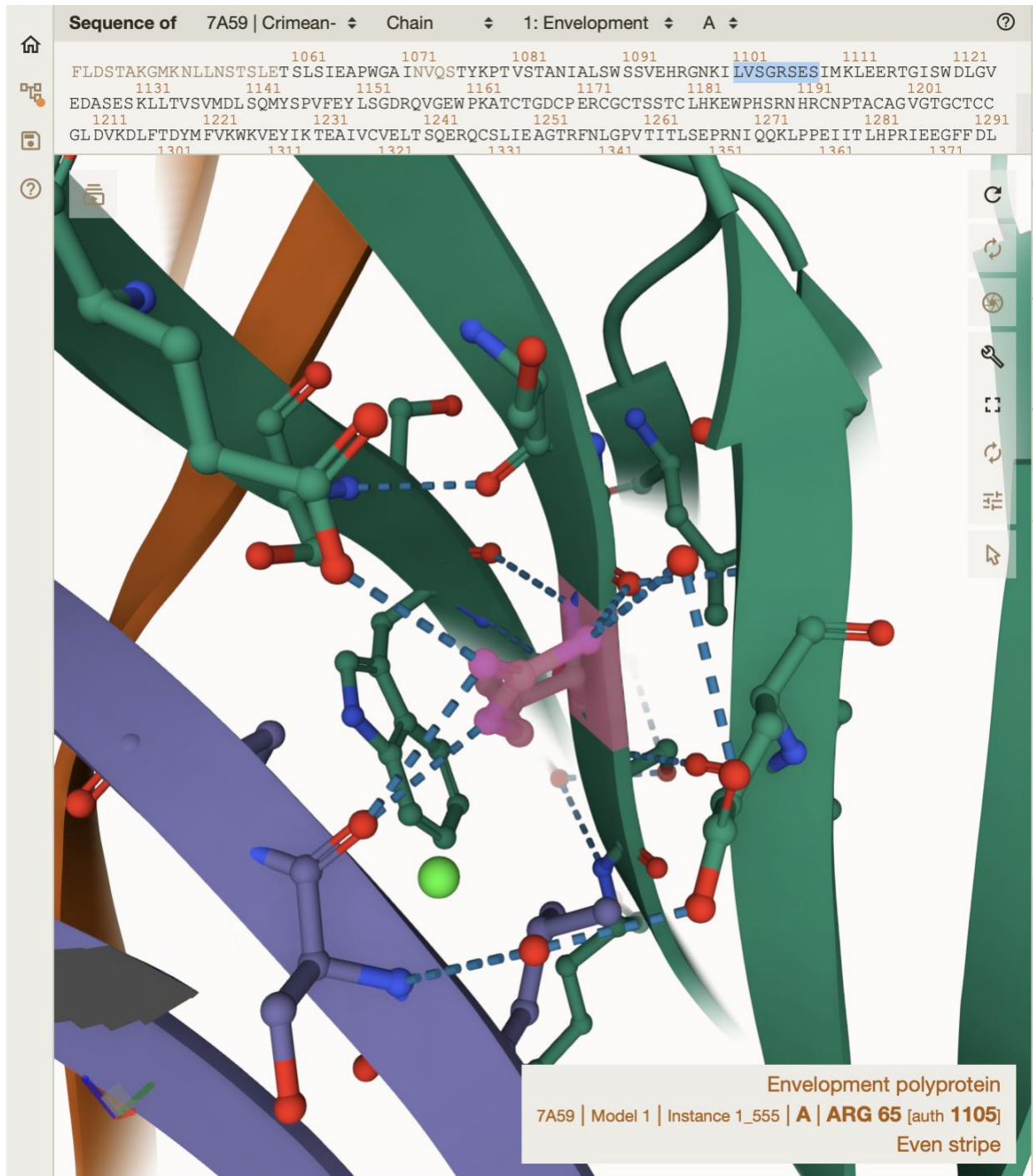
BV-BRC BETA | ORGANISMS | SEARCHES | TOOLS & SERVICES | WORKSPACES | All Data Types | Find a gene, genome, microan

MCL19F1-1812
 NIV1614467
 NIV1727150
 ArD8194
 DAK8194
 UG3010
 UG3010
 UG3010
 Nakiwogo
 YL04057
 79121M18
 Malko Tarnovo-BG2012-T1302
 Malko Tarnovo-BG2012-T1362
Malko Tarnovo-BG2012-T1303
 Pentalofos-Greece-2015
 JD-206
 ArD39554
 37-R-2013
 52-R-2014

E H R G N K I L V S G R S E S I M K L E E R T G I S W S L G V E D A S E S K T L T V S
 E H R G N K I L V S G R S E S I M K L E E R T G I S W S L G V E D A S E S K T L T V S
 E H R G N K I L V S G R S E S I M K L E E R T G I S W C L G V E D A S E S K T L T V S
 E H K G N K I L V S G R S E S I M K L E E R T G I S W C L G V E D A S E S K T L T V S
 E H R G N K I L V S G R S E S I M K L E E R T G I S W S L G V E D A S E S K T L T V S
 E H R G N K I L V S G R S E S I M K L E E R T G I S W S L G V E D A S E S K T L T V S
 E H R G N K I L V S G R S E S I M K L E E R T G I S W S L G V E D A S E S K T L T V S
 E H R G N K I L V S G R S E S I M K L E E R T G V S W N L G V E D A S E S K T L T V S
 E H R G N K I L V S G R S E S I M K L E E R T G V S W N L G V E D A S E S K T L T V S
 E H R G N K I L V S G R S E S I M K L E E R T G M S W N L G V E D A S E S R T L T I S
 E H R G N K I L V S G R S E S I M K L E E R T G M S W N L G V E D A S E S R T L T I S
 E H R G N K I L V S G G S E S I M K L E E R T G M S W N L G V E D A S E S R T L T I S
 E H R G N K I L V S G R S E S I M K L E E R T G I S W N L G V E D A S E S R T L T I S
 E H R G N K V L V S G R S E S I M K L E E R T G I S W N L G V E D A S E S R M L T V S
 E H R G N K I L V T G R S E S I M K L E E R T G I S W N L G V E D A S E S K T L T V S
 E H R G N K I L V T G R S E S I M K L E E R T G V S W D L G V D D A S E S K L L T V S
 E H R G N K I L V T G R S E S I M K L E E R T G V S W D L G V D D A S E S K L L T V S

DETAILS | COLORS | ID TYPE | Filter | DWNLD

- Note the R to G substitution responsible for impaired infectivity of human cells reported in the paper.
- Go to the Protein Structure tab, select 7A59 and "Structure" function in green bar



- Highlight LVSGRSES sequence to add ball and stick structure to ribbon view
- Find ARG at position 1105 in structure
- Note that it coordinates interactions between 4 different beta strands

Use Case 2: Are there consistent amino acid difference in viruses isolated from different hosts or different geographic regions?

Step 1. Sequence selection and quality control

- In Taxonomy tab, select Crimean-Congo hemorrhagic fever orthonairovirus as before
- Select the "**Genomes**" view option in the green vertical bar
- Select filtering criteria
 - Complete
 - M segment
 - Human and tick
- Select "**APPLY**" option above table
- Go to Proteins tab
- Select all records except the "putative....." proteins
- Select the "MSA" function in the green bar, and the "Amino Acids" option
- Note the problematic sequence in the middle of the alignment as candidates for removal



- **Remove the problematic sequences before further analysis (Note that in order to save time during the demo, the sequence curation has already been performed and the curated sequence records have been made available in the VEME public folder - CCHF complete M human+tick curated)**

Step 2. Phylogenetic analysis

- Go to Tools & Services and select Gene Tree service

The screenshot shows the BV-BRC BETA website interface. The top navigation bar includes 'ORGANISMS', 'SEARCHES', 'TOOLS & SERVICES', and 'WORKSPACES'. A dropdown menu for 'TOOLS & SERVICES' is open, displaying various analysis tools. The 'Gene Tree' option under the 'Protein Tools' section is highlighted with a red box. Below the menu, a table lists genome data for Crimean-Congo hemorrhagic fever virus (CCHFV) sequences.

Genome Name	Accession	Feature Type	BRC ID
<input checked="" type="checkbox"/> Crimean-Congo hemorrhagic fe	KJ682807	CDS	fig 1980519.74
<input checked="" type="checkbox"/> Crimean-Congo hemorrhagic fe	KJ682810	CDS	fig 1980519.74
<input checked="" type="checkbox"/> Crimean-Congo hemorrhagic fe	KJ682814	CDS	fig 1980519.74
<input checked="" type="checkbox"/> Crimean-Congo hemorrhagic fe	KJ682806	CDS	fig 1980519.74
<input checked="" type="checkbox"/> Crimean-Congo hemorrhagic fe	KJ682812	CDS	fig 1980519.74
<input checked="" type="checkbox"/> Crimean-Congo hemorrhagic fe	KJ682805	CDS	fig 1980519.74
<input checked="" type="checkbox"/> Crimean-Congo hemorrhagic fe	KJ682809	CDS	fig 1980519.74
<input checked="" type="checkbox"/> Crimean-Congo hemorrhagic fe	KJ682813	CDS	fig 1980519.74

- In the Add/Select Genome Group option, go to the Public Workspace folder and then VEME folder and select the “**CCHF_complete_M_human+tick_curated**” Genome Group
- Click the + button to select
- Perform phylogenetic inferencing using RAXML and the HKY85 model of evolution
- Select “**Output Folder**” and assign an “**Output Name**”

Services

Gene Tree



The Gene Tree Service is being tested. For further explanation, please see the Gene Tree Service [Quick Reference Guide](#) and [Tutorial](#).

Gene Tree



Choose fasta file or features for tree.

DNA PROTEIN



DNA/PROTEIN ALIGNED FASTA


 



UNALIGNED GENE FASTA


FEATURE GROUP

AND/OR SELECT GENOME GROUP 


 

SELECTED FILE/FEATURE TABLE


CCHF_complete_M_human+tick_curated 

Alignment Parameters

TRIM ENDS OF ALIGNMENT THRESHOLD




REMOVE GAPPY SEQUENCES THRESHOLD




Tree Parameters

RAXML PHYML FASTTREE

MODEL



OUTPUT FOLDER







OUTPUT NAME

Reset
Submit





- Monitor progress in the Jobs page

 Uploads
Jobs  0  2  21

Job Status All Services  0 queued  2 running  21 completed  3 failed

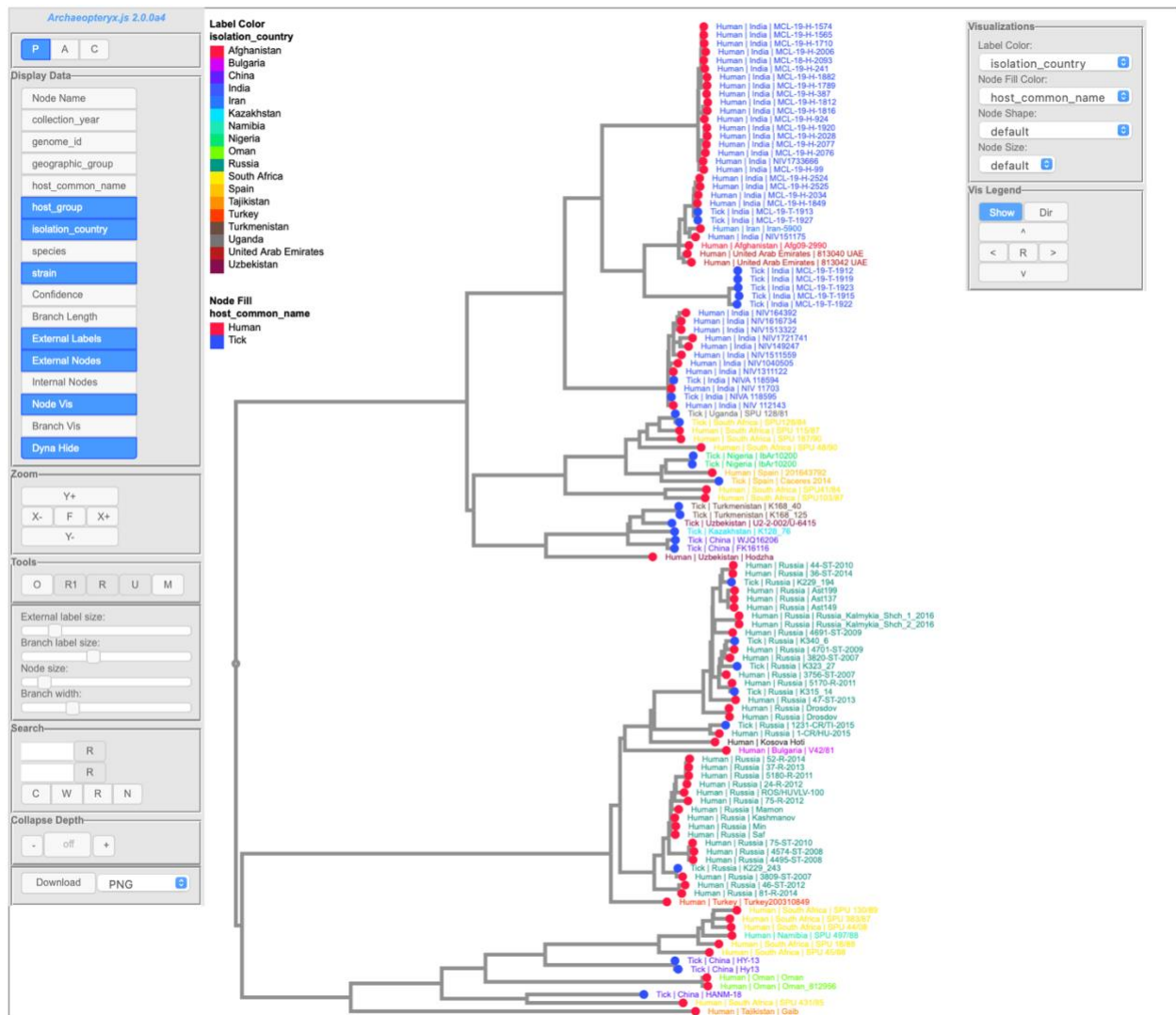
Last updated: 5:11:21

Status	ID	Service	Output Name	Submit	Start	Completed
running	9221237	GeneTree	CCHF_complete_M_hui	8/11/22, 5:05 PM		
completed	9221041	MetaCATS	CCHF_complete_M_hui	8/11/22, 1:29 PM	8/11/22, 1:29 PM	8/11/22, 1:31 PM
completed	9220601	GeneTree	CCHF_complete_M_hui	8/11/22, 10:52 AM	8/11/22, 12:25 PM	8/11/22, 12:36 PM
completed	9220544	MetaCATS	CCHF_complete_M_hui	8/11/22, 9:33 AM	8/11/22, 9:34 AM	8/11/22, 9:34 AM
completed	9220081	GeneTree	CCHF_complete_M_hui	8/10/22, 3:56 PM	8/11/22, 12:10 PM	8/11/22, 12:27 PM
completed	9220080	MSA	CCHF_complete_M_hui	8/10/22, 3:52 PM	8/11/22, 1:55 AM	8/11/22, 1:56 AM
completed	9220072	MetaCATS	CCHF_complete_M_hui	8/10/22, 3:07 PM	8/11/22, 1:55 AM	8/11/22, 1:56 AM

 HIDE
 VIEW
 REPORT
ISSUE...
 RERUN

- When job is complete, select “View” function in green bar
- Select the .phyloxml file and the “View” function in the green bar
- Select the following Display Data options to adjust node labels
 - Host_group
 - Isolation_country
 - Strain
- Adjust External Label Size

- In Visualization:
 - Label color by Isolation_country
 - Node fill color by host common name
- Adjust color as desired



- **Note that the hosts tend to be intermingled whereas the isolation countries match the phylogenetic structures**
- **Note the main segregation between isolates from India and Russia**

Step 3. Comparative genomic of protein sequences between geographic regions

To determine which amino acids differ between CCHF GPC protein sequences from Indian and Russian isolates, we will use the metaCATS service ([MetaCATS, Documentation](#)), which performs statistical analysis of each position in a multiple sequence alignment between selected groups of sequences.

- Use a similar workflow to select protein sequence based on the following criteria for India:
 - Complete
 - M segment
 - Human and Tick
 - India and Russia
- Select “**APPLY**” option above table
- Go to Proteins tab, select GPC proteins
- View MSA – Amino Acids
- **Curate protein selections to remove problematic sequences (done in advance)**
- Save as a Feature Group

- Go to Tools & Services and select Meta-CATS service
- Select “Output Folder” and “Output Name”
- Select Auto Grouping radio button
- Select Isolation Country for Metadata and select the following files from the Public Workspace/VEEME/BV-BRC tutorial folder with the + sign:
 - CCHF_complete_M_human+tick_curated_India_Russia_GPC
- Review group column in table
- Submit job

Metadata-driven Comparative Analysis Tool (meta-CATS)

The meta-CATS tool looks for positions that significantly differ between user-defined groups of sequences. However, biological biases due to covariation, codon biases, and differences in genotype, geography, time of isolation, or others may affect the robustness of the underlying statistical assumptions. For further explanation, please see Metadata-driven Comparative Analysis Tool (meta-CATS) Service [Quick Reference Guide](#) and [Tutorial](#).

Parameters

P-VALUE

OUTPUT FOLDER

OUTPUT NAME

Input

AUTO GROUPING
 FEATURE GROUPS
 ALIGNMENT FILE

METADATA

SELECT FEATURE GROUP

DNA
 PROTEIN

GROUP NAMES Change group

Delete Rows *Max groups 10. Current 2 groups.*

GROUPS GRID

<input type="checkbox"/>	BRC ID	Metadata	Group	Genome ID
<input checked="" type="checkbox"/>	fig 198051	India	India	198051
<input checked="" type="checkbox"/>	fig 198051	India	India	198051
<input checked="" type="checkbox"/>	fig 198051	India	India	198051
<input checked="" type="checkbox"/>	fig 198051	India	India	198051
<input checked="" type="checkbox"/>	fig 198051	India	India	198051
<input checked="" type="checkbox"/>	fig 198051	India	India	198051

- Monitor progress in the Jobs page
- When job is complete, select “View” function in green bar
- View the GPC_India_vs_Russia_metaCATS-chisqTable.tsv file
- Sort table by Chi-square value from largest to smallest

First Row Contains Column Headers

Position	Chi-square_value	P-value	Degrees_of_freedom	Fewer_5	India	Russia
<input type="checkbox"/> 126	79.993974135199	4.26117356163036e-18	2	Y	42 R	1 N, 37 S
<input type="checkbox"/> 153	79.993974135199	4.26117356163036e-18	2	Y	42 T	1 P, 37 S
<input type="checkbox"/> 325	79.993974135199	4.26117356163036e-18	2	Y	42 G	1 N, 37 S
<input type="checkbox"/> 369	79.993974135199	4.26117356163036e-18	2	Y	42 N	1 S, 37 T
<input type="checkbox"/> 1610	79.993974135199	4.26117356163036e-18	2	Y	42 I	1 A, 37 V
<input type="checkbox"/> 40	79.9939741346757	4.26117356274532e-18	2	Y	41 D, 1 N	38 T
<input type="checkbox"/> 35	79.9939721391984	4.26117781428486e-18	2	Y	42 M	2 I, 36 T
<input type="checkbox"/> 70	79.9939721391984	4.26117781428486e-18	2	Y	42 L	2 A, 36 T
<input type="checkbox"/> 108	79.9939721391984	4.26117781428486e-18	2	Y	42 S	36 D, 2 N
<input type="checkbox"/> 159	79.9939721391984	4.26117781428486e-18	2	Y	42 E	2 A, 36 T
<input type="checkbox"/> 166	79.9939721391984	4.26117781428486e-18	2	Y	42 S	36 P, 2 T
<input type="checkbox"/> 227	79.9939721391984	4.26117781428486e-18	2	Y	42 Q	36 M, 2 V
<input type="checkbox"/> 298	79.9939721391984	4.26117781428486e-18	2	Y	42 T	36 D, 2 E
<input type="checkbox"/> 107	79.9939721381098	4.26117781660424e-18	2	Y	40 A, 2 V	38 T
<input type="checkbox"/> 24	79.9939714751559	4.26117922908684e-18	2	Y	42 H	3 P, 35 S
<input type="checkbox"/> 208	79.9939714751559	4.26117922908684e-18	2	Y	42 L	3 P, 35 S
<input type="checkbox"/> 876	79.9939714734553	4.26117923271e-18	2	Y	39 M, 3 T	38 V
<input type="checkbox"/> 1483	79.9939711447461	4.26117993305444e-18	2	Y	42 D	4 N, 34 S
<input type="checkbox"/> 216	79.9939709479988	4.26118035224225e-18	2	N	42 S	5 L, 33 P
<input type="checkbox"/> 19	79.9939709449127	4.26118035881761e-18	2	N	5 E, 37 K	38 G
<input type="checkbox"/> 392	79.9939709449127	4.26118035881761e-18	2	N	37 I, 5 V	38 T
<input type="checkbox"/> 448	79.9939709449127	4.26118035881761e-18	2	N	37 A, 5 T	38 P
<input type="checkbox"/> 367	79.9939708182263	4.26118062873445e-18	2	N	42 T	6 L, 32 S

- Note positions that differ most between the two groups
- Go to Public – VEME folder
- Select the CCHF_complete_M_human+tick_curated_India_Russia_GPC Feature Broup
- Run MSA – Amino Acids
- Change ID Type to isolation_country
- View significant positions (e.g., 126)

Use Case 3: Isolation and characterization of an “unknown” Nairovirus

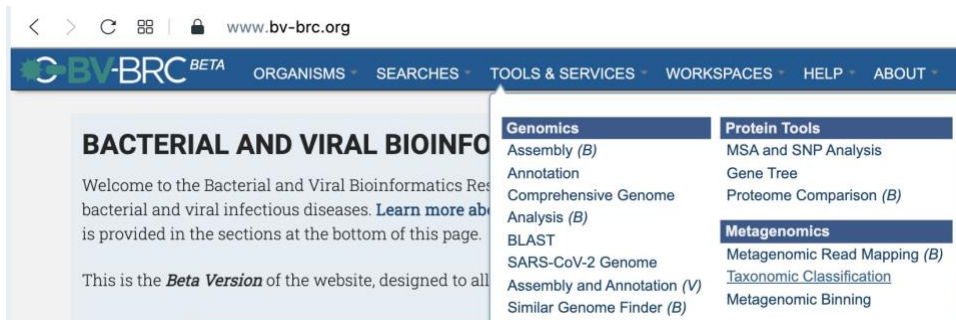
Often, researchers or clinicians encounter patients with symptoms of hemorrhagic fever but with unknown etiology. In this case, diagnostic measures may include whole genome sequencing of a patient sample, to try to detect the causative infectious disease agent. For the purposes of this exercise, we will pretend that the following raw reads deposited in the Sequence Read Archive (SRA) database, are from just such a sample. Run number:

[SRR10769498](https://www.ncbi.nlm.nih.gov/sra/SRR10769498)

Step 3a) SRA file analysis ([Taxonomic classifier](#), [Documentation](#))

In order to assess read content in this sample, we will use taxonomic classification.

- Navigate to the “Taxonomic Classification” tool underneath the “Metagenomics” header in the “TOOLS & SERVICES” tab.




- Input the above SRA run number (SRR10769498) into the appropriate box (red rectangle below), then click the indicated arrow (red circle below) to move the dataset to the “Selected Libraries” box.


Services


Taxonomic Classification


The Taxonomic Classification Service computes taxonomic classification for read data. For further explanation, please see the [Taxonomic Classification Service Quick Reference Guide](#) and [Tutorial](#).


Start With: 


READ FILE ASSEMBLED CONTIGS


Input File 


PAIRED READ LIBRARY 

↓ ↕ READ FILE 1 


↓ ↕ READ FILE 2 

SINGLE READ LIBRARY 



↓ ↕ READ FILE 

SRA RUN ACCESSION 

SRR10769498

Selected libraries 

Place read files here using the arrow buttons.

SRR10769498  

- Select the appropriate parameters as shown below, specifying your desired “output folder” and “output name”. Once selected, the “submit” button can be clicked to launch the job.

Parameters ⓘ

ALGORITHM

DATABASE

SAVE CLASSIFIED SEQUENCES
 NO YES

SAVE UNCLASSIFIED SEQUENCES
 NO YES

OUTPUT FOLDER

OUTPUT NAME

- To view results, click on the job status box in the bottom right corner of the webpage.

☰ Queued | ▶ Running | ✓ Completed

📁 Uploads 📁 Jobs ☰ 0 | ▶ 0 | ✓ 138

- Navigate to your job and click on “View” to view results.

Job Status All Services ☰ 0 queued ▶ 0 running ✓ 138 completed ⚠ 35 failed

Last updated: 11:46:09

Status	ID	Service	Output Name	Submit	Start	Completed
completed	6987745	Annotation	Nairoviridae Annotation	3/22/22, 8:57 AM	3/22/22, 8:57 AM	3/22/22, 9:01 AM
completed	6987740	Homology	Test2	3/22/22, 8:55 AM	3/22/22, 8:55 AM	3/22/22, 8:55 AM
completed	6987734	GenomeAssembly2	Assembly_CCHFV	3/22/22, 8:51 AM	3/22/22, 8:52 AM	3/22/22, 8:52 AM
completed	6987727	FastqUtils	CCHFV_Align	3/22/22, 8:48 AM	3/22/22, 8:48 AM	3/22/22, 8:55 AM
completed	6987720	TaxonomicClassification	SRR10769498_Tax	3/22/22, 8:45 AM	3/22/22, 8:45 AM	3/22/22, 8:52 AM

👁 VIEW
🗨 REPORT ISSUE...
🔄 RERUN

- A list of output files is provided for the user (see explanations in the documentation). You may explore these individually or use the eye-shaped “View” icon (red circle below) to navigate to the “Taxonomic Report”.



TaxonomicClassification Job Result

Job ID	6987720
Start time	3/22/22, 8:45 AM
End time	3/22/22, 8:52 AM
Run time	7m15s
Parameters	

Name	Size	Owner	Members	Created
Parent folder				
TaxonomicReport.html	19.1 kB	me	Only me	3/22/22, 8:52 AM
chart.html	2.5 MB	me	Only me	3/22/22, 8:52 AM
classified.fastq.gz	55.1 MB	me	Only me	3/22/22, 8:52 AM
full_report.txt	2.0 MB	me	Only me	3/22/22, 8:52 AM
output.txt.gz	14.9 MB	me	Only me	3/22/22, 8:52 AM
report.txt	427.8 kB	me	Only me	3/22/22, 8:52 AM

- Results can be viewed either in a “Table format” or as an “Interactive chart” (see below).

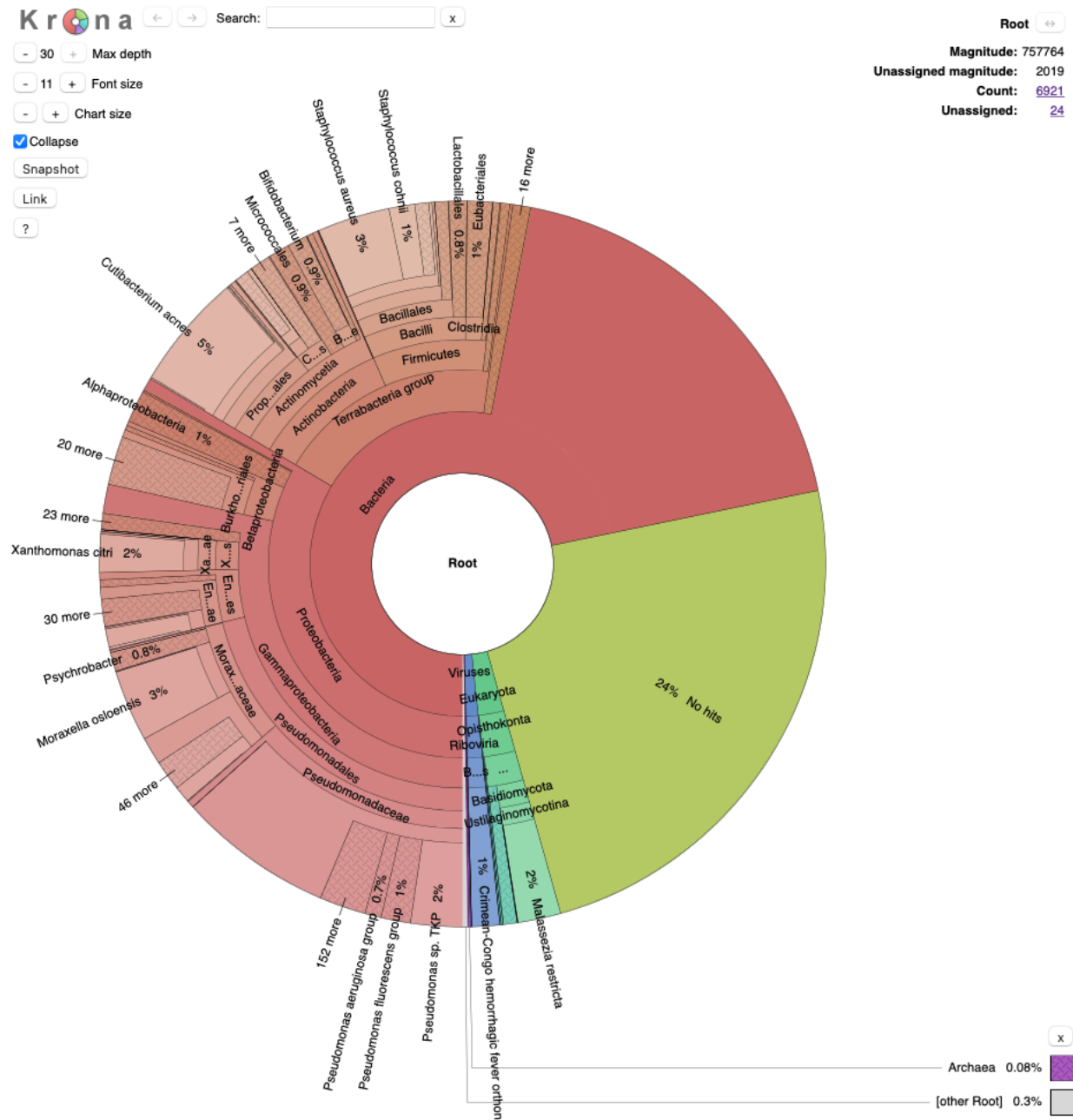
html file: TaxonomicReport.html

.....

In addition, the output file [output.txt.gz](#) contains information about each input sequence. Documentation on this format is available [here](#).

Pct Coverage	Frag in Clade	Frag in Taxon	Rank	NCBI Taxon ID	Scientific Name
1.29	9809	0	D	10239	Viruses
1.24	9418	0	D1	2559587	Riboviria
1.24	9386	0	K	2732396	Orthonavirae
1.24	9386	0	P	2497569	Negarnaviricota
1.24	9386	0	P1	2497571	Polyploviricotina
1.24	9386	0	C	2497576	Ellioviricetes
1.24	9386	0	O	1980410	Bunyavirales
1.24	9378	0	F	1980415	Nairoviridae
1.24	9378	0	G	1980517	Orthonairovirus
1.24	9378	9378	S	1980519	Crimean-Congo hemorrhagic fever orthonairovirus

html file: [chart.html](#) 



- Results for this SRA run number indicate the presence of Crimean-Congo Hemorrhagic Fever Virus reads.

Step 3b) Read QC and mapping ([Fastq Utilities](#), [Documentation](#))

After detection of a virus of interest in a sample, users may want to assemble their reads into viral contigs for further analysis. While reference-based sequencing is not yet available for viral sequences (*pipeline coming soon!*), users can utilize a combination of our “Fastq Utilities” and *de novo* sequence “Assembly” services to extract and assemble their viral reads (see below).

- Navigate to the “Fastq Utilities” tool underneath the “Utilities” header in the “TOOLS & SERVICES” tab.



- Enter the SRA run accession number from the previous step (don't forget to press the arrow to transfer the data to the “Selected libraries” box as in the previous step!) - Specify your desired “output folder” and “output name”.
- Select the appropriate analysis pipelines, “FastQC” and “Align”, as shown below, and click the “+” button to add these services.
- Select the appropriate “Target Genome”, in this case, CCHFV.
- Once selected, the “submit” button can be clicked to launch the job.

Services

Fastq Utilities ? 🔗

The Fastq Utilities Service provides capability for aligning, measuring base call quality, and trimming fastq read files. For further explanation, please see the Fastq Utilities Service Quick Reference Guide and Tutorial.

Parameters ⓘ

OUTPUT FOLDER

OUTPUT NAME

Paired read library ⓘ

READ FILE 1

READ FILE 2

Single read library ⓘ

READ FILE

SRA run accession ⓘ

SRR ACCESSION

Pipeline ⓘ

Align +

FastQC x

Align x

TARGET GENOME

Crimean-Congo Hemorrhagic Fever Genome name.

- Crimean-Congo hemorrhagic fever orthonairovirus IbAr10200 [1980519.4148]
- Crimean-Congo hemorrhagic fever orthonairovirus Semunya [1980519.5881]
- Crimean-Congo hemorrhagic fever orthonairovirus Iran/IR-T1-HAS/2014 [1980519.5895]
- Crimean-Congo hemorrhagic fever orthonairovirus Iran/IR-T2-HAN/2014 [1980519.5896]
- Crimean-Congo hemorrhagic fever orthonairovirus Iran/IR-T3-HSU/2015 [1980519.5897]
- Crimean-Congo hemorrhagic fever orthonairovirus Iran/IR-T4-HAS/2015 [1980519.5898]
- Crimean-Congo hemorrhagic fever orthonairovirus SPU128/81/7 [1980519.5883]
- Crimean-Congo hemorrhagic fever orthonairovirus SPU128/81/7 [1980519.5893]
- Crimean-Congo hemorrhagic fever orthonairovirus SPU4/81 [1980519.5894]
- Crimean-Congo hemorrhagic fever orthonairovirus Iran/IR-T5-HAN/2014 [1980519.5899]

Reset Submit

- Once your job has completed and you have selected the appropriate job from the list you can view the results of either the “FastQC” or “Align” pipelines (See eye view icon below).

FastqUtils Job Result

Job ID	6987727
Start time	3/22/22, 8:48 AM
End time	3/22/22, 8:55 AM
Run time	6m9s
Parameters	

VIEW HIDE

Select View

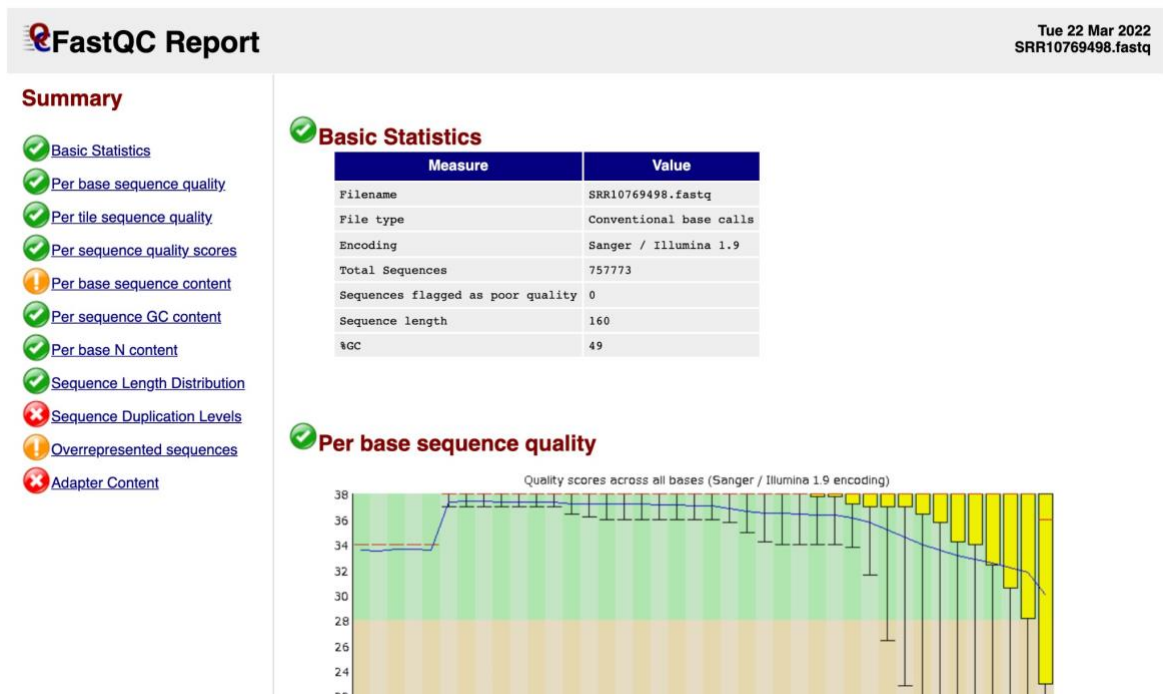
FastQC

Align

DWNLD

- The FastQC results summarize several quality control metrics for the sequence reads as shown below (for more information on each of these, see linked service documentation above).

html file: [SRR10769498_fastqc.html](#)

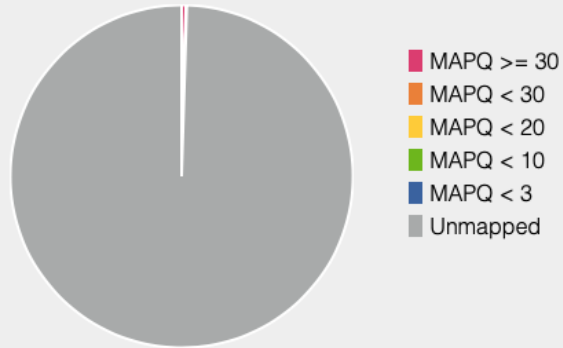


- For the results of the “Align” pipeline, a summary of reads, read length, and base quality is displayed (note: only 0.5% of reads are mapped to the target CCHFV genome previously specified).

SRR10769498.all.bam

757773 reads, size:63316044 bytes, created 2022-08-17 13:50:33

Mapping stats:



Number of alignments in various mapping quality (MAPQ) intervals and number of unmapped sequences.

	Number	Percentage
MAPQ >= 30	3013.0	0.4
MAPQ < 30	407.0	0.1
MAPQ < 20	0.0	0.0
MAPQ < 10	287.0	0.0
MAPQ < 3	0.0	0.0
Unmapped	754066.0	99.5
Total	757773.0	100.0

Number of alignments in various mapping quality (MAPQ) intervals and number of unmapped sequences.

- At this point, you may download the aligned reads or use them for further analysis, as shown in “**step 3c**”.

aniewiad1 / home / +TickTutorial / CCHFV_Align (8 items)

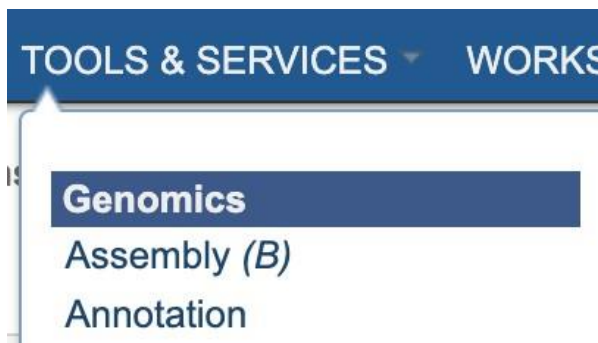
FastqUtils Job Result

Job ID	6987727
Start time	3/22/22, 8:48 AM
End time	3/22/22, 8:55 AM
Run time	6m9s
Parameters	

Name	Size	Owner	Members	Created
Parent folder				
SRR10769498.aligned.bam	52.7 kB	me	Only me	3/22/22, 8:54 AM
SRR10769498.aligned.bam.bai	96 B	me	Only me	3/22/22, 8:54 AM
SRR10769498.aligned.fq.gz	107.7 kB	me	Only me	3/22/22, 8:54 AM
SRR10769498.all.bam.samstat.html	232.2 kB	me	Only me	3/22/22, 8:55 AM
SRR10769498.unmapped.fq.gz	74.5 MB	me	Only me	3/22/22, 8:55 AM

Step 3c) de novo sequence assembly ([Genome Assembly Service, Documentation](#))

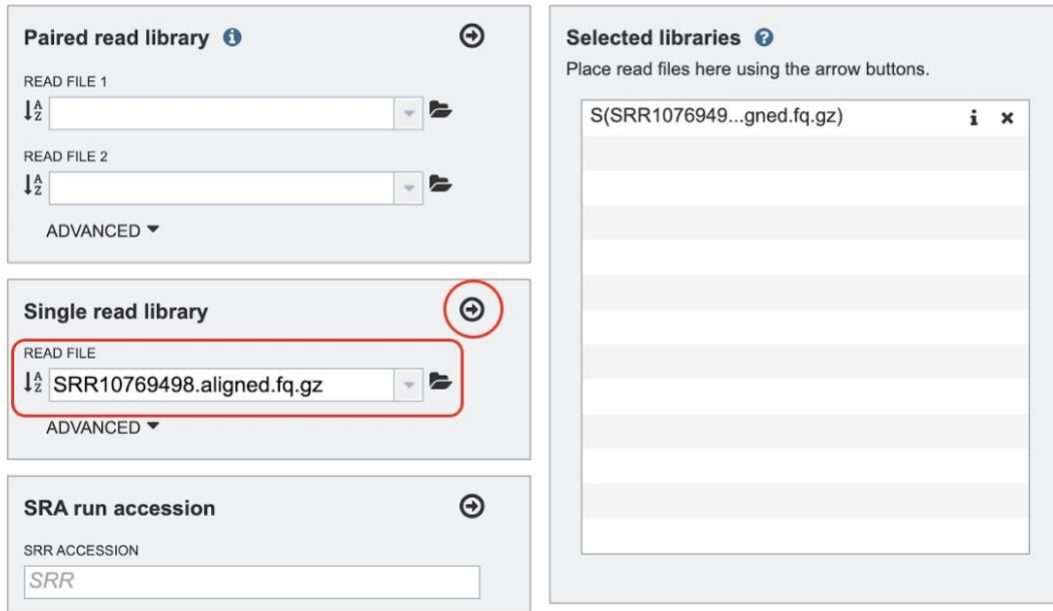
- To assemble the aligned reads from the previous step, navigate to the “Assembly” tool beneath the “Genomics” header underneath the “Tools & Services” tab.



- Select or upload as appropriate the “SRR10769498.aligned.fq.gz” from the previous step, under Single Read Library and click the arrow to move it to the “Selected Libraries” box.

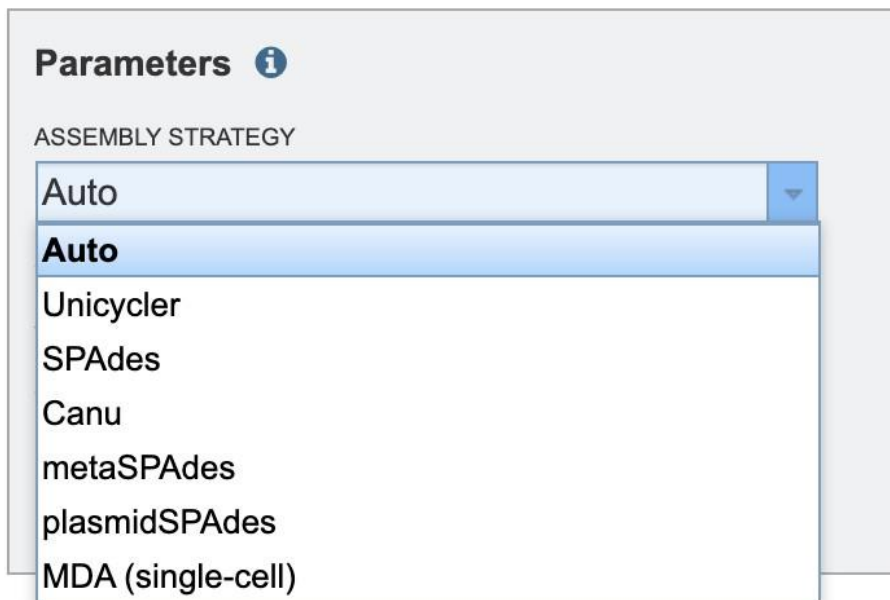
Genome Assembly

The Genome Assembly Service allows single or multiple assemblers to be invoked to compare results. The service attempts to select the best assembly. For further explanation, please see the [Genome Assembly Service Quick Reference Guide](#) and [Tutorial](#).



The screenshot shows the Genome Assembly service interface. On the left, there are three input sections: 'Paired read library' with two 'READ FILE' fields and an 'ADVANCED' dropdown; 'Single read library' with one 'READ FILE' field containing 'SRR10769498.aligned.fq.gz' and an 'ADVANCED' dropdown; and 'SRA run accession' with a text field containing 'SRR'. On the right, the 'Selected libraries' panel shows a list of read files, with the first one being 'S(SRR1076949...gned.fq.gz)'. A red circle highlights the 'Single read library' section, and a red box highlights the 'READ FILE' input field.

- Specify your desired “output folder” and “output name”.
- You may select your desired assembly strategy, or for the purposes of this exercise leave the “Auto” option selected.



The screenshot shows the 'Parameters' section of the interface. The 'ASSEMBLY STRATEGY' dropdown menu is open, showing the following options: 'Auto', 'Unicycler', 'SPAdes', 'Canu', 'metaSPAdes', 'plasmidSPAdes', and 'MDA (single-cell)'. The 'Auto' option is currently selected.

- Once all of the appropriate criteria have been specified, the “submit” button can be clicked to launch the job.

- After completion, users can view an assembly report as shown below.

aniewiad1 / home / +TickTutorial / Assembly_CCHFV (3 items)

GenomeAssembly2 Job Result

Job ID	6987734
Start time	3/22/22, 8:52 AM
End time	3/22/22, 8:52 AM
Run time	23s
Parameters	

Name	Size	Owner	Members	Created
Parent folder				
Assembly_CCHFV_assembly_report.html	17.5 kB	me	Only me	3/22/22, 8:52 AM
Assembly_CCHFV_contigs.fasta	4.7 kB	me	Only me	3/22/22, 8:52 AM
details		me	Only me	3/22/22, 8:52 AM

- Alternatively, users can download assembled contigs for further downstream analysis as shown below.

GenomeAssembly2 Job Result

Job ID	6987734
Start time	3/22/22, 8:52 AM
End time	3/22/22, 8:52 AM
Run time	23s
Parameters	

Download

Name	Size	Owner	Members	Created
Parent folder				
Assembly_CCHFV_assembly_report.html	17.5 kB	me	Only me	3/22/22, 8:52 AM
Assembly_CCHFV_contigs.fasta	4.7 kB	me	Only me	3/22/22, 8:52 AM
details		me	Only me	3/22/22, 8:52 AM

Step 3d) Blast against viral database ([BLAST](#), [Documentation](#))

Next, we will utilize the BLAST service to search the BV-BRC databases for the genomes most similar to our assembled contigs.

- Navigate to the “BLAST” tool underneath the “Genomics” header in the “TOOLS & SERVICES” tab.

Genomics

- Assembly (B)
- Annotation
- Comprehensive Genome
- Analysis (B)
- BLAST

- Select the “BLASTN” program (given that our contigs are nucleotide sequences).
- Next we can input our query either by copying and pasting contigs downloaded from the previous step (as shown below), or by directly selecting the fasta file from your workspace.

Services

BLAST ⓘ ⓘ

The BLAST service integrates the BLAST (Basic Local Alignment Search Tool) algorithms to perform searches against public or private genomes or other reference databases using DNA or protein sequence(s). For further explanation, please see [BLAST Service Quick Reference Guide and Tutorial](#).

Search program ⓘ

- BLASTN (nucleotide > nucleotide database)
- BLASTP (protein > protein database)
- BLASTX (translated nucleotide > protein database)
- tBLASTn (protein > translated nucleotide database)

Query source

- Enter sequence
- Select FASTA file
- Select feature group

```
>SRR10769498_aligned_assembly_contig_1 length 3405 coverage 59.1 normalized_cov 0.96
AGTAGTCTAGGTCACAACCATCCAGGACATCCAGGAGGTGTTGAAGTGTGGCTCAATTT
TATGAATTAGATGTCCATTTACCTTATCCCTTTTAATAAATTTCCGATGTGGTAGACCT
GTAGGTCTCCTGGCACACCATGTGTGCAACTCTGCAACTTACACACTGTGCTTGCCGATA
ACACCTTTTGCACATGCATCAGGTCAAAAAACCTTCTTCGATCCTAGGATGCAGTGTGA
TTATTTTCAGGAGGGAGTTTTTGTGGATGTTTCTTGGTCTGACAGTGTGATGGTCACAG
GACCTAAATTGAACCTTGTGCCCGCTTCAATCAAGCTACACTGCCTTTCCTGACTAGTAA
GTTCTACACACACTATGGCCTCTGTCTTGATGTATTCAACTTCCACTTGACAAACATAT
AATCTGTAAAAGGCTTTTCACATCTAATCCACAACAGGTGCAGCCAGTCCCTACACCCC
-----
```

- Next, select your desired query database. Appropriate options for this query include:
 - “Reference and representative genomes (virus)”
 - “Search within a genome group (*searches within a user-compiled dataset of viral genomes*)”
 - “Search within a taxon”
 - “Search within a selected fasta file”

Database Source

Reference and representative genomes (virus) [v]

Reference and representative genomes (bacteria, archaea)

Reference and representative genomes (virus)

Search within selected genome list

Search within selected genome group

Search within selected feature group

Search within a taxon

Search within selected fasta file

Database Type

Contigs (NT) [v]

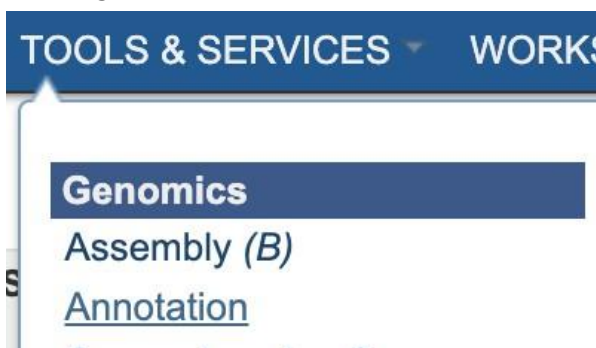
- In the example below, I have selected a saved “Genome group” that I previously compiled, containing all the complete Nairovirus M segments that are in the BV-BRC. Results are displayed as shown below.

<input type="checkbox"/>	Query ID	Genome	Subject ID	Product	Identit (%)	Query cover (%)	Subje cover (%)	Query Length	Subject Length	Score	E + value
<input type="checkbox"/>	▶ SRR10769498_aligned_asse	Crimean-Congo hemorrhagic fever ortho	NC_005300	segment	100	100.00	63.27	3405	5366	6219	0
<input type="checkbox"/>	▶ SRR10769498_aligned_asse	Crimean-Congo hemorrhagic fever ortho	KY484035	segment	100	100.00	63.27	3405	5366	6202	0
<input type="checkbox"/>	▶ SRR10769498_aligned_asse	Crimean-Congo hemorrhagic fever ortho	KX013451	segment	98	100.00	63.62	3405	5336	5864	0
<input type="checkbox"/>	▶ SRR10769498_aligned_asse	Crimean-Congo hemorrhagic fever ortho	MF287637	segment	97	100.00	63.25	3405	5368	5781	0
<input type="checkbox"/>	▶ SRR10769498_aligned_asse	Crimean-Congo hemorrhagic fever ortho	MF547416	segment	97	99.94	63.23	3405	5366	5633	0
<input type="checkbox"/>	▶ SRR10769498_aligned_asse	Crimean-Congo hemorrhagic fever ortho	KY484045	segment	96	100.00	63.28	3405	5365	5521	0
<input type="checkbox"/>	▶ SRR10769498_aligned_asse	Crimean-Congo hemorrhagic fever ortho	KY484042	segment	96	100.00	63.28	3405	5365	5499	0
<input type="checkbox"/>	▶ SRR10769498_aligned_asse	Crimean-Congo hemorrhagic fever ortho	KJ682814	segment	96	100.00	63.29	3405	5364	5493	0
<input type="checkbox"/>	▶ SRR10769498_aligned_asse	Crimean-Congo hemorrhagic fever ortho	AY900141	segment	96	100.00	67.16	3405	5055	5486	0
<input type="checkbox"/>	▶ SRR10769498_aligned_asse	Crimean-Congo hemorrhagic fever ortho	KJ682813	segment	96	100.00	63.29	3405	5364	5454	0

Step 3e) Annotate my genome ([Genome Annotation Service, Documentation](#))

Now that we have our assembled contigs, we can further characterize our viral genomes by annotating the proteins they code for. For this, we will utilize the “Genome annotation service”.


- Navigate to the “Annotation” tool underneath the “Genomics” header in the “TOOLS & SERVICES” tab.




- Upload or select your fasta formatted contig file as appropriate
- Select the desired annotation recipe; in this case “Viruses”
- Enter the appropriate Taxon name for annotation; in this case I have selected the entire *Nairoviridae* family, however users can also select CCHFV.
- Specify the appropriate output folders and names, and click “Annotate” to launch the job.


Genome Annotation

The Genome Annotation Service provides annotation of genomic features using the RAST tool kit (RASTtk) for bacteria and VIGOR4 for viruses. The service accepts a FASTA formatted contig file and an annotation recipe based on taxonomy to provide an annotated genome. For further explanation, please see the Genome Annotation Service Quick Reference Guide and Tutorial.

Parameters 


CONTIGS
 

ANNOTATION RECIPE




TAXONOMY NAME  TAXONOMY ID

MY LABEL

OUTPUT NAME


OUTPUT FOLDER
 

- Results can be viewed in a variety of ways, including “Genome View”, “CDS view”, as well as in the “Genome Browser”..




anewiad1 / home / +TickTutorial / Nairoviridae Annotation (24 items)
  

GenomeAnnotation Job Result

- Examples of the genome view and the protein list are shown below. Given that our sample only contained fragments of the genome, viewing it in the Genome Browser is not appropriate.

 Genome View
 Viruses » Negamaviricota » Ellioviricetes » Bunyavirales » Nairoviridae » **Nairoviridae Annotation**

Overview | Genome Browser | Proteins | Protein Structures | Domains and Motifs | Experiments | Interactions




 **Nairoviridae Annotation**  Edit Genomic Features 

Length: 4492bp, Contigs: 2

General Info	
Genome ID	1980415.25
Genome Name	Nairoviridae Annotation

Taxonomy Info	
Taxon ID	1980415
Superkingdom	Viruses
Kingdom	Orthomavirae
Phylum	Negamaviricota
Class	Ellioviricetes
Order	Bunyavirales
Family	Nairoviridae

	PATRIC	RefSeq
CDS	2	0

Services   

External Tools

BEI Resources

Recent PubMed Articles

• No recent articles found.

Genome View
 Viruses » Negamaviricota » Ellioviricetes » Bunyavirales » Nairoviridae » **Nairoviridae Annotation**

Overview Genome Browser **Proteins** Protein Structures Domains and Motifs Experiments Interactions

DOWNLOAD KEYWORDS ADV Search CDS PATRIC

FEATURE_TYPE x ANNOTATION x

The "Interactions" tab shows a list of protein-protein interactions, inferred using computational and laboratory methods.

Genome Name	Accession	Feature Type	BRC ID	RefSeq Locus Tag	Start	End	Strand	Length (NA)	Gene Symbol	Product
<input type="checkbox"/> Nairoviridae Annotation	1980415.25.cor	CDS	fig 1980415.25.		1	3405	-	3401	GPC	putative Pre-glycoprotein polyprotein C
<input type="checkbox"/> Nairoviridae Annotation	1980415.25.cor	CDS	fig 1980415.25.		1	951	+	951	GPC	putative Pre-glycoprotein polyprotein C

THANKS FOR FOLLOWING ALONG, AND PLEASE CONTACT US WITH YOUR QUESTIONS AT BV-BRC.ORG!

BV-BRC BETA ORGANISMS SEARCHES TOOLS & SERVICES WORKSPACES HELP ABOUT

BACTERIAL AND VIRAL BIOINFORMATICS RESOURCE CENTER

Welcome to the Bacterial and Viral Bioinformatics Resource Center (BV-BRC), an information system for bacterial and viral infectious diseases. [Learn more about BV-BRC](#). Introductory material for training is provided in the sections at the bottom of this page.

This is the *Beta Version* of the website, designed to allow users to explore, try out features, and

- Quick Start
- Quick References
- Tutorials
- Common Tasks
- CLI Tutorial
- Webinars
- Instructional Videos
- Workshops
- Contact Us**