BACTERIAL AND VIRAL BIOINFORMATICS RESOURCE CENTER

OVERVIEW

AUGUST 23, 2022

AGENDA

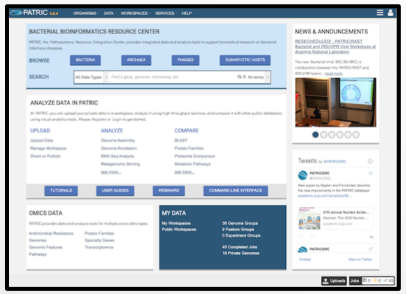
- BV-BRC Overview Presentation – 10 Mins
- BV-BRC Website Tour – 20 Mins
- Case Study: Crimean-Congo Hemorrhagic Fever comparative genomics – 25 Mins
- Q and A / Discussion – 5 Mins

BV-BRC MISSION

- **To provide a one-stop-shop for data about human pathogenic bacteria and viruses integrated into a scalable resource for comparative bioinformatics, large-scale integrative data mining, multi-scale systems biology exploration, and machine learning that will support the evolving needs of the infectious disease research community and serve as a paradigm for research data resources**

NIAID BIOINFORMATIC RESOURCE CENTERS (2019-2024)

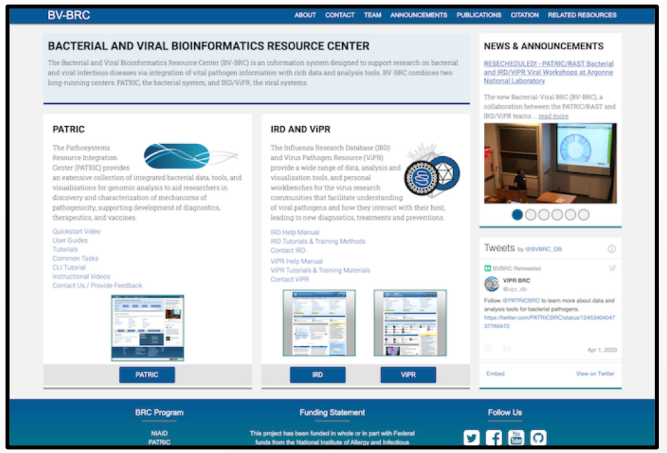
Viruses:
IRD/ViPR
Northrop, JCVI



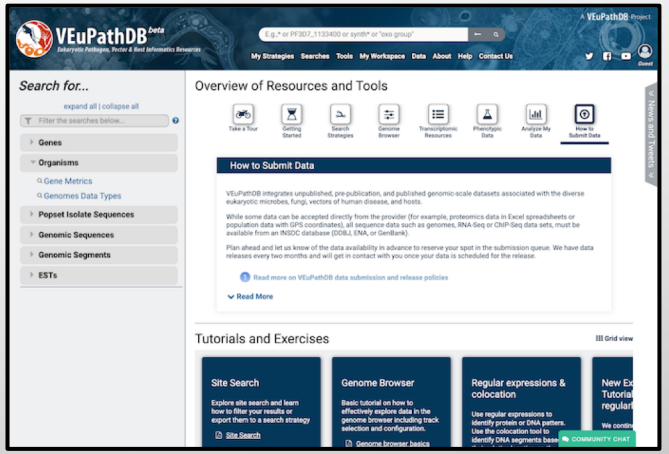
Bacteria:
PATRIC
UChicago, VT

Eukaryotes:
EuPathDB
UPenn, UGA

Vectors:
Notre Dame
EBI/EMBL



Bacteria and Viruses:
BV-BRC
UChicago, JCVI, UVA



Eukaryotes and Vectors:
VEuPathDB
UPenn, UGA
Notre Dame, EBI/EMBL

KEY FEATURES

- Supporting diverse **bacterial and viral communities**
- **Unified data model** and **real-time data integration** processes
- Efficient data management to **support exponential growth** of pathogen data
- **Consistent and accurate genome annotations** and other derived data types
- Automated and manual **curation of metadata**
- Modular, interoperable, and compute-intensive **high-throughput data analysis services**
- Integrative **analysis of multi-omics systems biology data** using visual analytics tools
- Phylogenomic/epidemiological analysis for **rapid outbreak response**
- **Explainable AI / machine learning** based tools
- **Private user workspaces** for data analysis, sharing, and publishing
- **Programmatic and batch access** via APIs, Command-line Interface (CLI), and FTP
- **Outreach and education** using online tools and in-person workshops

BV-BRC
Bacterial and Viral
Bioinformatics Resource Center

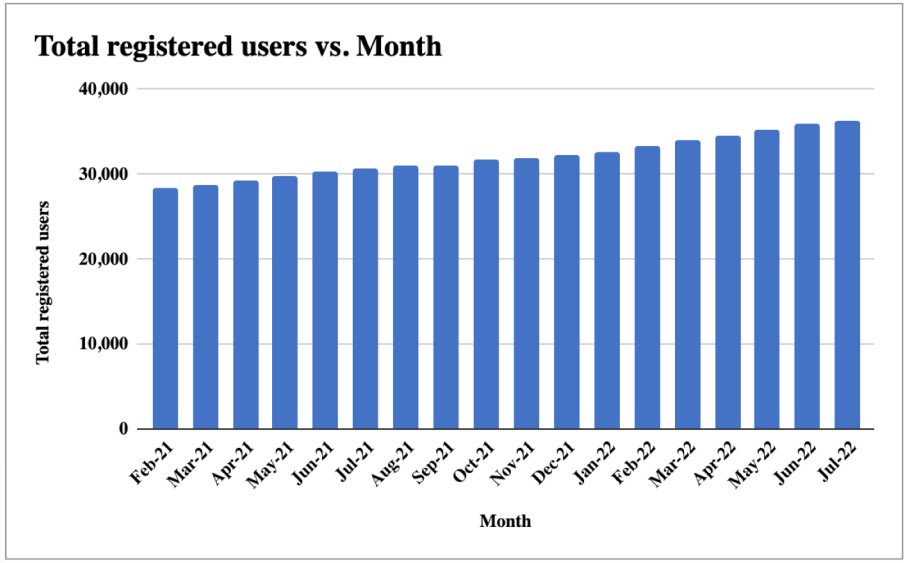
INTERNATIONAL USER COMMUNITY

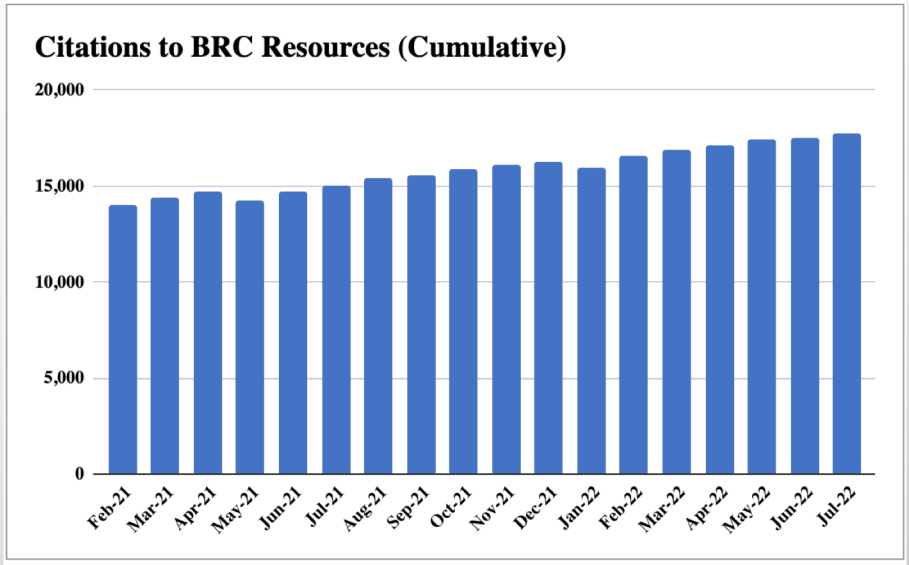
BV-BRC User Community

- **32,000+ Registered users**
- **40,000+ Unique visitors per month**
- **16,000+ Citations**

Diverse User Base

- Basic ID researchers / biologists
  - Academia, Govt labs, Non-profit, Industry, Pharma
- Clinical researchers
  - Clinical microbiologists, physician scientists
- Epidemiologists
- Bioinformatician and computational biologists
- Data scientists
- Other data resource providers





BV-BRC
Bacterial and Viral
Bioinformatics Resource Center

BV-BRC DATA

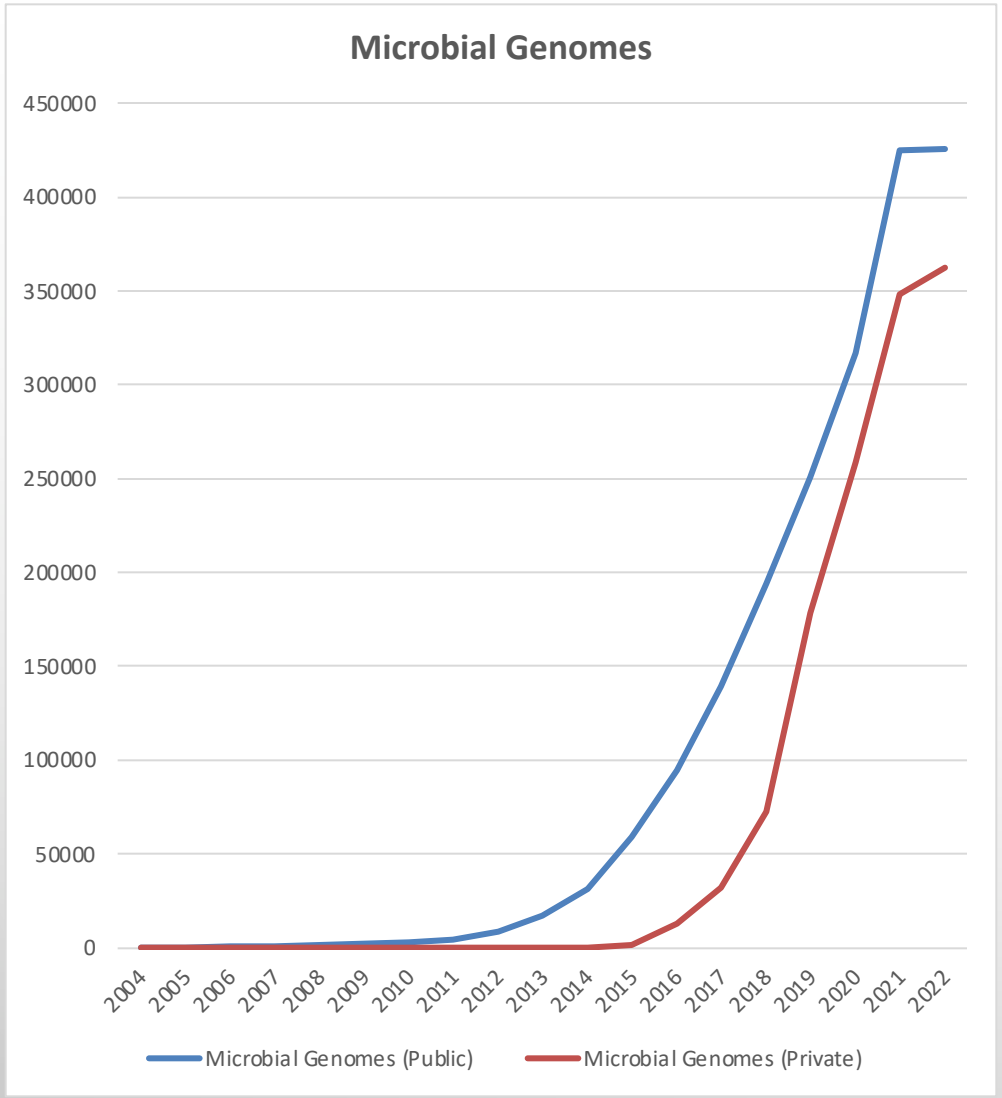
- **Total 9.0M+ microbial and viral genomes**
  - 8.0M+ viral genomes, including 6.2M SARS-CoV-2
  - 17K+ bacteriophage genomes
  - 10 eukaryotic host genomes
- **Uniform annotations across all genomes using VIGOR4 for viruses**
  - **111M+ genes/proteins/mat peptides,** and other genomic features
  - Protein functions, GO terms, EC numbers, protein families, domains and motifs
- **Curated metadata**
  - Host, collection date, geo location, other clinical and env metadata
- **220K+ experimentally characterized viral epitopes**
- **11.5K+ viral protein structures**
- **900+ curated Omics datasets**
  - 6500+ curated biosets from differential expression analysis
  - 111 host response datasets from bacterial and viral infections
- **56M+ microbial and host-pathogen PPIs**
- **1.6M surveillance and 47K serology records**

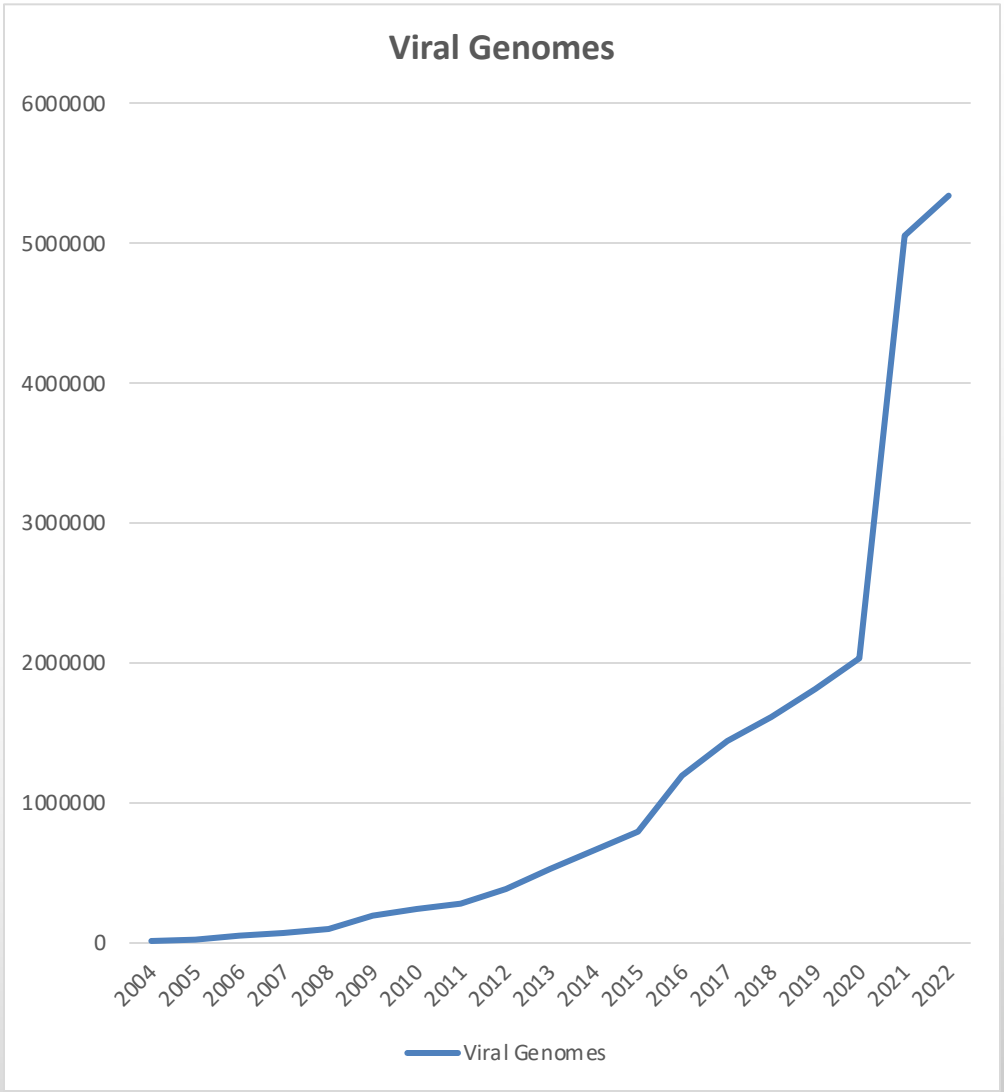
Data Sources

- Data aggregated from **Public Archives**
- Data from **NIAID Programs**, e.g., CEIRS, SysBio, functional and structural genomics programs
- **Derived Data** from internal pipelines
- **Data Integration**
- **Private Data** in user workspace

BV-BRC
Bacterial and Viral
Bioinformatics Resource Center

GENOME GROWTH





BRC
NIAID Bioinformatics Resource Centers

BV-BRC
Bacterial and Viral Bioinformatics Resource Center

BV-BRC SERVICES

Genomics Analysis
- Genome Assembly
- Genome Annotation
- Comprehensive Genome Analysis
- SARS-CoV-2 Genome Assembly and Annotation
- Similar Genome Finder
- BLAST
- Phylogenetic Tree
- Genome Alignment
- Primer Design
- Variation Analysis

Gene/Protein Analysis
- MSA and SNP Analysis
- Gene Tree
- Meta-CATS
- Proteome Comparison

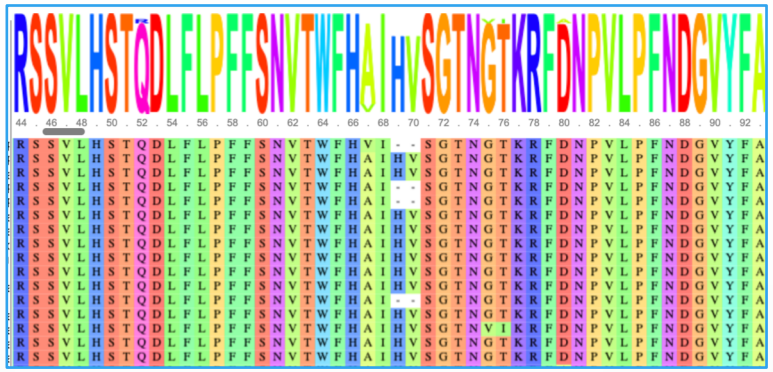
Metagenomics Analysis
- Taxonomic Classification
- Metagenomic Binning
- Metagenomic Read Mapping

Transcriptomics Analysis
- Expression Import
- RNA-Seq Analysis

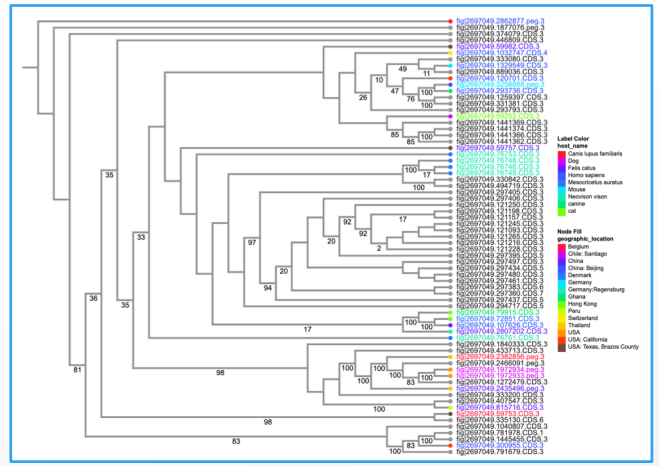
Utility Tools
- ID Mapper
- Fastq Utilities

Services in blue, support bacterial AND viral use cases

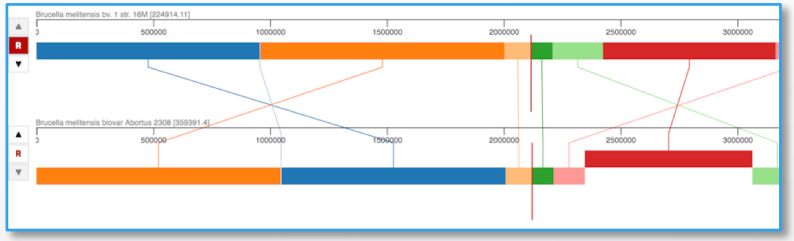
Multiple Sequence Alignment



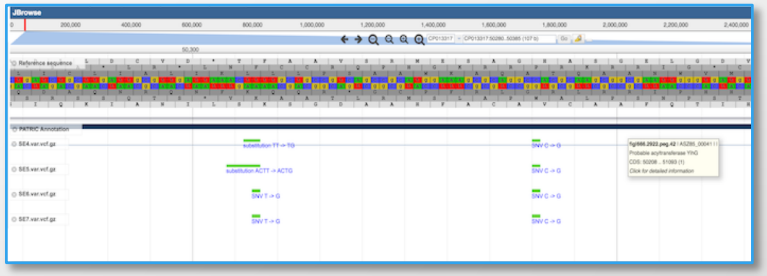
Phylogenetic Analysis



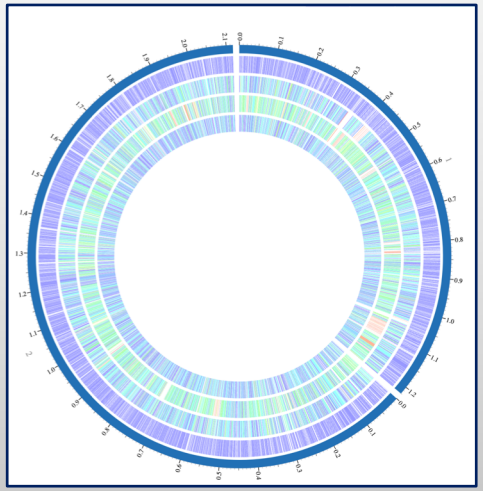
Whole Genome Alignment



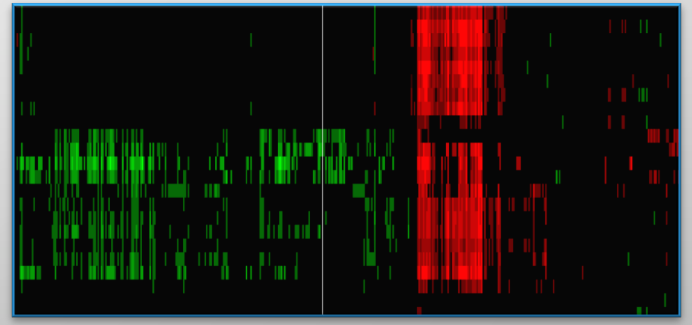
Variation Analysis



Proteome Comparison



Transcriptomic Analysis



NEW TOOLS FOR IRD/VIPR USERS

- SARS-COV-2 TOOLS
  - SARS-COV-2 GENOME ASSEMBLY AND ANNOTATION
  - SARS-COV-2 VARIANT TRACKER
- NGS SERVICES:
  - FASTQ UTILITIES
  - DE NOVO ASSEMBLY
- METAGENOMIC
  - TAXONOMIC CLASSIFICATION
  - METAGENOMIC BINNING
- TRANSCRIPTOMICS
  - RNA-SEQ ANALYSIS
  - EXPRESSION IMPORT
- 3-D PROTEIN STRUCTURE VIEWER (NEW)

Services

SARS-CoV-2 Genome Assembly and Annotation

The SARS-CoV-2 Genome Assembly and Annotation Service provides a streamlined "meta-service" that accepts raw reads and performs genome assembly, annotation, and variation analysis. For further explanation, please see the SARS-CoV-2 Genome Assembly and Annotation Service Quick Reference Guide and Tutorial.

Start With:

READ FILE ASSEMBLED CONTIGS

Input File

PAIRED READ LIBRARY

READ FILE 1

READ FILE 2

Select a platform...

SINGLE READ LIBRARY

READ FILE

Select a platform...

SRA RUN ACCESSION

SRR

Selected libraries

Place read files here using the arrow buttons.

Parameters

STRATEGY

Auto

TAXONOMY NAME

Severe acute respiratory syndrome coror

TAXONOMY ID

2697049

MY LABEL

My identifier123

OUTPUT FOLDER

OUTPUT NAME

Taxonomy + My Label

Reset Submit

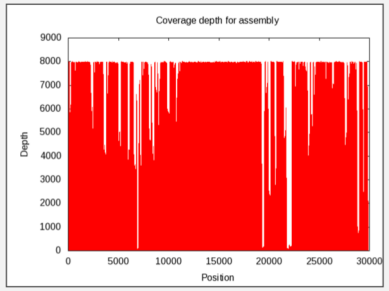
SARS2 Analysis Report for Severe acute respiratory syndrome coronavirus 2 SRR15315387

Protein SNPs from assembly

Gene	SNPs
S	T19R T95I Y144S Y145N L452R T478K D614G P681R
ORF1ab	F924F F924F A1306S A1306S P2046L P2046L P2287S P2287S A2618T A2618T N2829N N2829N D2907D D2907D V2930L V2930L T3255I T3255I T3646A T3646A V3689V V3689V P314L G662S P1000L A1918V N2328N
ORF3a	S26L E239Q
M	G79G I82T
ORF7a	G38G V82A T120I
ORF8	G77S D119V F120L
N	D63G R203M G215C D377Y
ORF7b	T40I

Assembly statistics

Job ID	assembly_32331
Elapsed time	954 sec
User-selected Recipe	auto
Platform	ILLUMINA
Library type	illumina
Chosen recipe	onecodex
Number of Ns in assembly	422
Number of N blocks in assembly	4



Variation data:

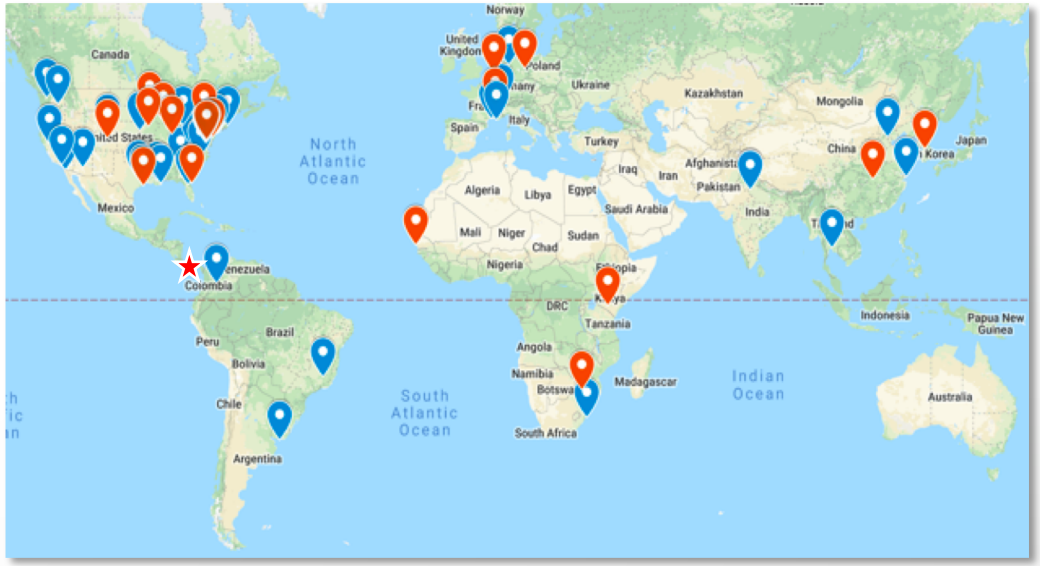
REGION	POS	REF	ALT	REF_DP	REF_RV	REF_QUAL	ALT_DP	ALT_RV	ALT_QUAL	ALT_FREQ	TOTAL_DP
MN908947.3	210	G	T	111	42	37	5406	2031	38	0.979703	5518
MN908947.3	241	C	T	2	1	55	5347	1964	39	0.999626	5349
MN908947.3	3037	C	T	2	0	37	7672	3479	36	0.999349	7677
MN908947.3	3037	C	T	2	0	37	7672	3479	36	0.999349	7677
MN908947.3	4181	G	T	218	105	37	7381	2920	37	0.970673	7604
MN908947.3	4181	G	T	218	105	37	7381	2920	37	0.970673	7604
MN908947.3	6402	C	T	171	58	39	7013	2846	37	0.97579	7187
MN908947.3	6402	C	T	171	58	39	7013	2846	37	0.97579	7187
MN908947.3	7124	C	T	89	36	37	2881	1232	37	0.970034	2970

Features annotated:

ID	Start	Strand	Length	Function
fig|2697049.3949155.CDS.1	255	+	6594	putative ORF1a polyprotein, N-terminal
fig|2697049.3949155.CDS.2	255	+	6594	putative ORF1ab polyprotein, N-terminal
fig|2697049.3949155.mat_peptide.1	255	+	540	leader protein
fig|2697049.3949155.mat_peptide.2	795	+	1914	nsp2
fig|2697049.3949155.mat_peptide.3	2709	+	4140	putative nsp3, N-terminal

BV-BRC: OUTREACH

- Virtual outreach
  - Online MOOC training courses
    - **5000+ enrolled in Bacterial Bioinformatics (Coursera)**
  - Half-day virtual mini workshops / webinar series
    - **SARS-CoV-2 Outbreak Response (Aug-Dec 2020)**
    - **Respiratory Pathogens (Feb-Jun 2021)**
    - **Tick-borne pathogens (March 2022)**
  - Instructional videos addressing common user questions
  - Social media: Twitter, Facebook, YouTube, and Reddit
- Tutorials and user guides for newly developed tools
- Rapid response to user questions / helpdesk
- Attendance at national meetings
  - Presentations, booths and workshops (ASM Microbe, ASV, GVN)
- Strategic workshops (2-3 per year)
- Publications on new resource functionality



Locations of the 90+ workshops conducted by PATRIC (Blue) and ViPR/IRD (Red) over past 10 years

BRC
NIAID Bioinformatics Resource Centers

BV-BRC Website: https://www.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 designed to support research on bacterial and viral infectious diseases. **Learn more about BV-BRC**. Introductory material for transitioning PATRIC and IRD/ViPR users 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 **provide feedback to the BV-BRC team**.

SEARCH All Data Types | *Find a gene, genome, microarray, etc* | All terms

BROWSE BACTERIA | ARCHAEA | VIRUSES | EUKARYOTIC HOSTS

ANALYZE DATA IN BV-BRC

Upload and analyze your data in the private workspace.

SEARCH	ANALYZE	MANAGE DATA	BATCH ACCESS
Taxa	Assembly	Access Private Workspace	Command-Line Interface
Genomes	Annotation	Save Search Results	Data API
Proteins	BLAST	Upload Data	FTP
Specialty Genes	MSA	Access Analysis Jobs	
All Searches...	All Tools & Services...	Share and Publish	

QUICK START | REFERENCE GUIDES | TUTORIALS

Information for PATRIC Users

The new BV-BRC system is built on the PATRIC system. ***Viral*** data and analysis tools and services from IRD/ViPR have now been integrated into the resource.

BV-BRC Data and Functionality Overview

Information for IRD/ViPR Users

The new BV-BRC system is built on the PATRIC Bacterial BRC system and incorporates data and tools from IRD and ViPR. The following materials are designed to aid IRD and ViPR users in using BV-BRC.

IRD/ViPR to BV-BRC Mapping
Common Tasks in BV-BRC

NEWS & ANNOUNCEMENTS

BV-BRC Beta Release

The Bacterial and Viral Bioinformatic Resource Center (BV-BRC) has released its new Beta website for researchers to use. ... read more

BV-BRC

BVBRC
A Twitter list by @BVBRC_DB

VIPR BRC @vipr_db
Exciting news! Hopefully this will be efective.
engadget.com/moderna-mrna-h…

Moderna begins early-stage tr…
Moderna has begun early-stage…
engadget.com

21h

PATRICBRC @PATRICBRC
Syphilis - USA:
Syphilis cases have steadily become more prevalent in Pennsylvania over the last two decades, data from the Pennsylvania Department of Health states.

Embed | View on Twitter

TICK-BORN VIRUSES IN BUNYAVIRALES ORDER

Virus View

Viruses (8012774 Genomes)

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

Virus Data Summary

Families	74
Genera	1685
Species	21452
Genomes / Segments	7996562
Protein Coding Genes (CDS)	144923049
Mature Peptides	140971003
3D Protein Structures (PDB)	11502

Virus Families

Single-Stranded Positive-Sense RNA

Caliciviridae

Coronaviridae

Flaviviridae

Hepeviridae

Picornaviridae

Togaviridae

Single-Stranded Negative-Sense RNA

Bunyavirales

Filoviridae

Paramyxoviridae

Orthomyxoviridae

Pneumoviridae

Rhabdoviridae

Double-Stranded RNA

Reoviridae

Single-Stranded DNA

Parvoviridae

Partially Double-Stranded DNA

Hepadnaviridae

Double-Stranded DNA

Adenoviridae

Asfarviridae

Herpesviridae

Polyomaviridae

Poxviridae

Featured Viruses

Bacteriophages

Dengue virus

Ebolavirus

Enterovirus

Hepatitis C virus

Influenza A virus

Lassa Virus

Monkeypox Virus

SARS-CoV-2

Zika Virus

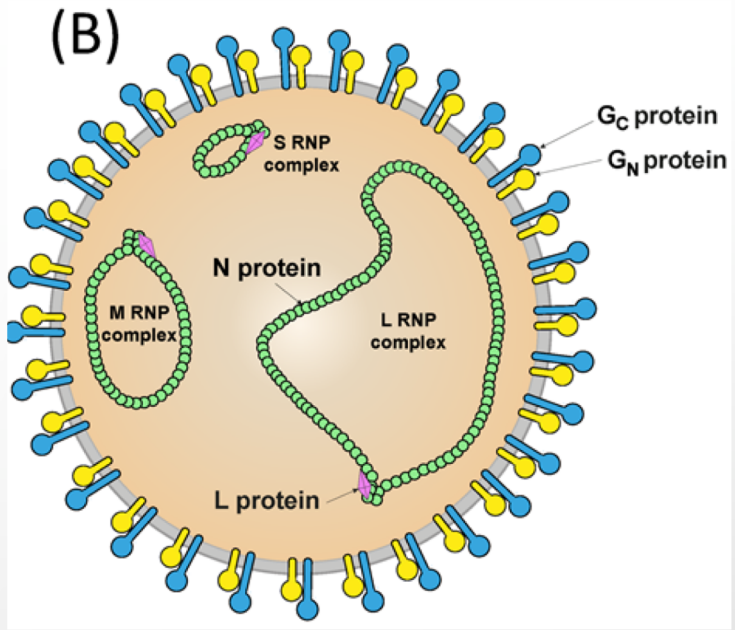
BRC NIAID Bioinformatics Resource Centers

BV-BRC

NAIROVIRIDAE SPECIES THAT INFECT HUMANS (**BOLD**)

Abu Hammad virus
Artashat orthonairovirus
Bandia virus
Beiji nairovirus
Burana virus
Chim orthonairovirus
Clo Mor viru
Crimean-Congo hemorrhagic fever orthonairovirus
Dera Ghazi Khan orthonairovirus
Dugbe orthonairovirus
Estero Real virus
Farallon virus
Geran virus
Grotenhout virus
Hazara virus
Hughes orthonairovirus
Kasokero orthonairovirus
Kupe virus
Nairobi sheep disease virus
Nairovirus sp. Hc652
Nayun tick nairovirus
Norway nairovirus 1
Pacific coast tick nairovirus
Paramushir virus
Punta Salinas virus
Pustyn virus
Qalyub orthonairovirus
Rondonia orthonairovirus
Sakhalin orthonairovirus
Saphire II virus
Soldado virus
Songling virus
South Bay virus
Tacheng Tick Virus
Taggert virus
Tamdy orthonairovirus
Tillamook virus
Tofla virus
Vinegar Hill virus
Yezo virus

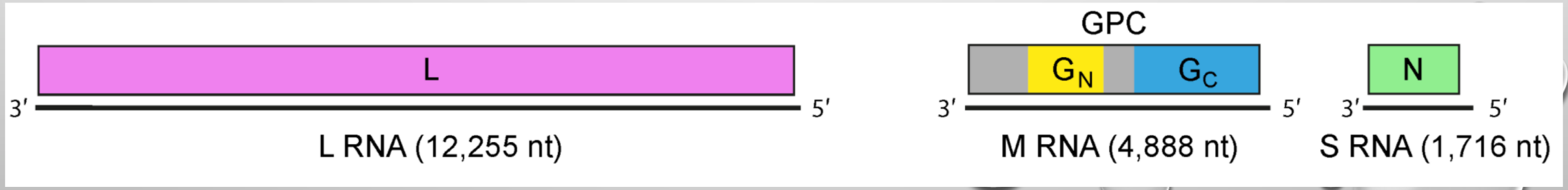
ORTHONAIROVIRUS GENOME & PROTEOME



Nucleoprotein (N): Structural virion protein (60–68 kD). Component of the RNP inside virions. Oligomerizes and encapsidates orthonairoviral genomic segments. Functions as an exoribonuclease.

Glycoprotein (GP): Structural virion protein consisting of two subunits (G_N 30–45 kD, G_C 72–84 kD). Produced via proteolytic cleavage from the GPC precursor. Inserts into virion membranes as GP spikes composed of G_N and G_C. As a putative class I fusion protein, GP mediates cell-surface and internal receptor binding, virion-cell membrane fusion and, thereby cell entry.

Large protein (L): Structural virion protein (250–450 kD) with RdRP, helicase, and endoribonuclease domains. Component of the RNP inside virions. Oligomerizes and mediates transcription and replication of orthonairoviral RNA segments. Mediates cap-snatching for viral mRNA capping.



THREE USE CASES

1) FINDING SPECIFIC MUTATION IN CRIMEAN CONGO HEMORRHAGIC FEVER VIRUS (CCHF) REPORTED IN THE LITERATURE
   - SEARCH, ASSEMBLE, AND CURATE RELEVANT DATASETS
   - CREATE MSA TO VERIFY AND LOOK FOR OTHER GENOMES
   - VISUALIZE RELATED DATA (3D PROTEIN STRUCTURES)

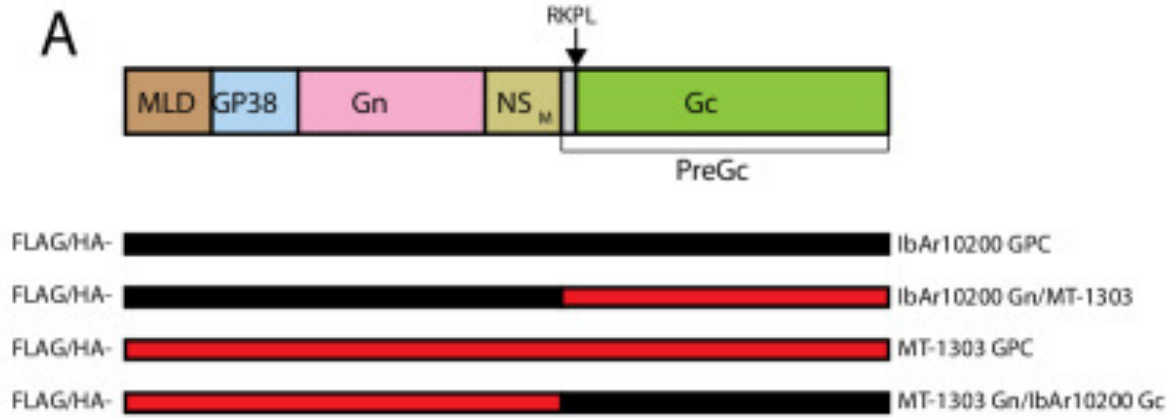
2) IDENTIFY SEQUENCE VARIATIONS IN CCHF THAT CORRELATE WITH HOST RANGE AND GEOGRAPHIC LOCATION
   - SEARCH, ASSEMBLE, AND CURATE DATASETS
   - CREATE MSA TO VERIFY AND LOOK FOR OTHER GENOMES
   - PHYLOGENETIC ANALYSIS TO REVEAL CORRELATION PATTERNS
   - USE METACATS TO SEARCH FOR HOST- OR LOCATION-SPECIFIC SITES

3) ISOLATION AND CHARACTERIZATION OF AN "UNKNOWN" NAIROVIRUS.
   - SRA FILE ANALYSIS (TAXONOMIC CLASSIFIER)
   - READ QC AND MAPPING (FASTQ UTILS)
   - *DE NOVO* SEQUENCE ASSEMBLY (ASSEMBLY)
   - BLAST AGAINST VIRAL DATABASE (BLAST)
   - ANNOTATE MY GENOME (ANNOTATION)

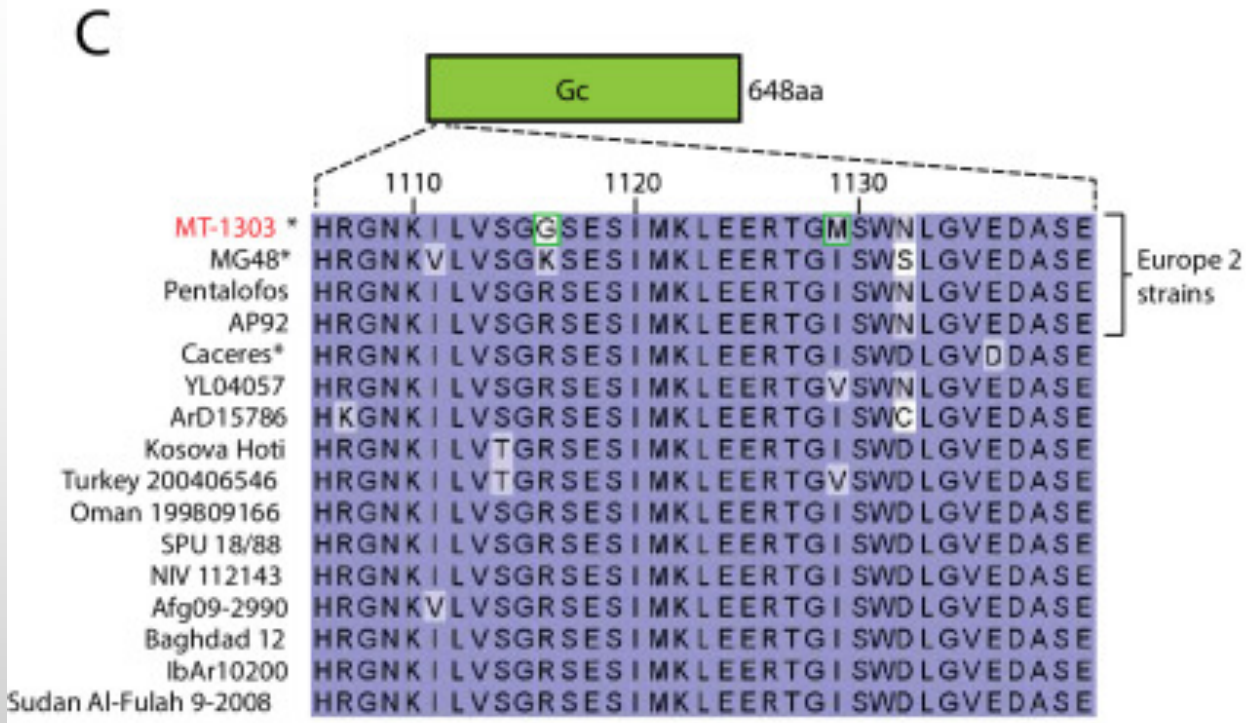
› Elife. 2020 Oct 21;9:e50999. doi: 10.7554/eLife.50999.

A single mutation in Crimean-Congo hemorrhagic fever virus discovered in ticks impairs infectivity in human cells

Brian L Hua [1], Florine Em Scholte [1], Valerie Ohlendorf [2 3], Anne Kopp [2 3], Marco Marklewitz [2 3], Christian Drosten [2 3], Stuart T Nichol [1], Christina Spiropoulou [# 1], Sandra Junglen [# 2 3], Éric Bergeron [# 1]



- single amino acid change appears to make the virus less able to infect human cells
- mutation prevents viral-cell (human) fusion step
- may explain why this strain and others in the Europe 2 group do not cause severe human disease



Strain names:

ref seq: IbAr10200

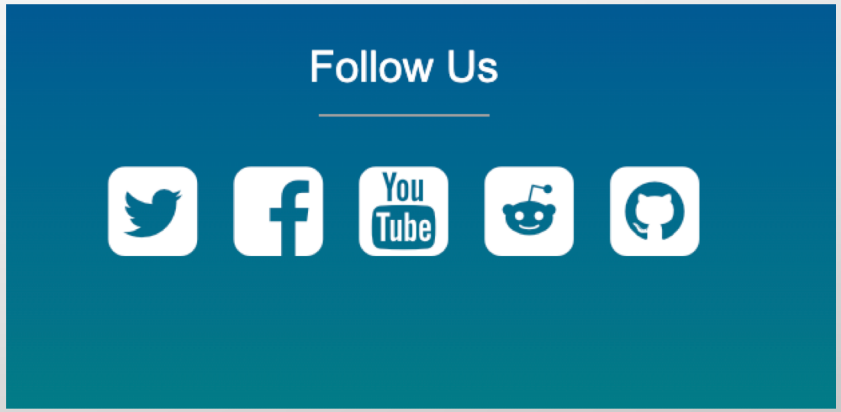
mutant: Malko Tarnovo-BG2012-T1303 (MT-1303)

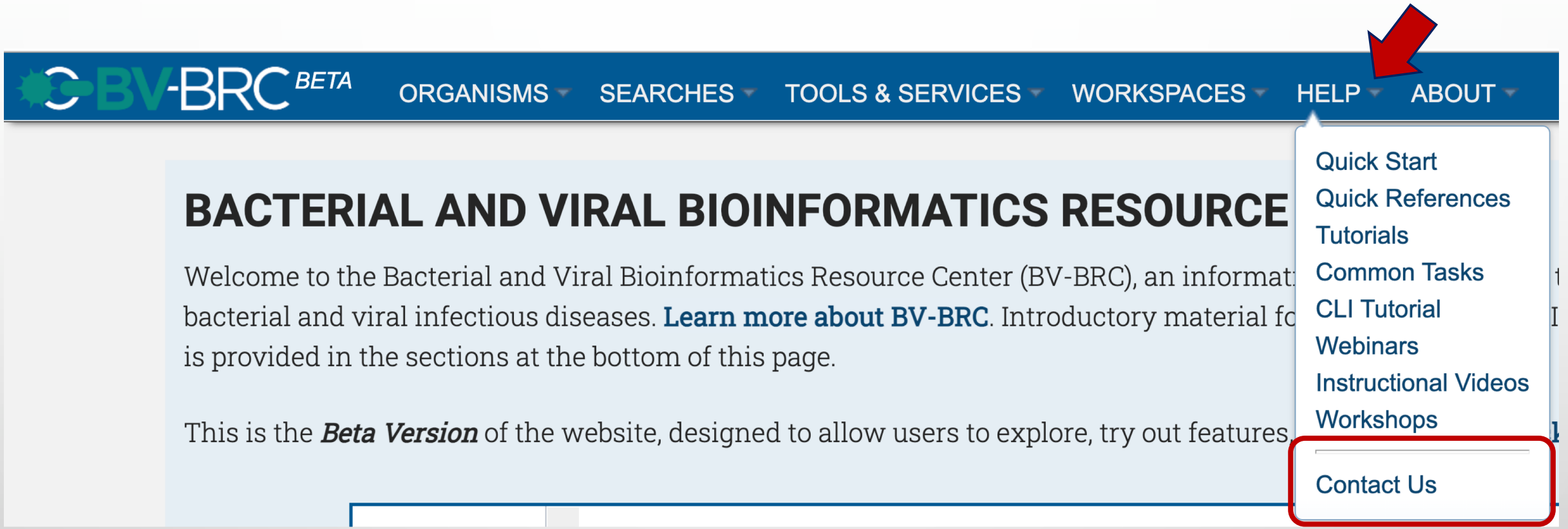
Mutation:

R1116G

WHERE CAN I FIND OUT MORE?

- INSTRUCTIONAL VIDEOS ON YOUTUBE
- TUTORIALS AND USER GUIDES
- HELPDESK
- SOCIAL MEDIA
  - TWITTER @BVBRC_DB
  - FACEBOOK @BVBRC19
  - REDDIT R/BRC_USERS




BV-BRC BETA
ORGANISMS
SEARCHES
TOOLS & SERVICES
WORKSPACES
HELP
ABOUT
Quick Start
Quick References
Tutorials
Common Tasks
CLI Tutorial
Webinars
Instructional Videos
Workshops
Contact Us
BACTERIAL AND VIRAL BIOINFORMATICS RESOURCE
Welcome to the Bacterial and Viral Bioinformatics Resource Center (BV-BRC), an informat
bacterial and viral infectious diseases. Learn more about BV-BRC. Introductory material fo
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,