

# BV-BRC Test Report

## A6. Service – Homology / BLAST – Short Peptide Search

<b>Item to test</b>	Homology / BLAST Service using short peptide search
<b>URL</b>	<a href="https://www.bv-brc.org/app/Homology">https://www.bv-brc.org/app/Homology</a>
<b>Prerequisites</b>	Short peptide query sequence
<b>References</b>	<a href="https://www.bv-brc.org/docs/quick_references/services/blast.html">https://www.bv-brc.org/docs/quick_references/services/blast.html</a> <a href="https://www.bv-brc.org/docs/tutorial/blast/blast.html">https://www.bv-brc.org/docs/tutorial/blast/blast.html</a>
<b>Tester(s)</b>	Zach Wallace, Maulik Shukla
<b>Test date</b>	06-May-2022 (follow-up from original test)
<b>Test result</b>	<b>Passed</b>

### Overview

- Test the Homology / BLAST service using exemplar short peptide query.
- Test various BLAST databases, i.e. reference genome databases, taxon level databases, and genome group databases.
- For each job submitted, verify successful completion of the job, availability of the output files in the workspace, and quality of the results.
- Test result selection and action from the result table, including MSA to visualize variations.

### Test Data

<b>Dataset</b>	<b>Rational</b>	<b>Input Format</b>	<b>Input</b>
SARS-CoV-2 Spike protein – short peptide	Viral protein of interest	FASTA NA and AA sequences	GKQGNFKNLREFVFKNIDGYFK

- All test datasets and corresponding job results are available in the following public workspace:  
<https://www.bv-brc.org/workspace/BVBRC@patricbrc.org/BVBRC%20Tests/BLAST>

### Test Results

- All jobs completed successfully, without any errors.
- All jobs resulted in expected output files in corresponding job output directory.
- For each of the jobs, the results were as expected, finding similar genes, proteins or genomic sequences from the selected reference genome set, taxon level, or genome level databases.
- All test datasets and corresponding job results are available in the following public workspace:  
<https://www.bv-brc.org/workspace/BVBRC@patricbrc.org/BVBRC%20Tests/BLAST>
- Below are a series of screenshots showing successful completion of the jobs, availability of the result files in the workspace, and the result table.

- Successful completion of all jobs using sars-cov-2 spile protein short peptide sequence as query and search databases of all reference genomes, taxon level, and single genome level.

## 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

```
>record_1
GKQGNFKNLREFVFKNIDGYFK
```

### Database Source

Reference and representative genomes (virus)

### Database Type

Proteins (AA)

### Output Folder

BLAST

### Output Name

Output Name

Name must be pr

ADVANCED OPTIONS ▾

Your job has been submitted successfully. Please visit your [Jobs List](#) to check the status of your job and access the results.

Status	ID	Service	Output Name	Submit	Start	Completed
completed	7746915	Homology	Short peptide - Rep - BLASTP	5/7/22, 8:18 AM	5/7/22, 8:18 AM	5/7/22, 8:18 AM
completed	7746914	Homology	Short peptide - Taxa - BLASTP	5/7/22, 8:17 AM	5/7/22, 8:17 AM	5/7/22, 8:18 AM
completed	7746913	Homology	Short peptide - Genome Group - BLASTP	5/7/22, 8:17 AM	5/7/22, 8:17 AM	5/7/22, 8:17 AM
completed	7746910	Homology	Short peptide - taxon - BLASTP	5/7/22, 8:08 AM	5/7/22, 8:08 AM	5/7/22, 8:09 AM
completed	7746909	Homology	Short peptide - Ref - BLASTP	5/7/22, 8:04 AM	5/7/22, 8:04 AM	5/7/22, 8:04 AM
completed	7746908	Homology	Short peptide - BLASTP	5/7/22, 8:04 AM	5/7/22, 8:04 AM	5/7/22, 8:04 AM

- BLAST results in the workspace directories.

Name	Size	Owner	Members	Created
Short peptide - Rep - BLASTP	7.3 kB	me	Public	5/7/22, 8:18 AM
Short peptide - Genome Group - BLASTP	7.4 kB	me	Public	5/7/22, 8:17 AM
SARSCoV2 Representative Genomes	22 genomes	me	Public	5/7/22, 8:16 AM
Short peptide - taxon - BLASTP	7.3 kB	me	Public	5/7/22, 8:09 AM
Short peptide - Ref - BLASTP	7.3 kB	me	Public	5/7/22, 8:04 AM
Short peptide - BLASTP	7.2 kB	me	Public	5/7/22, 8:04 AM

- BLAST Result table

Query Genome ID	Subject ID	RefSeq Locus Tag	Product	Length (NT)	Length (AA)	ALN Length	Identity (%)	Query cover (%)	Subject cover (%)	Score	E value
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2	fig 2697049.1329549.CDS.3	surface glycoprotein	3813	1270	1270	100	100.00	1.73	73	1e-18
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2 SAR	fig 2697049.120701.CDS.3	surface glycoprotein	3813	1270	1270	100	100.00	1.73	73	1e-18
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2 Wuhi	fig 2697049.107626.CDS.3 GU280_gp02	surface glycoprotein	3822	1273	1273	100	100.00	1.73	73	1e-18
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2 hCoV	fig 2697049.1032747.CDS.3	surface glycoprotein	3822	1273	1273	100	100.00	1.73	73	1e-18
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2 strain	fig 2697049.333200.CDS.3	surface glycoprotein	3822	1273	1273	100	100.00	1.73	73	1e-18
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2 SAR	fig 2697049.293793.CDS.3	surface glycoprotein	3822	1273	1273	100	100.00	1.73	73	1e-18
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2 strain	fig 2697049.335130.CDS.6	surface glycoprotein	3822	1273	1273	100	100.00	1.73	73	1e-18
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2 strain	fig 2697049.374079.CDS.3	surface glycoprotein	3819	1272	1272	100	100.00	1.73	73	1e-18
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2	fig 2697049.1272479.CDS.3	surface glycoprotein	3816	1271	1271	100	100.00	1.73	73	1e-18
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2 SAR	fig 2697049.300955.CDS.3	surface glycoprotein	3813	1270	1270	100	100.00	1.73	73	1e-18

Query Genome ID	Subject ID	RefSeq Locus Tag	Product	Length (NT)	Length (AA)	ALN Length	Identity (%)	Query cover (%)	Subject cover (%)	Score	E value
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2	fig 2697049.1329549.CDS.3	surface glycoprotein	3813	1270	1270	100	100.00	1.73	73	1e-18
<input checked="" type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2 SAR	fig 2697049.120701.CDS.3	surface glycoprotein	3813	1270	1270	100	100.00	1.73	73	1e-18

Query length: 22 Subject length: 1270  
Score: 73 bits(164) Expect: 1e-18  
Identities: 22/22(100%) Positives: 22/22(100%) Gaps: 0/22(0%)

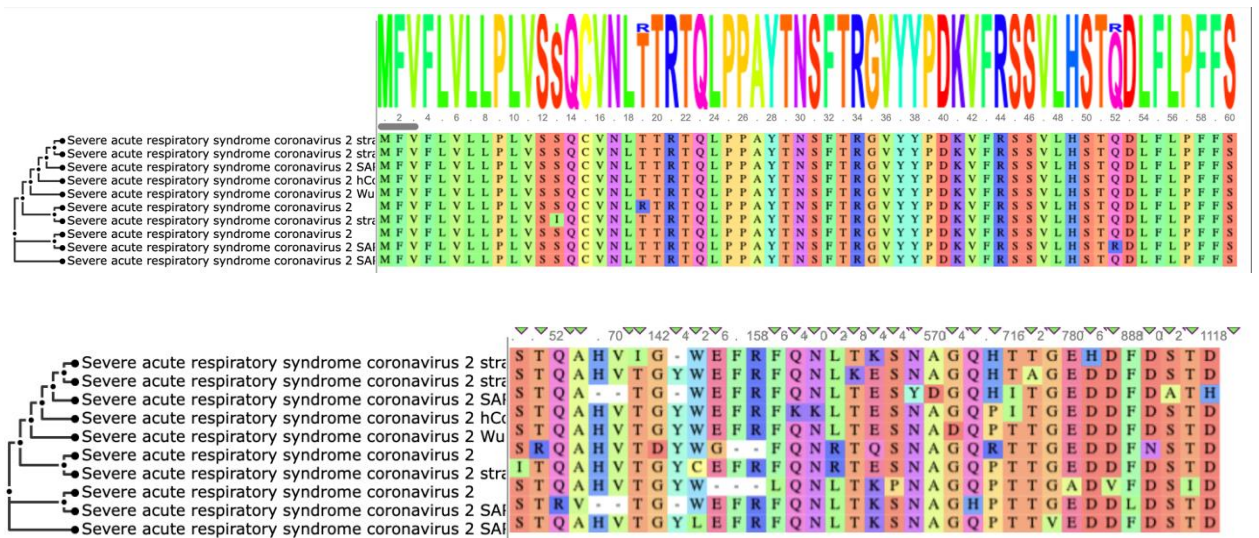
```

Query      1  GKQGNFKNLRFFVFNIDGYFK  22
           GKQGNFKNLRFFVFNIDGYFK
Sbjct    178 GKQGNFKNLRFFVFNIDGYFK  199

```

<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2 Wuhi	fig 2697049.107626.CDS.3 GU280_gp02	surface glycoprotein	3822	1273	1273	100	100.00	1.73	73	1e-18
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2 hCoV	fig 2697049.1032747.CDS.3	surface glycoprotein	3822	1273	1273	100	100.00	1.73	73	1e-18
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2 strain	fig 2697049.333200.CDS.3	surface glycoprotein	3822	1273	1273	100	100.00	1.73	73	1e-18
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2 SAR	fig 2697049.293793.CDS.3	surface glycoprotein	3822	1273	1273	100	100.00	1.73	73	1e-18
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2 strain	fig 2697049.335130.CDS.6	surface glycoprotein	3822	1273	1273	100	100.00	1.73	73	1e-18
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2 strain	fig 2697049.374079.CDS.3	surface glycoprotein	3819	1272	1272	100	100.00	1.73	73	1e-18
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2	fig 2697049.1272479.CDS.3	surface glycoprotein	3816	1271	1271	100	100.00	1.73	73	1e-18
<input type="checkbox"/>	record Severe acute respiratory syndrome coronavirus 2 SAR	fig 2697049.300955.CDS.3	surface glycoprotein	3813	1270	1270	100	100.00	1.73	73	1e-18

- Result selection and generation of MSA



## References

- [BLAST Service Quick Reference Guide](#)
- [BLAST Service Tutorial](#)