

BV-BRC Test Report

A16. Service – Gene Tree

Item to test	Gene Tree Service using bacterial and viral gene and protein sequences
URL	https://www.bv-brc.org/app/GeneTree
Prerequisites	Bacterial and viral feature groups and FASTA sequence files
References	https://www.bv-brc.org/docs/quick_references/services/genetree.html https://www.bv-brc.org/docs/tutorial/genetree/genetree.html
Tester(s)	Christian Zmasek, Yun Zhang, Maulik Shukla
Test date	08-May-2022 (follow-up from original test)
Test result	Passed

Overview

- Test the Gene Tree Service using exemplar bacterial and viral gene and protein sequences.
- Test input options, i.e., genome group, feature group, and fasta sequence files.
- Test different tree algorithms, i.e. RAXML, PHYML, and FASTTREE.
- For each job submitted, verify successful completion of the job, presence of output files in various formats, review resulting gene trees.
- View tree using interactive phylogenetic tree viewer and verify all functionality.

Test Data

Dataset	Rational	Input Format	Input
SARSCoV2 spike proteins	Viral proteins of interest	Feature group, FASTA file	SARSCoV2_Spike_protein.fasta
MTB ahpD proteins	Bacterial proteins of interest	Feature group, FASTA file	MTB_ahpD_protein.fasta

- 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/Multiple%20Sequence%20Alignment>

Test Results

- All MSA jobs completed successfully, without any errors.
- All jobs resulted in expected output files in corresponding job output directory, including alignment in afa, nexus, phy, and pir formars and consensus sequence in fasta format.
- SNPs were present as a tsv file in the expected format and the tsv viewer allow filtering and sorting of the table based on the SNP scores.

- The link to MSA viewer loaded the MSA and all functionality worked as expected, including changing node labels, and showing / hiding positions based on conservation.
- 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/Multiple%20Sequence%20Alignment>
- Below are a series of screenshots showing successful completion of the jobs, availability of the result files in the workspace, and the tree viewer.

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

SARSCoV2 Spike protein - AA.afa	x
---------------------------------	---

Alignment Parameters ?

TRIM ENDS OF ALIGNMENT THRESHOLD

REMOVE GAPPY SEQUENCES THRESHOLD

Tree Parameters ?

RAXML PHYML FASTTREE

MODEL

OUTPUT FOLDER

OUTPUT NAME

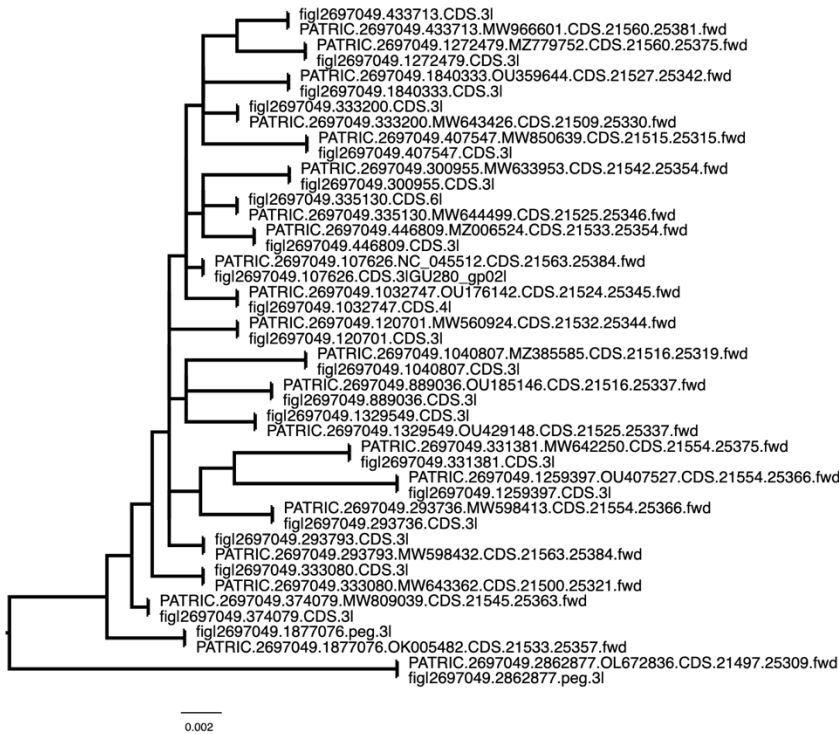
Your job has been submitted successfully. Please visit your [Jobs List](#) to check the status of your job and access the results.

Status	ID	Service	Output Name	Submit	Start	Completed
completed	7747840	GeneTree	MTB ahpD protein - AA	5/8/22, 5:26 PM	5/8/22, 5:26 PM	5/8/22, 5:27 PM
completed	7747841	GeneTree	MTB ahpD protein - NA	5/8/22, 5:26 PM	5/8/22, 5:26 PM	5/8/22, 5:26 PM
completed	7747842	GeneTree	SARSCoV2 Spike protein - NA	5/8/22, 5:26 PM	5/8/22, 5:26 PM	5/8/22, 5:27 PM
completed	7747843	GeneTree	SARSCoV2 Spike protein - AA	5/8/22, 5:28 PM	5/8/22, 5:28 PM	5/8/22, 5:29 PM
completed	7747844	GeneTree	SARSCoV2 Spike protein - AA - PHYML	5/8/22, 5:28 PM	5/8/22, 5:28 PM	5/8/22, 5:29 PM
completed	7747845	GeneTree	SARSCoV2 Spike protein - AA - FASTTREE	5/8/22, 5:28 PM	5/8/22, 5:28 PM	5/8/22, 5:29 PM
completed	7747846	GeneTree	SARSCoV2 Spike protein - AA - fasta	5/8/22, 5:29 PM	5/8/22, 5:29 PM	5/8/22, 5:30 PM
completed	7747855	GeneTree	SARSCoV2 Spike protein - AA - Alignment	5/8/22, 5:30 PM	5/8/22, 5:30 PM	5/8/22, 5:32 PM

Name	Size	Owner	Members	Created
MTB ahpD protein	48 features	me	Public	5/8/22, 5:24 PM
MTB ahpD protein - AA	5.4 kB	me	Public	5/8/22, 5:27 PM
MTB ahpD protein - NA	5.4 kB	me	Public	5/8/22, 5:26 PM
MTB_ahpD_protein.fasta	15.6 kB	me	Public	5/8/22, 5:24 PM
Parent folder			-	
SARSCoV2 Spike protein	22 features	me	Public	5/8/22, 5:24 PM
SARSCoV2 Spike protein - AA	5.4 kB	me	Public	5/8/22, 5:53 PM
SARSCoV2 Spike protein - AA - Alignment	5.9 kB	me	Public	5/8/22, 5:53 PM
SARSCoV2 Spike protein - AA - FASTTREE	5.6 kB	me	Public	5/8/22, 5:53 PM
SARSCoV2 Spike protein - AA - PHYML	5.5 kB	me	Public	5/8/22, 5:53 PM
SARSCoV2 Spike protein - AA - fasta	5.7 kB	me	Public	5/8/22, 5:53 PM
SARSCoV2 Spike protein - AA.afa	29.2 kB	me	Public	5/8/22, 5:29 PM
SARSCoV2 Spike protein - NA	5.6 kB	me	Public	5/8/22, 5:27 PM
SARSCoV2_Spike_protein.fasta	31.8 kB	me	Public	5/8/22, 5:24 PM

Name	Size	Owner	Members	Created
Parent folder			-	
SARSCoV2 Spike protein - AA - fasta.xml	42.1 kB	me	Public	5/8/22, 5:53 PM
SARSCoV2 Spike protein - AA - fasta_RAxML_log.txt	5.7 kB	me	Public	5/8/22, 5:53 PM
SARSCoV2 Spike protein - AA - fasta_RAxML_tree_rell.nwk	3.9 kB	me	Public	5/8/22, 5:53 PM
SARSCoV2 Spike protein - AA - fasta_RAxML_tree_rell.svg	22.3 kB	me	Public	5/8/22, 5:53 PM
SARSCoV2 Spike protein - AA - fasta_aligned.fasta	58.9 kB	me	Public	5/8/22, 5:53 PM
SARSCoV2 Spike protein - AA - fasta_gene_tree_report.html	34.5 kB	me	Public	5/8/22, 5:53 PM
SARSCoV2 Spike protein - AA - fasta_metadata.txt	7.2 kB	me	Public	5/8/22, 5:53 PM

Gene Tree Output



Analysis Steps

Start time Sun May 8 17:29:25 2022

Gather Sequence Data

- reading feature_group /BVBR@patricbrc.org/BVBR Tests/Gene Tree/SARSCoV2 Spike protein
- retrieving sequences for feature group /BVBR@patricbrc.org/BVBR Tests/Gene Tree/SARSCoV2 Spike protein
- number of elements = 22
- number of sequences retrieved: 22
- reading feature_protein_fasta /BVBR@patricbrc.org/BVBR Tests/Gene Tree/SARSCoV2_Spike_protein.fasta
- number of sequences retrieved: 22
- sequences need aligning

Organize Metadata

- feature metadata fields: product, accession, feature_id, genome_id genome metadata fields: species, strain, geographic_group, isolation_country, host_group, host_common_name, collection_year, subtype, lineage, clade

Align with mafft

Duration 20 seconds.

Phylogenetic Inference with RAxML

- command = raxmlHPC-PTHREADS-SSE3 -T 8 -p 12345 -m PROTCATLG -s SARSCoV2 Spike protein - AA - fasta_aligned.fa -n SARSCoV2 Spike protein - AA - fasta -f D

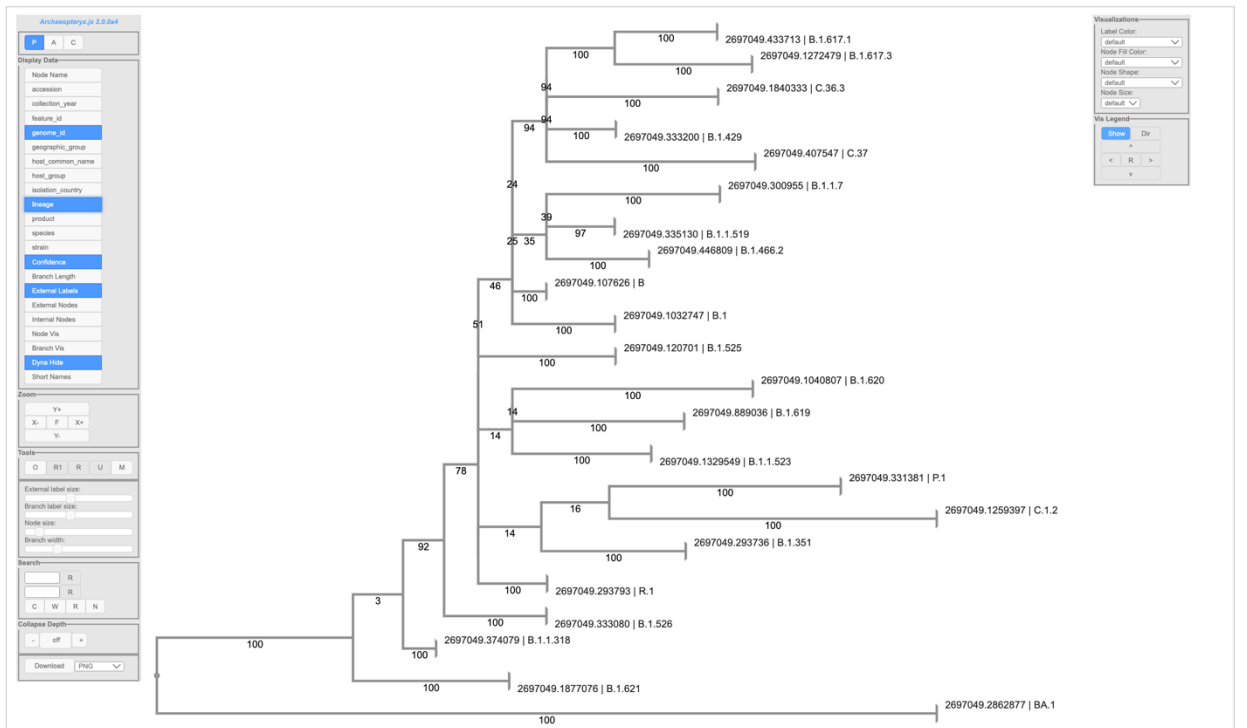
Details:

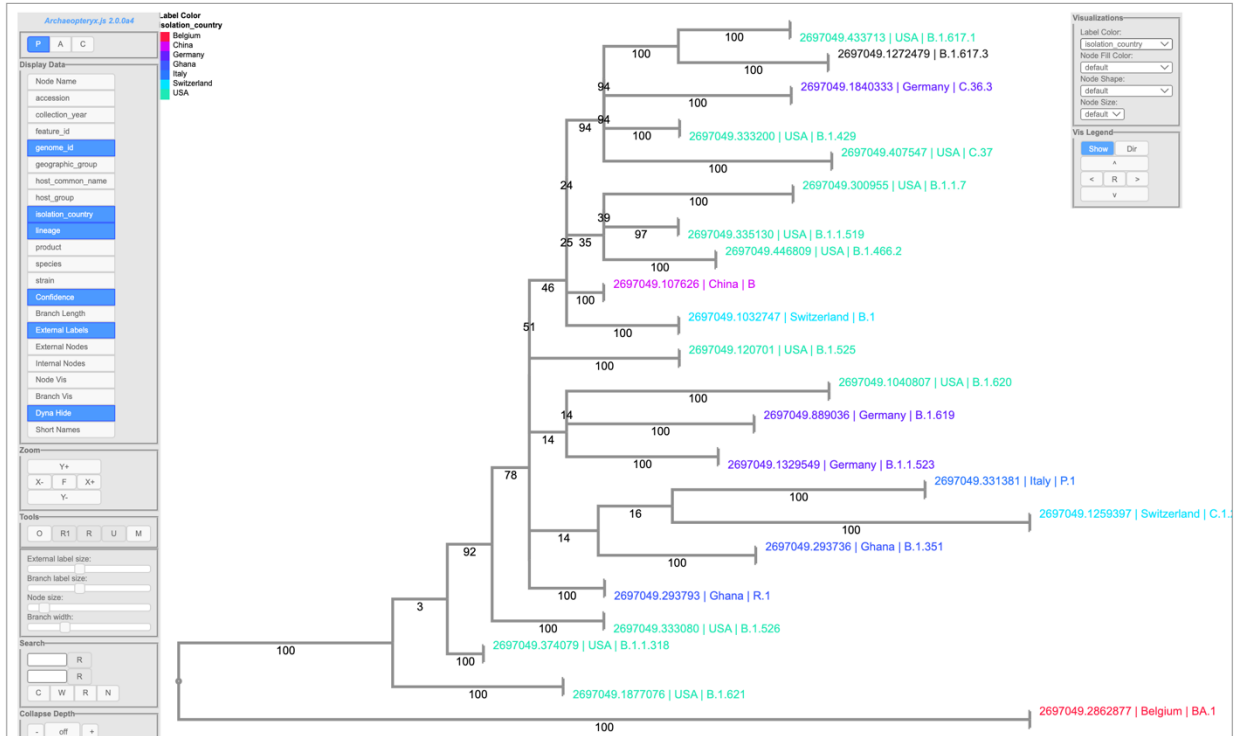
Duration 36 seconds.

Generate Tree Graphic

- figtree -graphic SVG -height 660 SARSCoV2 Spike protein - AA - fasta_RAxML_tree_rell.nex SARSCoV2 Spike protein - AA - fasta_RAxML_tree_rell.svg

Write PhyloXML





References

- [Gene Tree Service Quick Reference Guide](#)
- [Gene Tree Service Tutorial](#)