BV-BRC Test Report

A9. Service – MetaCATS

Item to test	MetaCATS Service using exemplar datasets
URL	https://www.bv-brc.org/app/MetaCATS
Prerequisites	Bacterial Fasta contig files in Workspace
References	https://www.bv-brc.org/docs/quick_references/services/metacats.html https://www.bv-brc.org/docs/tutorial/metacats/metacats.html
Tester(s)	Rebecca Wattam, Maulik Shukla
Test date	21-Apr-2022 (follow-up from original test)
Test result	**Pass (minor issue reported)**

Overview

- Test the service using exemplar datasets
- Test input options, i.e., feature groups and alignment files
- Test auto grouping using various metadata attributes
- For each job submitted, verify successful completion of the job, availability of the output files in the workspace, and quality of the results.

Test Data

Dataset	Rational	Input Format	Input
Spike protein – nonhuman mammals	Interesting use case	Single genome group	SARSCov2 Spike Protein - Nonhuman Mammals
Spike protein – nonhuman mammals	Interesting use case	Multiple genome groups	SARSCov2 Spike Protein - Cat SARSCov2 Spike Protein - Deer SARSCov2 Spike Protein - Minks
Spike protein – nonhuman mammals	Interesting use case	Alignment file	SARSCov2 Spike Protein - Nonhuman Mammals - MSA.afa pike_protein_nonhuman_mammal_metadata.tsv

- 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/MetaCATS

Test Results

- All MetaCATS jobs completed successfully, without any errors.
- All jobs resulted in expected output files in corresponding job output directory, including

- chisqTable.tsv - a tab separated value file with results for a “Chi-square Goodness” of fit test result, i.e. positions that have significant non-random distribution between the specified groups
- mcTable.tsv - a tab separated value file with results for adjusted p-values for multiple comparisons.

- The interactive chi-square table provided option for filtering for only those positions that vary significantly across host groups.
- 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/MetaCATS
- Below are a series of screenshots showing successful completion of the jobs, availability of the result files in the workspace, the interactive chi-square table positions that vary significantly across host groups.

- Input genomes groups and jobs results in the workspace.

Name	Size	Owner	Members	Created
Parent folder			-	
SARSCov2 Spike Protein - Cat	41 features	me	Public	5/7/22, 5:00 PM
SARSCov2 Spike Protein - Deer	15 features	me	Public	5/7/22, 5:02 PM
SARSCov2 Spike Protein - Minks	136 features	me	Public	5/7/22, 5:03 PM
SARSCov2 Spike Protein - Nonhuman Mammals	221 features	me	Public	5/7/22, 4:56 PM
SARSCov2 Spike Protein - Nonhuman Mammals - MSA	7.1 kB	me	Public	5/7/22, 5:06 PM
Spike protein - nonhuman mammal - autogroup	10.8 kB	me	Public	5/7/22, 5:01 PM
Spike protein - nonhuman mammals - custom groups	6.4 kB	me	Public	5/7/22, 5:04 PM

- Spike protein – nonhuman mammals – auto grouping based on host names in genome metadata

Parameters

P-VALUE

0.05

OUTPUT FOLDER

MetaCATS

OUTPUT NAME

Spike protein - nonhuman mammal - autogroup

Input

AUTO GROUPING FEATURE GROUPS ALIGNMENT FILE

METADATA

Host Name

SELECT FEATURE GROUP

SARSCov2 Spike Protein - Nonhuman Mammals

DNA PROTEIN

GROUP NAMES

Change group

Delete Rows *9 groups.*

Name	Size	Owner	Members	Created
Parent folder			-	
Spike protein - nonhuman mammal - autogroup-chisqTable.tsv	34.1 kB	me	Public	5/7/22, 5:01 PM
Spike protein - nonhuman mammal - autogroup-mcTable.tsv	1.6 MB	me	Public	5/7/22, 5:01 PM
Spike protein - nonhuman mammal - autogroup.mafft.log	3.6 kB	me	Public	5/7/22, 5:01 PM

BVBRC / BVBRC Tests / MetaCATS / .Spike protein - nonhuman mammal - autogroup / Spike protein - nonhuman mammal - autogroup-chisqTable.tsv

KEYWORDS | Significant | Filter | Reset

First Row Contains Column Headers

Position	Chi-square_value	P-value	Significant	Degrees_of_fi	Fewer_5	Dog	Felis catus	Gorilla gorilla gorilla	Mink	Mustela lutreola	Neogale vison	Neovison vison	Odocoileus virginianus	cat
19	24.7856169259	0.0016900306	Y	8	Y	1 R	1 T	1 T	7 T	2 T	2 T	8 T	1 T	2 T
69	24.7856169259	0.0016900306	Y	8	Y	1 H	1 H	1 -	7 H	2 H	2 H	8 H	1 H	2 H
70	24.7856169259	0.0016900306	Y	8	Y	1 V	1 V	1 -	7 V	2 V	2 V	8 V	1 V	2 V
112	24.7856169259	0.0016900306	Y	8	Y	1 L	1 S	1 S	7 S	2 S	2 S	8 S	1 S	2 S
142	17.4718410778	0.0255544069	Y	8	Y	1 D	1 G	1 G	7 D	2 G	2 G	5 D, 3 G	1 G	2 G
144	24.9766750196	0.0015687685	Y	8	Y	1 Y	1 Y	1 -	7 -	2 Y	2 Y	8 Y	1 Y	2 Y
156	24.7856169259	0.0016900306	Y	8	Y	1 G	1 E	1 E	7 E	2 E	2 E	8 E	1 E	2 E
157	24.7856169259	0.0016900306	Y	8	Y	1 -	1 F	1 F	7 F	2 F	2 F	8 F	1 F	2 F
158	24.7856169259	0.0016900306	Y	8	Y	1 -	1 R	1 R	7 R	2 R	2 R	8 R	1 R	2 R
245	24.7856169259	0.0016900306	Y	8	Y	1 H	1 H	1 H	7 H	2 H	2 H	8 H	1 Y	2 H
262	24.8962154910	0.0016187657	Y	8	Y	1 A	1 A	1 A	7 A	2 A	2 S	8 A	1 A	2 A
314	24.8962154910	0.0016187657	Y	8	Y	1 Q	1 Q	1 Q	7 Q	2 Q	2 K	8 Q	1 Q	2 Q
328	24.8962154910	0.0016187657	Y	8	Y	1 R	1 R	1 R	7 R	2 R	2 X	8 R	1 R	2 R
452	24.7856169259	0.0016900306	Y	8	Y	1 R	1 L	1 L	7 L	2 L	2 L	8 L	1 L	2 L
478	24.7856169259	0.0016900306	Y	8	Y	1 K	1 T	1 T	7 T	2 T	2 T	8 T	1 T	2 T
486	24.9789672689	0.0015673663	Y	8	Y	1 F	1 F	1 F	7 L	2 F	2 L	8 F	1 F	2 F
501	41.9399017900	0.0004028224	Y	16	Y	1 N	1 N	1 Y	7 T	2 N	2 N	3 N, 5 T	1 N	2 N
570	24.7856169259	0.0016900306	Y	8	Y	1 A	1 A	1 D	7 A	2 A	2 A	8 A	1 A	2 A
681	49.5409737231	2.7121695336	Y	16	Y	1 R	1 P	1 H	7 P	2 P	2 P	8 P	1 P	2 P
716	24.7856169259	0.0016900306	Y	8	Y	1 T	1 T	1 I	7 T	2 T	2 T	8 T	1 T	2 T
950	24.7856169259	0.0016900306	Y	8	Y	1 N	1 D	1 D	7 D	2 D	2 D	8 D	1 D	2 D
982	24.7856169259	0.0016900306	Y	8	Y	1 S	1 S	1 A	7 S	2 S	2 S	8 S	1 S	2 S
1118	24.7856169259	0.0016900306	Y	8	Y	1 D	1 D	1 H	7 D	2 D	2 D	8 D	1 D	2 D

1 - 24 of 24 results

- Spike protein – nonhuman mammals – manual grouping by using host specific genome groups

Parameters

P-VALUE: 0.05

OUTPUT FOLDER: MetaCATS

OUTPUT NAME: Spike protein - nonhuman mammals - custom groups

Input

AUTO GROUPING | FEATURE GROUPS | ALIGNMENT FILE

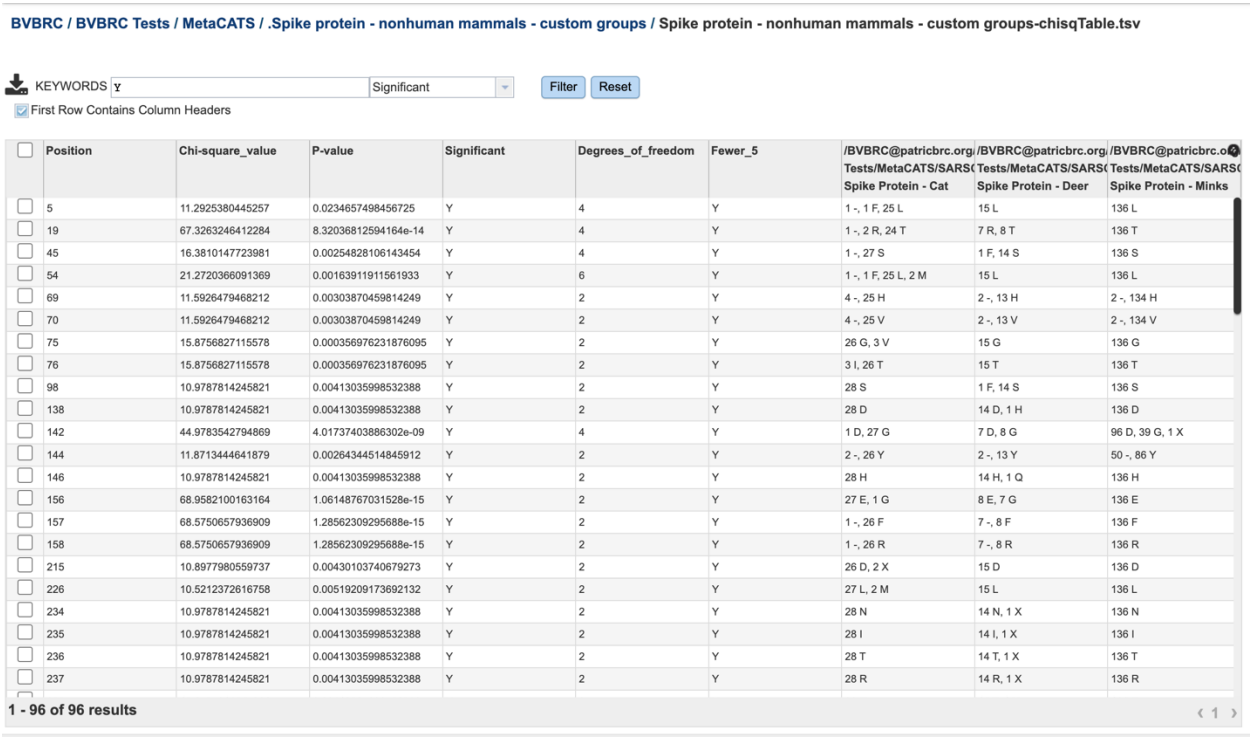
SELECT FEATURE GROUP: SARSCov2 Spike Protein - Minks

DNA | PROTEIN

SELECTED GROUPS TABLE

SARSCov2 Spike Protein - Minks
SARSCov2 Spike Protein - Deer
SARSCov2 Spike Protein - Cat

Reset | Submit



BVBRC / BVBRC Tests / MetaCATS / .Spike protein - nonhuman mammals - custom groups / Spike protein - nonhuman mammals - custom groups-chisqTable.tsv

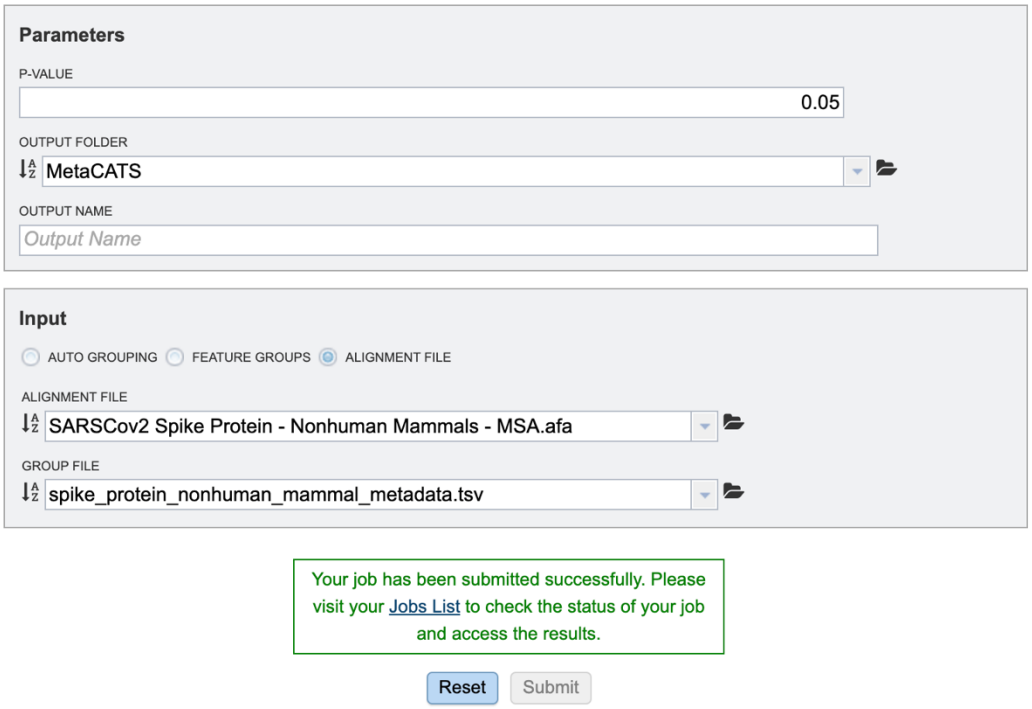
KEYWORDS Y Significant Filter Reset

First Row Contains Column Headers

Position	Chi-square_value	P-value	Significant	Degrees_of_freedom	Fewer_5	/BVBRC@patricbrc.org/Tests/MetaCATS/SARS(Spike Protein - Cat	/BVBRC@patricbrc.org/Tests/MetaCATS/SARS(Spike Protein - Deer	/BVBRC@patricbrc.org/Tests/MetaCATS/SARS(Spike Protein - Minks
5	11.2925380445257	0.0234657498456725	Y	4	Y	1 -, 1 F, 25 L	15 L	136 L
19	67.3263246412284	8.32036812594164e-14	Y	4	Y	1 -, 2 R, 24 T	7 R, 8 T	136 T
45	16.3810147723981	0.00254828106143454	Y	4	Y	1 -, 27 S	1 F, 14 S	136 S
54	21.2720366091369	0.00163911911561933	Y	6	Y	1 -, 1 F, 25 L, 2 M	15 L	136 L
69	11.5926479468212	0.00303870459814249	Y	2	Y	4 -, 25 H	2 -, 13 H	2 -, 134 H
70	11.5926479468212	0.00303870459814249	Y	2	Y	4 -, 25 V	2 -, 13 V	2 -, 134 V
75	15.8756827115578	0.0003569762318760 95	Y	2	Y	26 G, 3 V	15 G	136 G
76	15.8756827115578	0.0003569762318760 95	Y	2	Y	3 I, 26 T	15 T	136 T
98	10.9787814245821	0.00413035998532388	Y	2	Y	28 S	1 F, 14 S	136 S
138	10.9787814245821	0.00413035998532388	Y	2	Y	28 D	14 D, 1 H	136 D
142	44.9783542794869	4.01737403886302e-09	Y	4	Y	1 D, 27 G	7 D, 8 G	96 D, 39 G, 1 X
144	11.8713444641879	0.00264344514845912	Y	2	Y	2 -, 26 Y	2 -, 13 Y	50 -, 86 Y
146	10.9787814245821	0.00413035998532388	Y	2	Y	28 H	14 H, 1 Q	136 H
156	68.9582100163164	1.06148767031528e-15	Y	2	Y	27 E, 1 G	8 E, 7 G	136 E
157	68.5750657936909	1.28562309295688e-15	Y	2	Y	1 -, 26 F	7 -, 8 F	136 F
158	68.5750657936909	1.28562309295688e-15	Y	2	Y	1 -, 26 R	7 -, 8 R	136 R
215	10.8977980559737	0.00430103740679273	Y	2	Y	26 D, 2 X	15 D	136 D
226	10.5212372616758	0.00519209173692132	Y	2	Y	27 L, 2 M	15 L	136 L
234	10.9787814245821	0.00413035998532388	Y	2	Y	28 N	14 N, 1 X	136 N
235	10.9787814245821	0.00413035998532388	Y	2	Y	28 I	14 I, 1 X	136 I
236	10.9787814245821	0.00413035998532388	Y	2	Y	28 T	14 T, 1 X	136 T
237	10.9787814245821	0.00413035998532388	Y	2	Y	28 R	14 R, 1 X	136 R

1 - 96 of 96 results ‹ 1 ›

- Spike protein – nonhuman mammals – manual grouping by using user supplied alignment file and group metadata file.



Note:

- Bug: The jobs failed when using user supplied alignment and group files as input.
- Resolution: The bug has been identified and being fixed. It will be deployed to production in the next release.

References

- Metadata-driven Comparative Analysis Tool (meta-CATS) Quick Reference Guide
- Metadata-driven Comparative Analysis Tool (meta-CATS) Tutorial