

BV-BRC Test Report

A20. Service – Metagenomic Binning

Item to test	Metagenomic Binning Service using metagenomic read sets
URL	https://www.bv-brc.org/app/MetagenomicBinning
Prerequisites	Metagenomic reads as SRA accession or fastq files
References	https://www.bv-brc.org/docs/quick_references/services/metagenomic_binning_service.html
Tester(s)	Rebecca Wattam, Maulik Shukla
Test date	08-May-2022 (follow-up from original test)
Test result	Passed

Overview

- Test the Metagenomic Binning Service using exemplar metagenomic reads sets.
- Test input options, i.e., single-end / paired-end read files and SRA accessions.
- For each job submitted, verify successful completion of the job and presence of output files.
- Review the resulting metagenomic bins and their quality, including completeness and contamination.

Test Data

Dataset	Rational	Input Format	Input
Louse Fly Metagenome	Workshop example	Fastq file, SRA accession	ERR969522 ERR969522_1.fastq.gz ERR969522_2.fastq.gz
ICU metagenome sample	Workshop example	Fastq file	UC.MICU.02.30.fastq

- 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/Metagenomic%20Binning>

Test Results

- All jobs completed successfully, without any errors.
- All jobs resulted in expected output files in corresponding job output directory, including HTML reports with list of bacterial and viral bins and assessment of their quality.
- The HTML report was informative and provided a concise summary of the bacterial and viral bins extracted, assessment of their quality and links to annotated bins.
- 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/Metagenomic%20Binning>
- Below are a series of screenshots showing successful completion of the jobs, availability of the result files in the workspace, excerpts of the bacterial and viral binning reports.

Start With: ?

READ FILE ASSEMBLED CONTIGS

Input File ?

PAIRED READ LIBRARY

ERR969522_1.fastq.gz

ERR969522_2.fastq.gz

SINGLE READ LIBRARY

READ FILE

SRA RUN ACCESSION

SRR

Selected libraries ?

Place read files here using the arrow buttons.

P(ERR96...tq.gz, ERR96...tq.gz)

Parameters ?

ASSEMBLY STRATEGY

METASPADES MEGAHIT AUTO

ORGANISMS OF INTEREST

BACTERIA/ARCHAEA VIRUSES BOTH

OUTPUT FOLDER

Metagenomic Binning

OUTPUT NAME

Output Name

GENOME GROUP NAME

My Genome Group

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

[Reset](#) [Submit](#)

Status	ID	Service	Output Name	Submit	Start	Completed
completed	7747936	MetagenomeBinning	Louse Fly Metagenome - SRA	5/8/22, 8:17 PM	5/8/22, 8:17 PM	5/9/22, 3:07 AM
completed	7747937	MetagenomeBinning	Louse Fly Metagenome - fastq	5/8/22, 8:18 PM	5/8/22, 8:18 PM	5/9/22, 12:12 AM
completed	7747938	MetagenomeBinning	ICU Metagenome	5/8/22, 8:19 PM	5/8/22, 8:19 PM	5/8/22, 8:54 PM

BVBRC / BVBRC Tests / Metagenomic Binning (6 items)

UPLOAD ADD FOLDER SHOW HIDDEN

Name	Size	Owner	Members	Created
Parent folder	-	-	-	-
ERR969522_1.fastq.gz	3.9 GB	me	Public	5/8/22, 8:08 PM
ERR969522_2.fastq.gz	3.6 GB	me	Public	5/8/22, 8:08 PM
ICU Metagenome	18.5 kB	me	Public	5/8/22, 8:54 PM
Louse Fly Metagenome - SRA	32.5 kB	me	Public	5/9/22, 3:07 AM
Louse Fly Metagenome - fastq	33.0 kB	me	Public	5/9/22, 12:12 AM
UC.MICU.02.30.fastq	8.3 GB	me	Public	5/8/22, 8:07 PM
Parent folder	-	-	-	-
BinningReport.html	4.2 kB	me	Public	5/9/22, 3:07 AM
ViralBinningReport.html	3.8 kB	me	Public	5/9/22, 3:06 AM
bin.1.1247024	32.8 kB	me	Public	5/9/22, 2:47 AM
bin.1.1247024.fa	4.9 MB	me	Public	5/9/22, 2:40 AM
bin.2.1239307	32.8 kB	me	Public	5/9/22, 2:55 AM
bin.2.1239307.fa	4.8 MB	me	Public	5/9/22, 2:40 AM
bin.3.244204	32.7 kB	me	Public	5/9/22, 2:58 AM
bin.3.244204.fa	1.9 MB	me	Public	5/9/22, 2:40 AM
bin.4.1094557	32.8 kB	me	Public	5/9/22, 3:01 AM
bin.4.1094557.fa	1.6 MB	me	Public	5/9/22, 2:40 AM
bins.json	390.4 kB	me	Public	5/9/22, 2:40 AM
bins.stats.txt	994 B	me	Public	5/9/22, 2:40 AM
contigs.fasta	194.9 MB	me	Public	5/9/22, 2:26 AM
coverage.stats.txt	785 B	me	Public	5/9/22, 2:27 AM
params.txt	1.1 kB	me	Public	5/9/22, 2:26 AM
spades.log	157.8 kB	me	Public	5/9/22, 2:26 AM

Generated bins

2 [good bins](#) and 2 [bad bins](#) were found out of 4.

The following bins met the quality criteria of completeness >= 80%, fine consistency >= 87%, contamination <= 10%, and have a single PheS protein of reasonable size:

Score	Genome ID	Genome Name	Reference Genome	Coarse consistency (%)	Fine consistency (%)	Completeness (%)	Contamination (%)	Hypothetical (%)	Contig count	DNA size (bp)	Contigs N50 (bp)	Mean Coverage	Potentially Problematic Roles	Good PheS
2670	1094557.53	Bartonella melophagi clonal population	1094557.5	98.6	97.7	99.7	2.1	27.84	85	1522342	143167	166.49	35 roles	Y
1576	1247024.60	Arsenophonus endosymbiont clonal population	1247024.12	94.4	93.8	100	7.7	91.87	1419	4757144	5640	33.81	84 roles	Y

The following bins did not meet the quality criteria:

Score	Genome ID	Genome Name	Reference Genome	Coarse consistency (%)	Fine consistency (%)	Completeness (%)	Contamination (%)	Hypothetical (%)	Contig count	DNA size (bp)	Contigs N50 (bp)	Mean Coverage	Potentially Problematic Roles	Good PheS
1725	1239307.66	Sodalis sp. HS1 clonal population	1239307.3	95.9	81.7	91.5	7.1	59.89	1261	4698806	5921	13.95	257 roles	Y
179	244204.1219	uncultured Wolbachia clonal population	244204.10	99.6	81.9	100	44.6	41.42	689	1878087	6474	6.52	308 roles	Y

Generated bins

This table lists all the known viruses found in the sample.

Bin	Virus ID	Taxon ID	Genome ID	Genome Name	Length	Completeness (%)	Error (%)	Coverage
1	GCA_004194195.1	2516434	Not annotated	Klebsiella phage ST405-OXA48phi.1	13232	37.17	7.59	23.31
3	GCA_003958845.1	2071659	Not annotated	Escherichia phage sp.	9548	24.37	7.83	2.92
8	GCA_000840645.2	173824	Not annotated	Methanothermobacter phage psiM100	4086	13.18	9	32.46
10	GCA_000858165.1	101570	Not annotated	Enterobacteria phage ES18	5322	11.85	6.01	157.05
13	GCA_000836965.1	10742	Not annotated	Escherichia virus HK022	4502	11.3	2.07	39.64
15	GCA_002630045.1	373126	Not annotated	Sodalis phage phiSG1	5418	10.61	8.37	148.22
2	DTR_731398	10239	Not annotated	ERS698801 NODE_6	13451	30.74	6.88	87.58
4	DTR_490511	10239	Not annotated	VincentC_2016_MM098.7_NODE_5	9737	20.81	3.72	34.31
5	DTR_072516	10239	Not annotated	3300006943_Ga0099822_1020790	1316	18.16	9.4	20.73
6	DTR_722076	10239	Not annotated	SRR6486882_NODE_49	8194	17.42	3.73	36.58
7	DTR_210116	10239	Not annotated	3300018411_Ga0194137_10000286	5031	14.25	5.89	4.53
9	DTR_387353	10239	Not annotated	CM_madagascar_A88_02_1FE_NODE_268	6525	12.11	9.15	2.19
11	DTR_153427	10239	Not annotated	3300011717_Ga0122754_100126	5500	11.76	4.21	22.45
12	DTR_226270	10239	Not annotated	3300020147_Ga0196976_1000072	7797	11.71	7.19	2.69
14	DTR_877368	10239	Not annotated	SRS476205_k119_232982	10585	11.03	7.03	3.37
16	DTR_378146	10239	Not annotated	BritoIL_2016_M1.16.ST_NODE_901	3922	10.31	3.57	277.26

- ICU Metagenome Analysis

Name	Size	Owner	Members	Created
Parent folder	-	-	-	-
BinningReport.html	3.0 kB	me	Public	5/8/22, 8:54 PM
ViralBinningReport.html	1.5 kB	me	Public	5/8/22, 8:54 PM
bin.1.548	32.1 kB	me	Public	5/8/22, 8:50 PM
bin.1.548.fa	5.4 MB	me	Public	5/8/22, 8:46 PM
bin.2.1308	32.3 kB	me	Public	5/8/22, 8:53 PM
bin.2.1308.fa	2.0 MB	me	Public	5/8/22, 8:46 PM
bins.json	24.4 kB	me	Public	5/8/22, 8:46 PM
bins.stats.txt	963 B	me	Public	5/8/22, 8:46 PM
contigs.fasta	20.8 MB	me	Public	5/8/22, 8:44 PM
coverage.stats.txt	808 B	me	Public	5/8/22, 8:44 PM
megahit.log	131.3 kB	me	Public	5/8/22, 8:44 PM
options.json	805 B	me	Public	5/8/22, 8:44 PM
unbinned.fasta	12.3 MB	me	Public	5/8/22, 8:46 PM
unplaced.fasta	12.4 MB	me	Public	5/8/22, 8:46 PM
vBin1.fa	5.0 kB	me	Public	5/8/22, 8:54 PM
vBin2.fa	2.4 kB	me	Public	5/8/22, 8:54 PM

Generated bins

1 [good bins](#) and 1 [bad bins](#) were found out of 2.

The following bins met the quality criteria of completeness $\geq 80\%$, fine consistency $\geq 87\%$, contamination $\leq 10\%$, and have a single PheS protein of reasonable size:

Score	Genome ID	Genome Name	Reference Genome	Coarse consistency (%)	Fine consistency (%)	Completeness (%)	Contamination (%)	Hypothetical (%)	Contig count	DNA size (bp)	Contigs N50 (bp)	Mean Coverage	Potentially Problematic Roles	Good PheS
2748	548.1056	[Enterobacter] aerogenes clonal population	548.198	99.8	97.5	100	0.6	27.69	74	5332294	163049	266.77	36 roles	Y

The following bins did not meet the quality criteria:

Score	Genome ID	Genome Name	Reference Genome	Coarse consistency (%)	Fine consistency (%)	Completeness (%)	Contamination (%)	Hypothetical (%)	Contig count	DNA size (bp)	Contigs N50 (bp)	Mean Coverage	Potentially Problematic Roles	Good PheS
2582	1308.1758	Streptococcus thermophilus clonal population	1308.966	99.4	98.0	99	3.1	30.11	200	1992528	29490	706.15	36 roles	

Generated bins

This table lists all the known viruses found in the sample.

Bin	Virus ID	Taxon ID	Genome ID	Genome Name	Length	Completeness (%)	Error (%)	Coverage
2	GCA_004776585.1	2558776	<i>Not annotated</i>	Streptococcus satellite phage Javan607	2340	20.57	1.15	16.86
1	DTR_346915	10239	<i>Not annotated</i>	3300028581__Ga0247840_10021653	4942	83.32	9.11	5.52
3	DTR_648371	10239	<i>Not annotated</i>	SRR2105903__NODE_5	924	14.9	1.22	1.47
4	DTR_170259	10239	<i>Not annotated</i>	3300012979__Ga0123348_10000191	5359	10.32	8.9	4

References

- [Metagenomic Binning Quick Reference Guide](#)
- [Metagenomic Binning Tutorial](#)