

BV-BRC Test Report

A2. Service – Genome Annotation - Bacteria

Item to test	Genome Annotation Service using bacterial contig files
URL	https://www.bv-brc.org/app/Annotation
Prerequisites	Bacterial Fasta contig files in Workspace
References	https://www.bv-brc.org/docs/quick_references/services/genome_annotation_service.html https://www.bv-brc.org/docs/tutorial/genome_annotation/genome_annotation.html
Tester(s)	Rebecca Wattam, Maulik Shukla
Test date	21-Apr-2022 (follow-up from original test)
Test result	Passed

Overview

- Test the Genome Annotation Service using exemplar genome sequences for bacterial genomes.
- Test input options, i.e., contig file from the user's machine or from the user's workspace.
- For each job submitted, verify successful completion of the job, availability of the output files in the workspace, and quality of the annotations by comparing them with the same, or closely related, public genome in PATRIC.
- Verify successful integration of the genome in BV-BRC by reviewing the genome overview pages and other genome-level tabs.

Test Data

Genome Name	Rational	Input File
Mycobacterium tuberculosis H37Rv	Reference genome	Mycobacterium_tuberculosis_H37Rv.fasta
Escherichia coli MG1655	Reference genome	Escherichia_coli_MG1655.fasta
Brucella suis 1330	Reference genome	Brucella_suis_1330.fasta
Buchnera aphidicola Tuc7	Workshop example	Buchnera_aphidicola_strain Tuc7_SRR4240359.fasta
Escherichia coli MRSN388634	Workshop example	Escherichia_coli_MRSN388634.fasta
Staphylococcus aureus VB4283	Workshop example	Staphylococcus_aureus_VB4283.fna

- 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/Genome%20Annotation/Bacteria>

Test Results

- All annotation jobs completed successfully, without any errors.
- All jobs resulted in expected output files in corresponding job output directory, providing annotations in various standard file formats.

- All genomes were indexed successfully in the Solr database and were accessible from the BV-BRC website.
- For each of the genomes, the genome overview page, genome browser, and all other genome level tabs loaded successfully as expected.
- For each of the genomes, annotation of genomic features, protein functions, specialty genes was as expected when compared to those in corresponding public genomes in PATRIC.
- 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/Genome%20Annotation/Bacteria>
- Below are a series of screenshots showing successful completion of the jobs, availability of the result files in the workspace, and comparison of the annotation between test genome and corresponding public genome.

Status	ID	Service	Output Name	Submit ▲	Start	Completed +
completed	7419704	Annotation	Buchnera aphidicola strain Tuc7	4/20/22, 2:48 PM	4/20/22, 2:48 PM	4/20/22, 2:53 PM
completed	7419707	Annotation	Escherichia coli MRSN388634	4/20/22, 2:48 PM	4/20/22, 2:49 PM	4/20/22, 2:58 PM
completed	7419711	Annotation	Staphylococcus aureus VB4283	4/20/22, 2:49 PM	4/20/22, 2:49 PM	4/20/22, 2:56 PM
completed	7419813	Annotation	Brucella suis 1330	4/20/22, 3:16 PM	4/20/22, 3:16 PM	4/20/22, 3:24 PM
completed	7419816	Annotation	Escherichia coli MG1655	4/20/22, 3:16 PM	4/20/22, 3:17 PM	4/20/22, 3:27 PM
completed	7419819	Annotation	Mycobacterium tuberculosis H37Rv	4/20/22, 3:17 PM	4/20/22, 3:17 PM	4/20/22, 3:27 PM

- Result verification: Brucella suis 1330

Job result directory: <https://www.bv-brc.org/workspace/BVBRC@patricbrc.org/BVBRC%20Tests/Genome%20Annotation/Bacteria/Brucella%20suis%201330>

Name	Size	Owner	Members	Created
Parent folder			-	
Brucella suis 1330.contigs.fasta	3.4 MB	me	Public	4/20/22, 3:21 PM
Brucella suis 1330.embl	7.3 MB	me	Public	4/20/22, 3:22 PM
Brucella suis 1330.feature_dna.fasta	3.4 MB	me	Public	4/20/22, 3:21 PM
Brucella suis 1330.feature_protein.fasta	1.3 MB	me	Public	4/20/22, 3:21 PM
Brucella suis 1330.features.txt	430.9 kB	me	Public	4/20/22, 3:21 PM
Brucella suis 1330.gb	7.0 MB	me	Public	4/20/22, 3:21 PM
Brucella suis 1330.genome	11.8 MB	me	Public	4/20/22, 3:21 PM
Brucella suis 1330.gff	458.6 kB	me	Public	4/20/22, 3:22 PM
Brucella suis 1330.merged.gb	7.0 MB	me	Public	4/20/22, 3:21 PM
Brucella suis 1330.tar.gz	1.9 MB	me	Public	4/20/22, 3:22 PM
Brucella suis 1330.txt	4.5 MB	me	Public	4/20/22, 3:22 PM
Brucella suis 1330.xls	5.0 MB	me	Public	4/20/22, 3:21 PM
GenomeReport.html	22.0 kB	me	Public	4/20/22, 3:21 PM
genome_quality_details.txt	28.0 kB	me	Public	4/20/22, 3:21 PM
load_files		me	Public	4/20/22, 3:22 PM
quality.json	308.9 kB	me	Public	4/20/22, 3:24 PM

- Annotation comparison with the public reference genome in PATRIC:

<https://patricbrc.org/view/Genome/204722.5>

Genomic Features			Genomic Features		
	PATRIC	RefSeq		PATRIC	RefSeq
CDS	3270	0	CDS	3432	3272
repeat_region	94	0	rRNA	55	55
rRNA	55	0	pseudogene	22	62
rRNA	9	0	rRNA	9	9
			misc_RNA	3	0

Protein Features			Protein Features		
	PATRIC	RefSeq		PATRIC	RefSeq
Hypothetical proteins	580	0	Hypothetical proteins	747	1148
Proteins with functional assignments	2690	0	Proteins with functional assignments	2685	2124
Proteins with EC number assignments	930	0	Proteins with EC number assignments	913	519
Proteins with GO assignments	796	0	Proteins with GO assignments	578	0
Proteins with Pathway assignments	715	0	Proteins with Pathway assignments	715	439
Proteins with Subsystem assignments	1239	0	Proteins with Subsystem assignments	1240	0
Proteins with PATRIC genus-specific family (PLfam) assignments	3170	0	Proteins with PATRIC genus-specific family (PLfam) assignments	3430	0
Proteins with PATRIC cross-genus family (PGfam) assignments	3195	0	Proteins with PATRIC cross-genus family (PGfam) assignments	3430	0
Proteins with FIGfam assignments	0	0	Proteins with FIGfam assignments	3381	0

Specialty Genes			Specialty Genes		
	Source	Genes		Source	Genes
Virulence Factor	Victors	225	Virulence Factor	Victors	226
Virulence Factor	VFDB	43	Virulence Factor	VFDB	42
Virulence Factor	PATRIC_VF	1	Virulence Factor	PATRIC_VF	1
Transporter	TCDB	26	Transporter	TCDB	26
Drug Target	DrugBank	4	Human Homolog	Human	78
Drug Target	TTD	1	Essential Gene	PATRIC	224

- Result verification: Mycobacterium tuberculosis H37Rv

Job result directory: <https://www.bv-brc.org/workspace/BVBRC@patricbrc.org/BVBRC%20Tests/Genome%20Annotation/Bacteria/Mycobacterium%20tuberculosis%20H37Rv>

Name	Size	Owner	Members	Created
Parent folder			-	
GenomeReport.html	12.7 kB	me	Public	4/20/22, 3:25 PM
Mycobacterium tuberculosis H37Rv.contigs.fasta	4.5 MB	me	Public	4/20/22, 3:26 PM
Mycobacterium tuberculosis H37Rv.embl	9.8 MB	me	Public	4/20/22, 3:25 PM
Mycobacterium tuberculosis H37Rv.feature_dna.fasta	4.7 MB	me	Public	4/20/22, 3:26 PM
Mycobacterium tuberculosis H37Rv.feature_protein.fasta	1.9 MB	me	Public	4/20/22, 3:26 PM
Mycobacterium tuberculosis H37Rv.features.txt	566.2 kB	me	Public	4/20/22, 3:25 PM
Mycobacterium tuberculosis H37Rv.gb	9.4 MB	me	Public	4/20/22, 3:26 PM
Mycobacterium tuberculosis H37Rv.genome	15.8 MB	me	Public	4/20/22, 3:25 PM
Mycobacterium tuberculosis H37Rv.gff	599.7 kB	me	Public	4/20/22, 3:26 PM
Mycobacterium tuberculosis H37Rv.merged.gb	9.4 MB	me	Public	4/20/22, 3:26 PM
Mycobacterium tuberculosis H37Rv.tar.gz	2.5 MB	me	Public	4/20/22, 3:26 PM
Mycobacterium tuberculosis H37Rv.txt	6.1 MB	me	Public	4/20/22, 3:26 PM
Mycobacterium tuberculosis H37Rv.xls	6.8 MB	me	Public	4/20/22, 3:26 PM
genome_quality_details.txt	54.8 kB	me	Public	4/20/22, 3:25 PM
load_files		me	Public	4/20/22, 3:26 PM
quality.json	554.7 kB	me	Public	4/20/22, 3:27 PM

Annotation comparison with the public reference genome in PATRIC:

<https://patricbrc.org/view/Genome/83332.12>

Genomic Features				Genomic Features			
	PATRIC	RefSeq			PATRIC	RefSeq	
CDS	4264	0		CDS	4367	3989	
repeat_region	96	0		tRNA	45	45	
classifier_predicted_region	46	0		pseudogene	26	8	
tRNA	44	0		rRNA	2	3	
crispr_repeat	43	0		misc_RNA	1	2	
crispr_spacer	41	0		misc_feature	0	1149	

Protein Features				Protein Features			
	PATRIC	RefSeq			PATRIC	RefSeq	
Hypothetical proteins	869	0		Hypothetical proteins	1033	1481	
Proteins with functional assignments	3395	0		Proteins with functional assignments	3334	2508	
Proteins with EC number assignments	1063	0		Proteins with EC number assignments	1055	1081	
Proteins with GO assignments	919	0		Proteins with GO assignments	935	0	
Proteins with Pathway assignments	814	0		Proteins with Pathway assignments	812	921	
Proteins with Subsystem assignments	1427	0		Proteins with Subsystem assignments	1409	0	
Proteins with PATRIC genus-specific family (PLfam) assignments	3981	0		Proteins with PATRIC genus-specific family (PLfam) assignments	4364	0	
Proteins with PATRIC cross-genus family (PGfam) assignments	4185	0		Proteins with PATRIC cross-genus family (PGfam) assignments	4364	0	
Proteins with FIGfam assignments	0	0		Proteins with FIGfam assignments	4290	0	

Specialty Genes				Specialty Genes			
	Source	Genes			Source	Genes	
Antibiotic Resistance	PATRIC	57		Antibiotic Resistance	PATRIC	57	
Antibiotic Resistance	CARD	29		Antibiotic Resistance	CARD	29	
Antibiotic Resistance	NDARO	3		Antibiotic Resistance	NDARO	3	
Drug Target	DrugBank	71		Drug Target	DrugBank	71	
Drug Target	TTD	29		Drug Target	TTD	29	
Transporter	TCDB	119		Essential Gene	PATRIC	215	

- Result verification: Escherichia coli MG1655

Job result directory: <https://www.bv-brc.org/workspace/BVBRC@patricbrc.org/BVBRC%20Tests/Genome%20Annotation/Bacteria/Escherichia%20coli%20MG1655>

Name	Size	Owner	Members	Created
Parent folder			-	
Escherichia coli MG1655.contigs.fasta	4.7 MB	me	Public	4/20/22, 3:24 PM
Escherichia coli MG1655.embl	10.2 MB	me	Public	4/20/22, 3:24 PM
Escherichia coli MG1655.feature_dna.fasta	4.8 MB	me	Public	4/20/22, 3:24 PM
Escherichia coli MG1655.feature_protein.fasta	1.9 MB	me	Public	4/20/22, 3:24 PM
Escherichia coli MG1655.features.txt	597.6 kB	me	Public	4/20/22, 3:24 PM
Escherichia coli MG1655.gb	9.8 MB	me	Public	4/20/22, 3:24 PM
Escherichia coli MG1655.genome	16.6 MB	me	Public	4/20/22, 3:23 PM
Escherichia coli MG1655.gff	636.8 kB	me	Public	4/20/22, 3:24 PM
Escherichia coli MG1655.merged.gb	9.8 MB	me	Public	4/20/22, 3:24 PM
Escherichia coli MG1655.tar.gz	2.7 MB	me	Public	4/20/22, 3:24 PM
Escherichia coli MG1655.txt	6.4 MB	me	Public	4/20/22, 3:24 PM
Escherichia coli MG1655.xls	7.1 MB	me	Public	4/20/22, 3:24 PM
GenomeReport.html	6.0 kB	me	Public	4/20/22, 3:23 PM
genome_quality_details.txt	44.9 kB	me	Public	4/20/22, 3:23 PM
load_files		me	Public	4/20/22, 3:24 PM
quality.json	448.0 kB	me	Public	4/20/22, 3:27 PM

Annotation comparison with the public reference genome in PATRIC:

<https://patricbrc.org/view/Genome/511145.12>

Genomic Features			Genomic Features		
	PATRIC	RefSeq		PATRIC	RefSeq
CDS	4506	0	CDS	4549	4321
repeat_region	94	0	tRNA	88	89
rRNA	88	0	rRNA	14	22
rRNA	22	0	misc_RNA	8	0
crispr_repeat	21	0	pseudogene	7	177
crispr_spacer	19	0	mat_peptide	0	488

Protein Features			Protein Features		
	PATRIC	RefSeq		PATRIC	RefSeq
Hypothetical proteins	268	0	Hypothetical proteins	358	27
Proteins with functional assignments	4238	0	Proteins with functional assignments	4191	4294
Proteins with EC number assignments	1267	0	Proteins with EC number assignments	1226	1046
Proteins with GO assignments	1046	0	Proteins with GO assignments	1092	2525
Proteins with Pathway assignments	887	0	Proteins with Pathway assignments	883	848
Proteins with Subsystem assignments	1802	0	Proteins with Subsystem assignments	1736	0
Proteins with PATRIC genus-specific family (PLfam) assignments	4450	0	Proteins with PATRIC genus-specific family (PLfam) assignments	4547	0
Proteins with PATRIC cross-genus family (PGfam) assignments	4452	0	Proteins with PATRIC cross-genus family (PGfam) assignments	4547	0
Proteins with FIGfam assignments	0	0	Proteins with FIGfam assignments	4413	0

Specialty Genes			Specialty Genes		
	Source	Genes		Source	Genes
Virulence Factor	Victors	229	Virulence Factor	Victors	230
Virulence Factor	PATRIC_VF	214	Virulence Factor	PATRIC_VF	215
Virulence Factor	VFDB	60	Virulence Factor	VFDB	60
Transporter	TCDB	893	Transporter	TCDB	892
Drug Target	DrugBank	388	Human Homolog	Human	55
Drug Target	TTD	58	Essential Gene	PATRIC	222

- Result verification: Buchnera aphidicola strain Tuc7

Job result directory: <https://www.bv-brc.org/workspace/BVBRC@patricbrc.org/BVBRC%20Tests/Genome%20Annotation/Bacteria/Buchnera%20aphidicola%20strain%20Tuc7>

Name	Size	Owner	Members	Created
Parent folder			-	
Buchnera aphidicola strain Tuc7.contigs.fasta	671.1 kB	me	Public	4/20/22, 2:50 PM
Buchnera aphidicola strain Tuc7.embl	1.4 MB	me	Public	4/20/22, 2:51 PM
Buchnera aphidicola strain Tuc7.feature_dna.fasta	660.8 kB	me	Public	4/20/22, 2:50 PM
Buchnera aphidicola strain Tuc7.feature_protein.fasta	260.6 kB	me	Public	4/20/22, 2:50 PM
Buchnera aphidicola strain Tuc7.features.txt	91.7 kB	me	Public	4/20/22, 2:50 PM
Buchnera aphidicola strain Tuc7.gb	1.4 MB	me	Public	4/20/22, 2:50 PM
Buchnera aphidicola strain Tuc7.genome	2.8 MB	me	Public	4/20/22, 2:50 PM
Buchnera aphidicola strain Tuc7.gff	101.3 kB	me	Public	4/20/22, 2:51 PM
Buchnera aphidicola strain Tuc7.merged.gb	1.4 MB	me	Public	4/20/22, 2:50 PM
Buchnera aphidicola strain Tuc7.tar.gz	371.6 kB	me	Public	4/20/22, 2:50 PM
Buchnera aphidicola strain Tuc7.txt	897.2 kB	me	Public	4/20/22, 2:50 PM
Buchnera aphidicola strain Tuc7.xls	993.8 kB	me	Public	4/20/22, 2:50 PM
GenomeReport.html	8.1 kB	me	Public	4/20/22, 2:50 PM
genome_quality_details.txt	27.8 kB	me	Public	4/20/22, 2:50 PM
load_files		me	Public	4/20/22, 2:51 PM
quality.json	283.4 kB	me	Public	4/20/22, 2:53 PM

Annotation comparison with corresponding public genome in PATRIC:

<https://patricbrc.org/view/Genome/9.164>

Genomic Features			Genomic Features		
	PATRIC	RefSeq		PATRIC	RefSeq
CDS	605	0	CDS	605	0
tRNA	32	0	tRNA	32	0
rRNA	3	0	rRNA	3	0

Protein Features			Protein Features		
	PATRIC	RefSeq		PATRIC	RefSeq
Hypothetical proteins	26	0	Hypothetical proteins	28	0
Proteins with functional assignments	579	0	Proteins with functional assignments	577	0
Proteins with EC number assignments	318	0	Proteins with EC number assignments	319	0
Proteins with GO assignments	267	0	Proteins with GO assignments	269	0
Proteins with Pathway assignments	241	0	Proteins with Pathway assignments	243	0
Proteins with Subsystem assignments	408	0	Proteins with Subsystem assignments	408	0
Proteins with PATRIC genus-specific family (PLfam) assignments	597	0	Proteins with PATRIC genus-specific family (PLfam) assignments	596	0
Proteins with PATRIC cross-genus family (PGfam) assignments	600	0	Proteins with PATRIC cross-genus family (PGfam) assignments	598	0
Proteins with FIGfam assignments	0	0	Proteins with FIGfam assignments	595	0

Specialty Genes			Specialty Genes		
	Source	Genes		Source	Genes
Virulence Factor	Victors	4	Virulence Factor	Victors	4
Virulence Factor	PATRIC_VF	3	Virulence Factor	PATRIC_VF	3
Transporter	TCDB	3	Transporter	TCDB	3
Drug Target	DrugBank	11	Drug Target	DrugBank	11
Drug Target	TTD	1	Drug Target	TTD	1
Antibiotic Resistance	PATRIC	17	Antibiotic Resistance	PATRIC	18

- Result verification: Escherichia coli MRSN388634

Job result directory: <https://www.bv-brc.org/workspace/BVBRC@patricbrc.org/BVBRC%20Tests/Genome%20Annotation/Bacteria/Escherichia%20coli%20MRSN388634>

Name	Size	Owner	Members	Created
Parent folder			-	
Escherichia coli MRSN388634.contigs.fasta	5.3 MB	me	Public	4/20/22, 2:56 PM
Escherichia coli MRSN388634.embl	11.5 MB	me	Public	4/20/22, 2:56 PM
Escherichia coli MRSN388634.feature_dna.fasta	5.3 MB	me	Public	4/20/22, 2:56 PM
Escherichia coli MRSN388634.feature_protein.fasta	2.1 MB	me	Public	4/20/22, 2:56 PM
Escherichia coli MRSN388634.features.txt	723.2 kB	me	Public	4/20/22, 2:56 PM
Escherichia coli MRSN388634.gb	11.1 MB	me	Public	4/20/22, 2:56 PM
Escherichia coli MRSN388634.genome	18.5 MB	me	Public	4/20/22, 2:56 PM
Escherichia coli MRSN388634.gff	750.8 kB	me	Public	4/20/22, 2:56 PM
Escherichia coli MRSN388634.merged.gb	11.1 MB	me	Public	4/20/22, 2:56 PM
Escherichia coli MRSN388634.tar.gz	2.9 MB	me	Public	4/20/22, 2:56 PM
Escherichia coli MRSN388634.txt	7.1 MB	me	Public	4/20/22, 2:56 PM
Escherichia coli MRSN388634.xls	7.9 MB	me	Public	4/20/22, 2:56 PM
GenomeReport.html	27.1 kB	me	Public	4/20/22, 2:56 PM
alignments		me	Public	4/20/22, 2:56 PM
genome_quality_details.txt	44.9 kB	me	Public	4/20/22, 2:56 PM
load_files		me	Public	4/20/22, 2:56 PM

Annotation comparison with corresponding public genome in PATRIC:

<https://patricbrc.org/view/Genome/562.10576>

Genomic Features				Genomic Features			
	PATRIC	RefSeq			PATRIC	RefSeq	
CDS	5133	0		CDS	5164	0	
rRNA	84	0		repeat_region	203	0	
rRNA	10	0		rRNA	84	0	
				rRNA	10	0	

Protein Features				Protein Features			
	PATRIC	RefSeq			PATRIC	RefSeq	
Hypothetical proteins	617	0		Hypothetical proteins	753	0	
Proteins with functional assignments	4516	0		Proteins with functional assignments	4411	0	
Proteins with EC number assignments	1345	0		Proteins with EC number assignments	1303	0	
Proteins with GO assignments	1119	0		Proteins with GO assignments	1117	0	
Proteins with Pathway assignments	954	0		Proteins with Pathway assignments	951	0	
Proteins with Subsystem assignments	1900	0		Proteins with Subsystem assignments	1617	0	
Proteins with PATRIC genus-specific family (PLfam) assignments	4961	0		Proteins with PATRIC genus-specific family (PLfam) assignments	5153	0	
Proteins with PATRIC cross-genus family (PGfam) assignments	5008	0		Proteins with PATRIC cross-genus family (PGfam) assignments	5153	0	
Proteins with FIGfam assignments	0	0		Proteins with FIGfam assignments	4763	0	

Specialty Genes				Specialty Genes			
	Source	Genes			Source	Genes	
Virulence Factor	Victors	232		Virulence Factor	Victors	232	
Virulence Factor	PATRIC_VF	207		Virulence Factor	PATRIC_VF	207	
Virulence Factor	VFDB	76		Virulence Factor	VFDB	76	
Transporter	TCDB	862		Transporter	TCDB	866	
Drug Target	DrugBank	392		Human Homolog	Human	58	
Drug Target	TTD	61		Drug Target	DrugBank	394	

References

- [Genome Annotation Protocol](#)
- [RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes](#)
- [Genome Annotation Service](#)
- [Genome Annotation Service Quick Reference Guide](#)
- [Genome Annotation Service Tutorial](#)