

BV-BRC Test Report

A8. Service – Similar Genome Finder - Bacteria

Item to test	Similar Genome Finder Service using bacterial genome, fasta and fastq files
URL	https://www.bv-brc.org/app/GenomeDistance
Prerequisites	Bacterial Fasta contig file and FastQ file in the workspace
References	https://www.bv-brc.org/docs/quick_references/services/similar_genome_finder_service.html https://www.bv-brc.org/docs/tutorial/similar_genome_finder/similar_genome_finder.html
Tester(s)	Rebecca Wattam, Maulik Shukla
Test date	07-May-2022 (follow-up from original test)
Test result	Passed

Overview

- Test the Similar Genome Finder Service using exemplar bacterial genome.
- Test input options, i.e., genome id / genome name, fasta contig file, and fastq file as input.
- Test different databases, i.e., reference / representative genomes and all public genomes.
- For each job submitted, verify successful completion of the job and the quality of the search results.
- Test the selection and actions from the search result table.

Test Data

Dataset	Rational	Input Format	Input
Escherichia coli MRSN388634	Workshop example	Genome Name	Escherichia coli MRSN388634
Escherichia coli MRSN388634	Workshop example	Fasta file	Escherichia_coli_MRSN388634.fasta
Escherichia coli MRSN388634	Workshop example	Fastq file	SRR3584989_1.fastq

- All test datasets are available in the following public workspace:
<https://www.bv-brc.org/workspace/BVBRC@patricbrc.org/BVBRC%20Tests/Similar%20Genome%20Finder>

Test Results

- All jobs completed successfully, without any errors.
- All jobs resulted in expected results, finding genomes that were most similar to the query genome.
- The results were presented as an interactive table and the table filters, selection and actions worked as expected.

- For the same query genome, input using genome name, contigs, and fastq file produced similar results.
- Below are a series of screenshots showing submission forms, parameters values and results from various searches.

- Escherichia coli MRSN388634 – genome name as input

Select a genome

SEARCH BY GENOME NAME OR GENOME ID

OR UPLOAD FASTA/FASTQ

ADVANCED OPTIONS

Parameters

MAX HITS:

P VALUE THRESHOLD:

DISTANCE:

SCOPE:

REFERENCE AND REPRESENTATIVE GENOMES

ALL PUBLIC GENOMES

Search

<input type="checkbox"/>	Genome Name	Strain	GenBank Accessions	Size	CDS	Collection Year	Isolation Country	Host Common Name	Distance	P value	K-mer Counts
<input type="checkbox"/>	Escherichia coli IAI39	IAI39	CU928164	5132068	5094				0.0173006	0	533/1000
<input type="checkbox"/>	Escherichia coli UMN026	UMN026	CU928163, CU92 5358200		5198				0.0245258	0	426/1000
<input type="checkbox"/>	Escherichia coli O83:H1 str. NRG 857C	NRG 857C	CP001855, CP00 4894879		4686				0.0250805	0	419/1000
<input type="checkbox"/>	Escherichia coli str. K-12 substr. MG1655	K12	U00096	4639675	4549	1922			0.026819	0	398/1000
<input type="checkbox"/>	Shigella sp. D9	D9	ACDL00000000	4764345	4575			Human	0.0280449	0	384/1000
<input type="checkbox"/>	Shigella sonnei Ss046	Ss046	CP000038, CP00 5055316		5693	1950	China		0.0284967	0	379/1000
<input type="checkbox"/>	Shigella sonnei strain H140920393	H140920393	CXAG00000000	4765223	5067	2014	United Kingdom	Human	0.0297077	0	366/1000
<input type="checkbox"/>	Escherichia coli O104:H4 str. 2011C-3493	2011C-3493	CP003289.1, CP 5437407		5479				0.0297077	0	366/1000
<input type="checkbox"/>	Escherichia coli O157:H7 str. Sakai	O157:H7	BA000007, AB01 5594477		5602	1997	Japan		0.0307758	0	355/1000
<input type="checkbox"/>	Shigella boydii Sb227	Sb227	CP000036, CP00 4646520		5327	1950	China		0.0315792	0	347/1000
<input type="checkbox"/>	Shigella flexneri 2a str. 301	301	AE005674, AF38 4828821		5527	1984	China	Human	0.0316813	0	346/1000
<input type="checkbox"/>	Shigella dysenteriae Sd197	Sd197	CP000034, CP00 4560911		5785	1950	China		0.0331502	0	332/1000
<input type="checkbox"/>	Shigella flexneri 1235-66	1235-66	AKNF00000000	8717850	9204			Human	0.056983	0	178/1000
<input type="checkbox"/>	Escherichia coli strain NCTC9001	NCTC9001	LN831040, LN83 1822639		1836	1952	Denmark	Human	0.0628995	0	154/1000
<input type="checkbox"/>	Escherichia fergusonii ATCC 35469	ATCC 35469	CU928158, CU92 4643861		4373			Human	0.0719629	0	124/1000
<input type="checkbox"/>	Escherichia albertii TW07627	TW07627	ABKX00000000	4746590	4839				0.0824989	0	97/1000
<input type="checkbox"/>	Citrobacter youngae ATCC 29220	ATCC 29220	ABWL00000000	5166865	4709			Human	0.170926	6.47373e-57	14/1000
<input type="checkbox"/>	Salmonella bongori NCTC 12419	NCTC 12419; culture collection SGSC:SARC11	FR877557	4460105	4276				0.174408	5.7527e-53	13/1000
<input type="checkbox"/>	Citrobacter freundii CFNIH1	CFNIH1	CP007557.1, CP 5371331		5151	2012	USA		0.174408	2.00037e-52	13/1000

<input type="checkbox"/>	Genome Name	Strain	GenBank Accessions	Size	CDS	Collection Year	Isolation Country	Host Common Name	Distance	P value	K-mer Counts
<input type="checkbox"/>	Escherichia coli strain MRSN388634	MRSN388634		5211994	5164	2016	USA	Human	0	0	1000/1000
<input type="checkbox"/>	Escherichia coli MRSN388634	MRSN388634		5211994	5158				0	0	1000/1000
<input type="checkbox"/>	Escherichia coli MRSN388634	MRSN388634		5211994	5133				0	0	1000/1000
<input type="checkbox"/>	Escherichia coli strain 2016C-3878	2016C-3878	CP029747,CP02	5272336	5399	2016	USA	Human	2.38274e-05	0	999/1000
<input type="checkbox"/>	Escherichia coli strain AR Bank #0349	AR Bank #0349	CP041996,CP04	5273509	5288			Human	2.38274e-05	0	999/1000
<input type="checkbox"/>	Escherichia coli strain P19_598	P19_598	CP067241,CP06	5208004	5126	2019	Switzerland		0.00229842	0	910/1000
<input type="checkbox"/>	Escherichia coli strain TUM9875	TUM9875	BGMG000000000	5613945	7217	2010	Japan	Human	0.00263042	0	898/1000
<input type="checkbox"/>	Escherichia coli strain 1	1	CAJZPT0000000C	5050962	4961			Human	0.00268639	0	896/1000
<input type="checkbox"/>	Escherichia coli strain 1	1	CAJZPT0000000C	5050962	4961			Human	0.00268639	0	896/1000
<input type="checkbox"/>	Escherichia coli strain 1	1	CAJZPT0000000C	5050962	4961			Human	0.00268639	0	896/1000
<input type="checkbox"/>	Klebsiella pneumoniae strain 170	170	CAJMXR0000000	5050962	4961	2018	Netherlands	Human	0.00268639	0	896/1000
<input type="checkbox"/>	Escherichia coli strain EC-PAG-733	EC-PAG-733	QKXE000000000	5066509	5385	2007	Mexico	Human	0.00274254	0	894/1000
<input type="checkbox"/>	Escherichia coli strain CoR-28	CoR-28	MKFS000000000	5165952	5109	2014	China	Human	0.00279888	0	892/1000
<input type="checkbox"/>	Escherichia coli strain JML100	JML100	BFPN000000000	4915572	4860	2015	Japan	Human	0.00279888	0	892/1000
<input type="checkbox"/>	Escherichia coli strain MH14-138M	MH14-138M	BGPJ000000000	4909017	4809	2014	Viet Nam	Human	0.0028554	0	890/1000
<input type="checkbox"/>	Escherichia coli strain BFR-15808	BFR-15808	SOOU000000000	5042316	5086	2018			0.0028554	0	890/1000
<input type="checkbox"/>	Escherichia coli strain BFR-15808	BFR-15808	SOOU000000000	5042316	5086	2018			0.0028554	0	890/1000
<input type="checkbox"/>	Escherichia coli strain NIFDS_EC2016_31	NIFDS_EC2016_31	WVUU000000000	4971459	4956	2016	South Korea		0.00288374	0	889/1000
<input type="checkbox"/>	Escherichia coli strain IM8c	IM8c	WSXA000000000	4877244	4893	2016	Czech Republic		0.00291212	0	888/1000
<input type="checkbox"/>	Escherichia coli strain CO1	CO1	RXLK000000000	5035070	5125	2016	Ecuador		0.00291212	0	888/1000
<input type="checkbox"/>	Escherichia coli strain PB461	PB461		5068782	4951	2015	Thailand	Human	0.00294054	0	887/1000
<input type="checkbox"/>	Escherichia coli strain 100	100	JAIZJU000000000	5329848	5400	2019	China	Mink	0.00296902	0	886/1000
<input type="checkbox"/>	Escherichia coli strain 100	100	JAIZJU000000000	5329848	5400	2019	China	Mink	0.00296902	0	886/1000
<input type="checkbox"/>	Escherichia coli strain 100	100	JAIZJU000000000	5329848	5400	2019	China	Mink	0.00296902	0	886/1000

1 - 51 of 51 results

< 1 >

- Escherichia coli MRSN388634 – fasta file as input

Select a genome

SEARCH BY GENOME NAME OR GENOME ID

OR UPLOAD FASTA/FASTQ

ADVANCED OPTIONS ▾

Parameters

MAX HITS:

P VALUE THRESHOLD:

DISTANCE:

SCOPE:

REFERENCE AND REPRESENTATIVE GENOMES

ALL PUBLIC GENOMES

Search

<input type="checkbox"/>	Genome Name	Strain	GenBank Accessions	Size	CDS	Collection Year	Isolation Country	Host Common Name	Distance	P value	K-mer Counts
<input type="checkbox"/>	Escherichia coli strain MRSN388634	MRSN388634		5211994	5164	2016	USA	Human	0	0	1000/1000
<input type="checkbox"/>	Escherichia coli MRSN388634	MRSN388634		5211994	5158				0	0	1000/1000
<input type="checkbox"/>	Escherichia coli MRSN388634	MRSN388634		5211994	5133				0	0	1000/1000
<input type="checkbox"/>	Escherichia coli strain 2016C-3878	2016C-3878	CP029747,CP02	5272336	5399	2016	USA	Human	2.38274e-05	0	999/1000
<input type="checkbox"/>	Escherichia coli strain AR Bank #0349	AR Bank #0349	CP041996,CP04	5273509	5288			Human	2.38274e-05	0	999/1000
<input type="checkbox"/>	Escherichia coli strain P19_598	P19_598	CP067241,CP06	5208004	5126	2019	Switzerland		0.00229842	0	910/1000
<input type="checkbox"/>	Escherichia coli strain TUM9875	TUM9875	BGMG00000000	5613845	7217	2010	Japan	Human	0.00263042	0	898/1000
<input type="checkbox"/>	Escherichia coli strain 1	1	CAJZPT000000C	5050962	4961			Human	0.00268639	0	896/1000
<input type="checkbox"/>	Escherichia coli strain 1	1	CAJZPT000000C	5050962	4961			Human	0.00268639	0	896/1000
<input type="checkbox"/>	Escherichia coli strain 1	1	CAJZPT000000C	5050962	4961			Human	0.00268639	0	896/1000
<input type="checkbox"/>	Klebsiella pneumoniae strain 170	170	CAJMKR000000	5050962	4961	2018	Netherlands	Human	0.00268639	0	896/1000
<input type="checkbox"/>	Escherichia coli strain EC-PAG-733	EC-PAG-733	QKXE00000000	5066509	5385	2007	Mexico	Human	0.00274254	0	894/1000
<input type="checkbox"/>	Escherichia coli strain CoR-28	CoR-28	MKFS00000000	5165952	5109	2014	China	Human	0.00279888	0	892/1000
<input type="checkbox"/>	Escherichia coli strain JML100	JML100	BFPN00000000	4915572	4860	2015	Japan	Human	0.00279888	0	892/1000
<input type="checkbox"/>	Escherichia coli strain MH14-138M	MH14-138M	BGPJ00000000	4909017	4809	2014	Viet Nam	Human	0.0028554	0	890/1000
<input type="checkbox"/>	Escherichia coli strain BIR-15808	BIR-15808	SOOU00000000	5042316	5086	2018			0.0028554	0	890/1000
<input type="checkbox"/>	Escherichia coli strain BIR-15808	BIR-15808	SOOU00000000	5042316	5086	2018			0.0028554	0	890/1000
<input type="checkbox"/>	Escherichia coli strain NIFDS_EC2016_31	NIFDS_EC2016_31	WVUU00000000	4971459	4956	2016	South Korea		0.00288374	0	889/1000
<input type="checkbox"/>	Escherichia coli strain IM8c	IM8c	WSXA00000000	4877244	4893	2016	Czech Republic		0.00291212	0	888/1000

- Escherichia coli MRSN388634 – fastq file as input

Select a genome

SEARCH BY GENOME NAME OR GENOME ID

OR UPLOAD FASTA/FASTQ

ADVANCED OPTIONS

Parameters

MAX HITS:

P VALUE THRESHOLD:

DISTANCE:

SCOPE:

REFERENCE AND REPRESENTATIVE GENOMES

ALL PUBLIC GENOMES

Search

<input type="checkbox"/>	Genome Name	Strain	GenBank Accessions	Size	CDS	Collection Year	Isolation Country	Host Common Name	Distance	P value	K-mer Counts
<input type="checkbox"/>	Escherichia coli strain MRSN388634	MRSN388634		5211994	5164	2016	USA	Human	0.0489202	0	218/1000
<input type="checkbox"/>	Escherichia coli strain 2016C-3878	2016C-3878	CP029747,CP02	5272336	5399	2016	USA	Human	0.0489202	0	218/1000
<input type="checkbox"/>	Escherichia coli MRSN388634	MRSN388634		5211994	5158				0.0489202	0	218/1000
<input type="checkbox"/>	Escherichia coli strain AR Bank #0349	AR Bank #0349	CP041996,CP04	5273509	5288			Human	0.0489202	0	218/1000
<input type="checkbox"/>	Escherichia coli MRSN388634	MRSN388634		5211994	5133				0.0489202	0	218/1000
<input type="checkbox"/>	Escherichia coli strain 100	100	JAIZUU0000000	5329848	5400	2019	China	Mink	0.0515304	0	204/1000
<input type="checkbox"/>	Escherichia coli strain 100	100	JAIZUU0000000	5329848	5400	2019	China	Mink	0.0515304	0	204/1000
<input type="checkbox"/>	Escherichia coli strain 100	100	JAIZUU0000000	5329848	5400	2019	China	Mink	0.0515304	0	204/1000
<input type="checkbox"/>	Escherichia coli strain CoR-28	CoR-28	MKFS00000000	5165952	5109	2014	China	Human	0.052117	0	201/1000
<input type="checkbox"/>	Escherichia coli strain B3866	B3866	RWVYJ00000000	5206483	5324				0.052117	0	201/1000
<input type="checkbox"/>	Escherichia coli strain P19_598	P19_598	CP067241,CP06	5208004	5126	2019	Switzerland		0.052117	0	201/1000
<input type="checkbox"/>	Escherichia coli strain SYAB3	SYAB3	LNUE00000000	5241410	5221	2012	China	Human	0.0523149	0	200/1000
<input type="checkbox"/>	Escherichia coli strain SYAB1	SYAB1	CM003776,LNUI	5241449	5213	2012	China	Human	0.0523149	0	200/1000
<input type="checkbox"/>	Escherichia coli strain PB185	PB185		5145397	5088	2015	Thailand	Human	0.0523149	0	200/1000
<input type="checkbox"/>	Escherichia coli strain ZTA15/00685EB1	ZTA15/00685EB1		5202454	5211	2015	Spain	Pig	0.0523149	0	200/1000
<input type="checkbox"/>	Escherichia coli strain SYAU2	SYAU2	CM003779,LNUI	5224350	5160	2012	China	Human	0.0525139	0	199/1000
<input type="checkbox"/>	Escherichia coli strain SYAU3	SYAU3	CM003777,LNUI	5174089	5119	2012	China	Human	0.0525139	0	199/1000
<input type="checkbox"/>	Escherichia coli strain SYAU1	SYAU1	LNUG00000000	5187611	5140	2012	China	Human	0.0525139	0	199/1000
<input type="checkbox"/>	Escherichia coli strain SYAB2	SYAB2	CM003775,LNUI	5237295	5209	2012	China	Human	0.0525139	0	199/1000
<input type="checkbox"/>	Escherichia coli strain SYAB4	SYAB4	CM003774,LNUI	5229355	5208	2012	China	Human	0.0525139	0	199/1000
<input type="checkbox"/>	Escherichia coli strain MRV15-117	MRV15-117	AP017617,AP01	5295071	5224	2012	Japan	Cow	0.0525139	0	199/1000
<input type="checkbox"/>	Escherichia coli strain B222	B222	NRTO00000000	5321608	5341	2016	China	Human	0.0525139	0	199/1000
<input type="checkbox"/>	Escherichia coli strain PB356	PB356		5155675	5125	2015	Thailand	Human	0.0525139	0	199/1000
<input type="checkbox"/>	Escherichia coli strain NIHE15-1569	NIHE15-1569	BPUH00000000	5190348	5205	2015	Viet Nam	Human	0.0525139	0	199/1000

1 - 51 of 51 results

< 1 >

References

- [Similar Genome Finder Service Quick Reference Guide](#)
- [Similar Genome Finder Service Tutorial](#)