BV-BRC Test Report

A13. Service – Variation Analysis - Bacteria

Item to test	Variation Analysis Service using bacterial read files and SRA accessions
URL	https://www.bv-brc.org/app/Variation
Prerequisites	Bacterial Fasta contig files in Workspace
References	https://www.bv-brc.org/docs/quick_references/services/variation_analysis_service.html https://www.bv-brc.org/docs/tutorial/variation_analysis/variation_analysis.html
Tester(s)	Rebecca Wattam, Maulik Shukla
Test date	08-May-2022 (follow-up from original test)
Test result	**Passed**

Overview

- Test the Variation Analysis Service using exemplar bacterial datasets.
- Test input options, i.e., read files in FASTQ format and SRA accessions.
- Test different alignment strategies, i.e. BWA-mem, BWA-mem-strict, Bowtie2 and LAST. Test different SNP callers, such as FreeBayes and SAMtools.
- For each job submitted, verify successful completion of the job, presence of output files, and the quality of results.
- Review SNPs and alignment files in interactive the genome browser.

Test Data

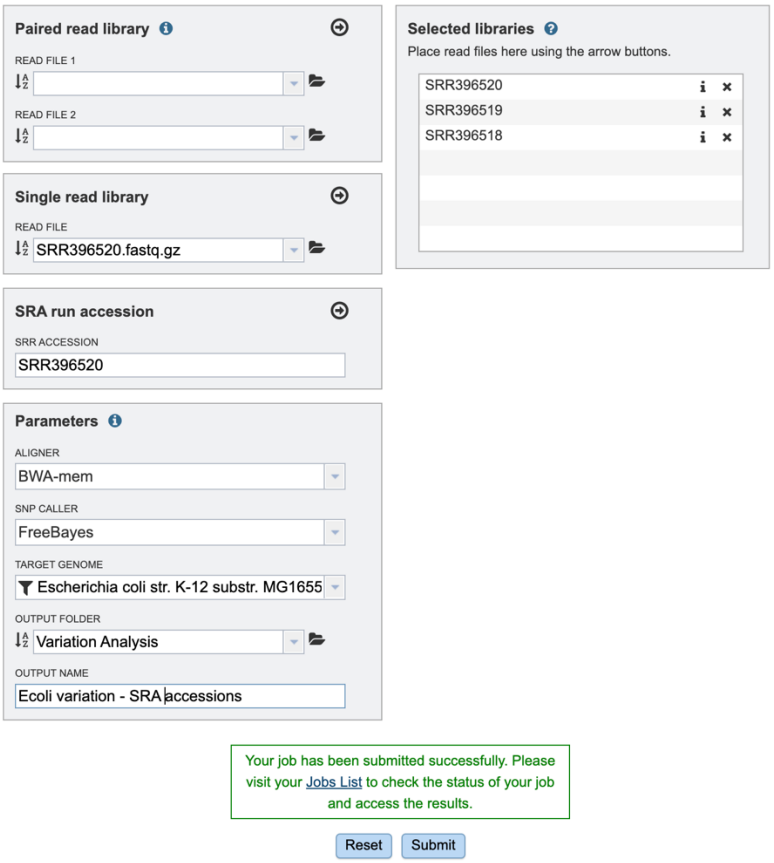
Dataset	Rational	Input Format	Input
Escherichia coli genomes	Workshop dataset	Fastq files and SRA accessions	SRR396518.fastq.gz SRR396519.fastq.gz SRR396520.fastq.gz

- 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/Variation%20Analysis

Test Results

- All variation analysis jobs completed successfully, without any errors.
- All jobs resulted in expected output files in corresponding job output directory, including BAM files, VCF files, and variation report.
- The variation report was informative and provided list of SNPs and indels as an HTML table. The table shown SNPs, their position, corresponding gene, NA and AA position, NA/AA change, and annotation of SNP effect and its significance.

- The link to genome browser worked as expected. It loaded all BAM and VCF files as separate tracks and allowed zooming in to a single SNP level. Clicking on a SNP provided all the details and annotation about that SNP.
- 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/Variation%20Analysis
- Below are a series of screenshots showing successful completion of the jobs, availability of the result files in the workspace, the variation table and genome browser views.



Status	ID	Service	Output Name	Submit	Start	Completed
completed	7747607	Variation Analysis	Ecoli variation - fastq files	5/8/22, 11:35 AM	5/8/22, 11:35 AM	5/8/22, 12:03 PM
completed	7747608	Variation Analysis	Ecoli variation - SRA accessions	5/8/22, 11:36 AM	5/8/22, 11:36 AM	5/8/22, 12:00 PM
completed	7747609	Variation Analysis	Ecoli variation - bwa-mem-strict	5/8/22, 11:37 AM	5/8/22, 11:37 AM	5/8/22, 11:49 AM
completed	7747610	Variation Analysis	Ecoli variation - bowtie2	5/8/22, 11:37 AM	5/8/22, 11:37 AM	5/8/22, 11:53 AM
completed	7747615	Variation Analysis	Ecoli variation - last	5/8/22, 11:40 AM	5/8/22, 11:40 AM	5/8/22, 12:17 PM
completed	7747616	Variation Analysis	Ecoli variation - samtools	5/8/22, 11:40 AM	5/8/22, 11:40 AM	5/8/22, 11:57 AM

BVBRC / BVBRC Tests / Variation Analysis (9 items)

UPLOAD ADD FOLDER SHOW HIDDEN

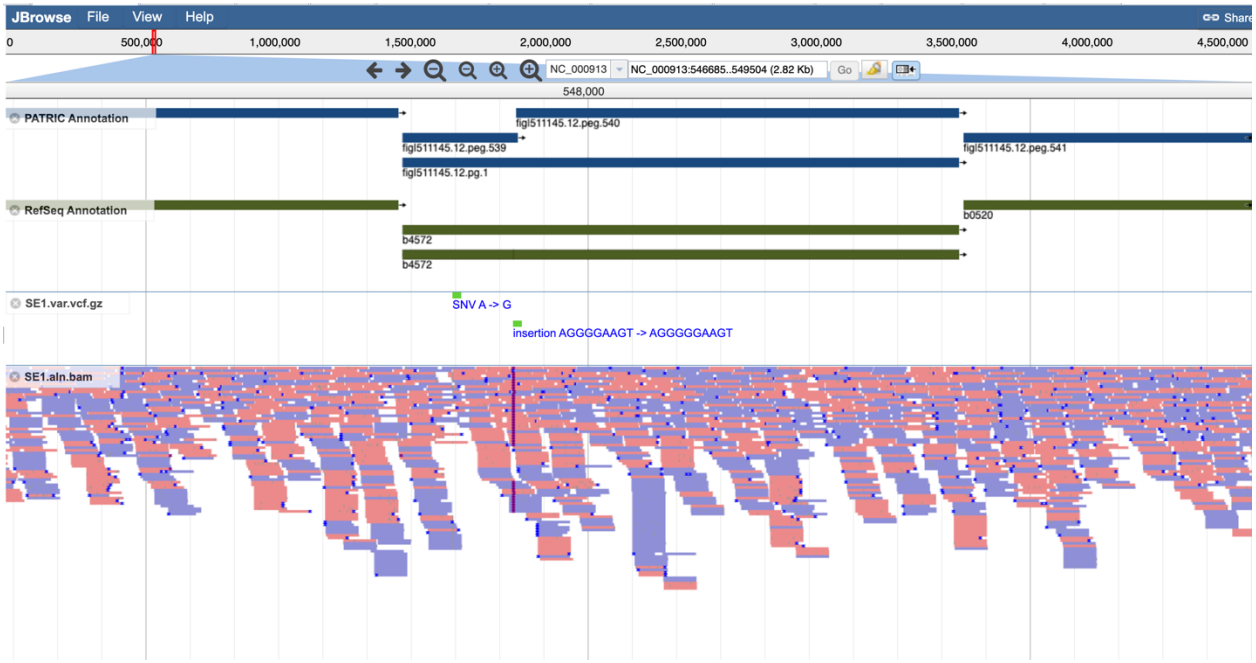
Name	Size	Owner	Members	Created
Parent folder			-	
SRR396520.fastq.gz	228.6 MB	me	Public	5/8/22, 11:33 AM
SRR396519.fastq.gz	276.7 MB	me	Public	5/8/22, 11:33 AM
SRR396518.fastq.gz	248.2 MB	me	Public	5/8/22, 11:33 AM
Ecoli variation - bwa-mem-strict	10.1 kB	me	Public	5/8/22, 11:49 AM
Ecoli variation - bowtie2	9.9 kB	me	Public	5/8/22, 11:53 AM
Ecoli variation - samtools	9.9 kB	me	Public	5/8/22, 11:57 AM
Ecoli variation - SRA accessions	10.1 kB	me	Public	5/8/22, 12:00 PM
Ecoli variation - fastq files	10.3 kB	me	Public	5/8/22, 12:03 PM
Ecoli variation - last	9.8 kB	me	Public	5/8/22, 12:17 PM

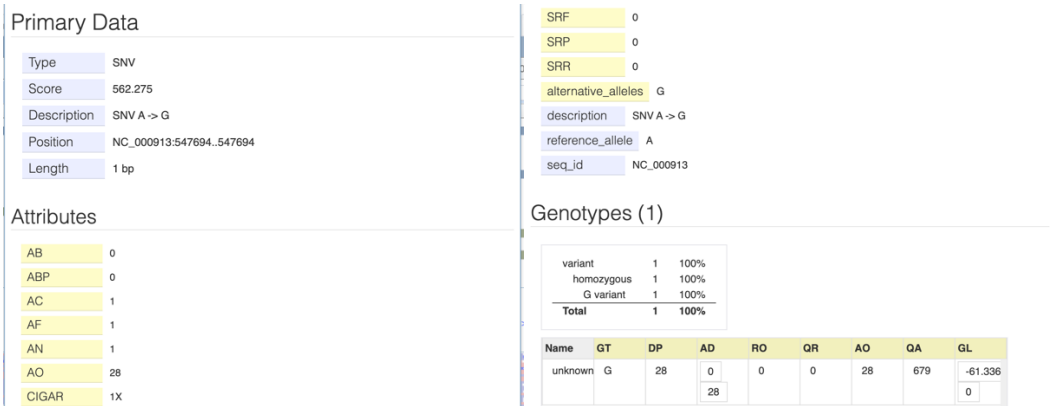
Name	Size	Owner	Members	Created
Parent folder			-	
SE1.aln.bam	191.8 MB	me	Public	5/8/22, 12:00 PM
SE1.aln.bam.bai	14.2 kB	me	Public	5/8/22, 12:00 PM
SE1.consensus.fa	4.7 MB	me	Public	5/8/22, 12:00 PM
SE1.var.annotated.tsv	3.0 kB	me	Public	5/8/22, 12:00 PM
SE1.var.snpEff.vcf	15.2 kB	me	Public	5/8/22, 12:00 PM
SE1.var.vcf	12.9 kB	me	Public	5/8/22, 12:00 PM
SE1.var.vcf.gz	3.4 kB	me	Public	5/8/22, 12:00 PM
SE1.var.vcf.gz.tbi	269 B	me	Public	5/8/22, 12:00 PM
SE2.aln.bam	209.2 MB	me	Public	5/8/22, 12:00 PM
SE2.aln.bam.bai	14.3 kB	me	Public	5/8/22, 12:00 PM
SE2.consensus.fa	4.7 MB	me	Public	5/8/22, 12:00 PM
SE2.var.annotated.tsv	5.8 kB	me	Public	5/8/22, 12:00 PM
SE2.var.snpEff.vcf	22.9 kB	me	Public	5/8/22, 12:00 PM
SE2.var.vcf	19.3 kB	me	Public	5/8/22, 12:00 PM
SE2.var.vcf.gz	4.6 kB	me	Public	5/8/22, 12:00 PM
SE2.var.vcf.gz.tbi	329 B	me	Public	5/8/22, 12:00 PM
SE3.aln.bam	178.2 MB	me	Public	5/8/22, 12:00 PM
SE3.aln.bam.bai	14.1 kB	me	Public	5/8/22, 12:00 PM
SE3.consensus.fa	4.7 MB	me	Public	5/8/22, 12:00 PM

BVBRC / BVBRC Tests / Variation Analysis / .Ecoli variation - SRA accessions / all.var.html

html file: all.var.html

Samples	Contig	Pos	Ref	Var	Score	Var_cov	Var_frac	Type	Ref_nt	Var_nt	Ref_nt_pos_change	Ref_aa_pos_change	Frameshift	Gene_ID	Locus_tag	Gene_name
1:SE2	NC_000913	1049100	C	T	11.4092	2.0	0.09									
1:SE2	NC_000913	1049669	A	G	49.2323	3.0	0.13	Synon	gga	ggG	294A>G	Gly98Gly		fig|511145.12.peg.1024	b0988	insB
1:SE2	NC_000913	1049673	A	G	51.2844	2.0	0.10	Nonsyn	aag	Gag	298A>G	Lys100Glu		fig|511145.12.peg.1024	b0988	insB
1:SE2	NC_000913	1049680	T	C	39.3747	2.0	0.11	Nonsyn	ctg	cCg	305T>C	Leu102Pro		fig|511145.12.peg.1024	b0988	insB
1:SE2	NC_000913	1049682	T	G	43.4713	2.0	0.12	Nonsyn	tcg	Gcg	307T>G	Ser103Ala		fig|511145.12.peg.1024	b0988	insB
1:SE2	NC_000913	1049685	T	G	42.3477	2.0	0.17	Nonsyn	ttc	Gtc	310T>G	Phe104Val		fig|511145.12.peg.1024	b0988	insB
1:SE3	NC_000913	1568836	C	G	57.3728	2.0	0.13	Nonsyn	ggc	Cgc	1234G>C	Gly412Arg		fig|511145.12.peg.1560	b1493	gadB
1:SE1	NC_000913	1568846	G	A	19.4249	2.0	0.09	Synon	gcc	gcT	1224C>T	Ala408Ala		fig|511145.12.peg.1560	b1493	gadB
1:SE3	NC_000913	1617114	TGCAACTAATTACTTGCCAGGG	TG	528.949	24.0	0.96	Deletion								
2:SE1,SE2	NC_000913	1617423	C	T	1011.0	50.5	0.96	Nonsyn	cgc	Tgc	280C>T	Arg94Cys		fig|511145.12.peg.1599	b1530	marR
1:SE3	NC_000913	1633932	C	A	58.363	2.0	0.12	Nonsyn	gat	Tat	460G>T	Asp154Tyr		fig|511145.12.peg.1618		
1:SE3	NC_000913	1718907	GCAATTGTATTGCTAAAAC	GC	341.821	16.0	0.94	Deletion								
3:SE1,SE2,SE3	NC_000913	1903785	G	A	1667.44	72.3	0.99	Nonsyn	ggt	gAt	74G>A	Gly25Asp		fig|511145.12.peg.1898	b1821	yebN





References

- [Variation Analysis Service Quick Reference Guide]
- [Variation Analysis Service Tutorial]