



“Snippy” Aracı ile Varyant Analizi

Hazırlayanlar: Şahin Öztürk



Nucleotide

Nucleotide

Search

Advanced

Help

GenBank

Send to

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-genome

NCBI Reference Sequence: NC_045512.2

[FASTA](#) [Graphics](#)

Go to

LOCUS NC_045512 29903 bp ss-RNA linear VRL 18-JUL-2020
DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome.
ACCESSION NC_045512
VERSION NC_045512.2
DBLINK BioProject: [PRJNA485481](#)
KEYWORDS RefSeq.
SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
ORGANISM [Severe acute respiratory syndrome coronavirus 2](#)
Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes; Nidovirales; Coronavirinae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus.
REFERENCE 1 (bases 1 to 29903)
AUTHORS Wu,F., Zhao,S., Yu,B., Chen,Y.M., Wang,W., Song,Z.G., Hu,Y., Tao,Z.W., Tian,J.H., Pei,Y.Y., Yuan,M.L., Zhang,Y.L., Dai,F.H., Liu,Y., Wang,Q.M., Zheng,J.J., Xu,L., Holmes,E.C. and Zhang,Y.Z.
TITLE A new coronavirus associated with human respiratory disease in China
JOURNAL Nature 579 (7798), 265-269 (2020)

- Complete Record
 - Coding Sequences
 - Gene Features
- Choose Destination
- File
 - Clipboard
 - Collections
 - Analysis Tool

Download 1 item.

Format

GenBank

Show GI

Create File

Related information

[Assembly](#)

[BioProject](#)

[Protein](#)

[PubMed](#)

Varyant analizinde kullanacağımız Covid okumaları için referans genom olarak kullanmak üzere NCBI veri bankasından covid'e ait "NC_045512.2" genomunu "genbank" formatında indiriyoruz.

Tools

search tools

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Datamash

GENOMIC FILE MANIPULATION

FASTA/FASTQ



















FASTQ Quality Control

SAM/BAM




Download from web or upload from disk

Regular Composite Collection Rule-based

You added 3 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
 sequence.gb	76.6 KB	Auto-de...  	unspecified (?) 		0% 
 ERR4082748_2.fastq	101 MB	Auto-de...  	unspecified (?) 		0% 
 ERR4082748_1.fastq	101.1 MB	Auto-de...  	unspecified (?) 		0% 

Type (set all): Auto-detect   Genome (set all): unspecified (?) 

 Choose local files  Choose remote files  Paste/Fetch data **Start** Pause Reset Close

- Kullanacağımız "COVID" okumalarını ve NCBI' dan indirmiş olduğumuz Covid'e ait "Genbank" dosyasını "choose local files" butonu ile bilgisayarımızdan galaxy'e yüklemek üzere seçiyoruz.
- Ardından "Start" ile yüklemeyi başlatıyoruz.

Galaxy

Workflow Visualize Shared Data Help User

Using 0%

Tools

Snippy

Upload Data

Show Sections

snippy Snippy finds SNPs between a haploid reference genome and your NGS sequence reads.

snippy-core Combine multiple Snippy outputs into a core SNP alignment

snippy-clean_full_aln Replace any non-standard sequence characters in snippy 'core.full.aln' file.

WORKFLOWS

All workflows

snippy Snippy finds SNPs between a haploid reference genome and your NGS sequence reads. (Galaxy Version 4.6.0+galaxy0)

Will you select a reference genome from your history or use a built-in index?

Use a genome from history and build index

Built-ins were indexed using default options. See 'Indexes' section of help below. If you would like to perform self-mapping select 'history' here, then choose your input file as reference.

Use the following dataset as the reference sequence

1: sequence.gb

You can upload a FASTA or FASTQ sequence to the history and use it as reference

Single or Paired-end reads

Paired

Select between paired and single end data

Select first set of reads

2: ERR4082748_1.fastq

Specify dataset with forward reads

Select second set of reads

3: ERR4082748_2.fastq

History

search datasets

Snippy

212 MB

3

3 : ERR4082748_2.fastq

2 : ERR4082748_1.fastq

1 : sequence.gb

- Dosyalar yüklendikten sonra varyant analizi yapmak için araçlar kısmında “snippy” aracını aratıp, ok ile gösterdiğimiz aracı açıyoruz.
- “Reference File” kısmında yüklemiş olduğumuz referans genomu seçiyoruz.
- İki yönlü okumalarımız olduğundan, “Paired” seçeneğini seçiyoruz.
- İlk olarak “Select first set of reads” kısmında ileri yönlü fastq dosyamızı seçiyoruz.
- Ardından “Select second set of reads” kısmında da geri yönlü olan ikinci fastq dosyamızı seçiyoruz.

Tools ☆ ▾

Snippy x

Upload Data

Show Sections

snippy Snippy finds SNPs between a haploid reference genome and your NGS sequence reads.

snippy-core Combine multiple Snippy outputs into a core SNP alignment

snippy-clean_full_aln Replace any non-standard sequence characters in snippy 'core.full.aln' file.

WORKFLOWS

All workflows

3: ERR4082748_2.fastq

Specify dataset with reverse reads

Advanced parameters 👁

Output selection

Select/Unselect all

- The final annotated variants in VCF format (Varyantlar VCF dosya formatında olsun.)
- The variants in GFF3 format
- A simple tab-separated summary of all the variants (Tüm varyantlar sekmelerle basitçe özetlensin.)
- A summary of the samples and mapping
- A log file with the commands run and their outputs
- A version of the reference but with - at position with depth=0 and N for 0 to depth to --mincov (does not have variants)
- A version of the reference genome with all variants instantiated
- The alignments in BAM format. Note that multi-mapping and unmapped reads are not present.
- Zipped files needed for input into snippy-core ("snippy-core" aracında kullanabilmek için sıkıştırılmış dosyalar oluşturulsun.)

Job Resource Parameters

Use default job resource parameters ▾

Email notification

No

Send an email notification when the job completes.

Execute

History + = ▾


search datasets ▾ x

Snippy ✎

212 MB 📍 3 ↻

- 3 : ERR4082748_2.fastq 👁 ✎ 🗑
- 2 : ERR4082748_1.fastq 👁 ✎ 🗑
- 1 : sequence.gb 👁 ✎ 🗑

Elde edeceğimiz “VCF” dosyasında inceleyeceğimiz kısımlar:

- CHROME: Varyantın bulunduğu kromozomu tanımlar.
 - POS: Varyantın bulunduğu kromozom pozisyonunu tanımlar.
 - REF: Referans genomda bulunan nükleotidi tanımlar.
 - ALT: Hizalanan genomdaki nükleotidi tanımlar.
 - EVIDENCE: Hizalanan fragmentlerdeki nükleotitlerin kaçının referans genomla eşleştiğini veya kaçının farklı eşleştiğini ifade eder.
 - FTYPE: Hizalanan fragment tipini belirtir.
 - NT_POS: Nükleotit pozisyonunu belirtir.
 - AA_POS: Aminoasit pozisyonunu belirtir.
 - EFFECT: Genomdaki varyantın ne olduğunu ifade eder.
- 

Galaxy Workflow Visualize Shared Data Help User Using 0%

Tools ☆

search tools x

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Datamash

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

Convert Formats

```
##commandline="freebayes -p 2 -P 0 -C 2 -F 0.05 --min-coverage 10 --min-repeat-entropy 1.0 -q 13 -m 60 --strict-vcf -f reference/ref.fa snps.bam -
##INFO= <ID=DP,Number=1,Type=Integer,Description="Total read depth at the locus">
##INFO= <ID=RO,Number=1,Type=Integer,Description="Count of full observations of the reference haplotype.">
##INFO= <ID=AO,Number=A,Type=Integer,Description="Count of full observations of this alternate haplotype.">
##INFO= <ID=QR,Number=1,Type=Integer,Description="Reference allele quality sum in phred">
##INFO= <ID=QA,Number=A,Type=Integer,Description="Alternate allele quality sum in phred">
##INFO= <ID=AB,Number=A,Type=Float,Description="Allele balance at heterozygous sites: a number between 0 and 1 representing the ratio of rea
##INFO= <ID=TYPE,Number=A,Type=String,Description="The type of allele, either snp, mnp, ins, del, or complex.">
##FORMAT= <ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT= <ID=GL,Number=G,Type=Float,Description="Genotype Likelihood, log10-scaled likelihoods of the data given the called genotype for e
##FORMAT= <ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT= <ID=RO,Number=1,Type=Integer,Description="Reference allele observation count">
##FORMAT= <ID=QR,Number=1,Type=Integer,Description="Sum of quality of the reference observations">
##FORMAT= <ID=AO,Number=A,Type=Integer,Description="Alternate allele observation count">
##FORMAT= <ID=QA,Number=A,Type=Integer,Description="Sum of quality of the alternate observations">
##bcftools_viewVersion=1.13+htslib-1.13
##bcftools_viewCommand=view --include 'FMT/GT="1/1" && QUAL>=100.0 && FMT/DP>=10 && (FMT/AO)/(FMT/DP)>=0.9' snps.raw.vcf; Date=M
##bcftools_annotateVersion=1.13+htslib-1.13
##bcftools_annotateCommand=annotate --remove ^INFO/TYPE,^INFO/DP,^INFO/RO,^INFO/AO,^INFO/AB,^FORMAT/GT,^FORMAT/DP,^FORMAT/
##SnpEffVersion="5.0e (build 2021-03-09 06:01), by Pablo Cingolani"
##SnpEffCmd="SnpEff -noStats -no-downstream -no-upstream -no-utr ref snps.filt.vcf "
##INFO= <ID=ANN,Number=.,Type=String,Description="Functional annotations: 'Allele | Annotation | Annotation_Impact | Gene_Name | Gene_ID | F
##INFO= <ID=LOF,Number=.,Type=String,Description="Predicted loss of function effects for this variant. Format: 'Gene_Name | Gene_ID | Number_o
##INFO= <ID=NMD,Number=.,Type=String,Description="Predicted nonsense mediated decay effects for this variant. Format: 'Gene_Name | Gene_ID
#CHROM POS ID REF ALT QUAL FILTER INFO
NC_045512 241 . C T 1628.74 . AB=0;AO=54;DP=54;QA=1990;QR=0;RO=0;TYPE=snp;ANN=T|intergenic_region|MC
NC_045512 3037 . C T 844.54 . AB=0;AO=27;DP=27;QA=1045;QR=0;RO=0;TYPE=snp;ANN=T|synonymous_variant|
```

History +

search datasets x

Snippy ✎

213 MB 6

6 : snippy on data 3, data 2, and data 1 dir for snippy cor e

5 : snippy on data 3, data 2, and data 1 snps table

4 : snippy on data 3, data 2, and data 1 snps vcf file

3 : ERR4082748_2.fastq

2 : ERR4082748_1.fastq

1 : sequence.gb

Elde ettiğimiz "VCF" dosyasında referans genomda olan ve hizalanan DNA'da karşısına gelen nükleotitleri gözlemleyerek varyant analizimizi yapabiliriz.

Tools

search tools

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Datamash

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

VCF/BCF

CHROM	POS	TYPE	REF	ALT	EVIDENCE	FTYPE	STRAND	NT_POS	AA_POS	EFFECT
NC_045512	241	snp	C	T	T:54 C:0	5'UTR	+			intergenic_region n.241C>T
NC_045512	3037	snp	C	T	T:27 C:0	mat_peptide	+	2772/21290	924/7095	synonymous_variant c.2772C>T
NC_045512	14408	snp	C	T	T:54 C:1	mat_peptide	+	14143/21290	4715/7095	synonymous_variant c.14143C>T
NC_045512	16289	snp	C	T	T:25 C:0	mat_peptide	+	16024/21290	5342/7095	missense_variant c.16024C>T p.L
NC_045512	23403	snp	A	G	G:34 A:0	CDS	+	1841/3822	614/1273	missense_variant c.1841A>G p.A
NC_045512	25429	snp	G	T	T:43 G:1	CDS	+	37/828	13/275	missense_variant c.37G>T p.Val1

History

search datasets

Snippy

213 MB

6

6 : snippy on data 3, data 2, and data 1 dir for snippy core

5 : snippy on data 3, data 2, and data 1 snps table

4 : snippy on data 3, data 2, and data 1 snps vcf file

3 : ERR4082748_2.fastq

2 : ERR4082748_1.fastq

1 : sequence.gb

Bu dosyamızda da aynı şekilde referans genomda yer alan ve hizalanan DNA'da karşısına gelen nükleotitleri gözlemleyerek varyant analizimizi yapabiliriz.

Tools

search tools

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Datamash

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

ua The global community has created a **continuously updated list** of laboratories that can host Ukrainian scientists at all career levels. If your lab can host a scientist -- add your name to the list here. In addition, Galaxy Project has a number of positions at its EU and US sites. Contact us at ukraine@galaxyproject.org

Світова наукова спільнота створила **список лабораторій**, що постійно оновлюється та які можуть прийняти українських науковців усіх рівнів, у тому числі й аспірантів. Якщо ваша лабораторія має можливість запросити -- ви можете додати ваше ім'я до списку тут. Крім того, Galaxy Project має відкриті вакансії у своїх європейських та американських осередках. Пишіть нам на ukraine@galaxyproject.org

Научное сообщество создало постоянно обновляемый **список лабораторий**, которые могут принять украинских ученых (включая аспирантов). К тому же, Galaxy Project имеет открытые позиции на своих европейских и американских сайтах. Контактуйте нас используя ukraine@galaxyproject.org

Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed.

The Galaxy Community and UseGalaxy.* initiative lost one of its

İndirmeler

Galaxy6-[snippy_on_data_3_data_2_and_data_1_dir_fo...z...
İndirme tamamlandı.

Temizle

Tümünü göster

213 MB

6

6 : snippy on data 3, data 2,
and data 1 dir for snippy cor
e

5 : snippy on data 3, data 2,
and data 1 snps table

4 : snippy on data 3, data 2,
and data 1 snps vcf file

3 : ERR4082748_2.fastq

2 : ERR4082748_1.fastq

1 : sequence.gb

Bu işlemler sonucunda "Snippy-core" aracında kullanabileceğimiz sıkıştırılmış dosyayı da elde etmiş olduk.