

# IBT 432 Aplikasi Bioinformatika

## Praktek: Analisis komparasi genomika

**Riza Arief Putranto**

# Identifikasi gen penyebab virulensi *Mycobacterium tuberculosis* strain pada TK-K-009 dengan komparasi pada strain pembanding

REVIEW

Virulence 4:1, 3–66; January 1, 2013; © 2013 Landes Bioscience

## Virulence factors of the *Mycobacterium tuberculosis* complex

Marina A. Forrellad,<sup>1</sup> Laura I. Klepp,<sup>1</sup> Andrea Gioffré,<sup>1</sup> Julia Sabio y García,<sup>1</sup> Hector R. Morbidoni,<sup>2</sup> María de la Paz Santangelo,<sup>1</sup> Angel A. Cataldi<sup>1</sup> and Fabiana Bigi<sup>1,\*</sup>

<sup>1</sup>Instituto de Biotecnología; CICVyA-INTA; Buenos Aires, Argentina; <sup>2</sup>Cátedra de Microbiología; Facultad de Ciencias Médicas; Universidad Nacional de Rosario; Rosario, Argentina

**Keywords:** *Mycobacterium tuberculosis*, virulence factors, virulence, pathogen, virulence genes

**Abbreviations:** ABC-transporter, ATP-binding cassette transporter; BMDM, bone marrow-derived macrophages; CFU, colony-forming unit; DAT, di-acylated trehalose; DC, dendritic cell; ECF, extracytoplasmic function; IFN- $\gamma$ , gamma interferon; LAM, lipoarabinomannane; Lpp, lipoprotein; MAMTs, mycolic acid methyl transferases; MHC-II, major histocompatibility complex class II; MTBC, *Mycobacterium tuberculosis* complex; ORF, open reading frame; PAT, poly-acylated trehalose; PDIM, phthiocerol dimycocerosate; PGL, phenolic glycolipid; *p*-HBADs, *p*-hydroxybenzoic acid derivatives; PIM, phosphatidylinositol mannoside; RD, region of difference; ROS, reactive oxygen species; RNS, reactive nitrogen species; SL, sulfolipid; STM, signature-tagged mutagenesis; TAG, triacylglycerides; TAT, tri-acylated trehalose; TB, tuberculosis; TCS, two-component system; TLR2, Toll-like receptor 2; TDM, trehalose di-mycolates; TMM, trehalose mono-mycolates; TNF- $\alpha$ , tumor necrosis factor-alpha; TraSH, transposon site hybridization; T7SS, type VII secretion system

Strain pembanding: *Mycobacterium bovine*

Strain target: *Mycobacterium tuberculosis* TKK\_04\_0029

# 1. Akses data transkriptomika dan genomika dalam *database* European Nucleotide Archive (ENA)

www.ebi.ac.uk/ena

EMBL-EBI

Services Research Training About us

ENA  
European Nucleotide Archive

Search

Examples: [BN000085](#), [histone](#)

Advanced Search Sequence

Home Search & Browse Submit & Update Software About ENA Support

## European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation. [More about ENA](#)

Access to ENA data is provided through the browser, through search tools, large scale file download and through the API.

### Text Search

Search

Examples: [BN000065](#), [histone](#)

Advanced search

### Sequence Search

Enter or paste a nucleotide sequence or accession number

### Popular

- [Submit and update](#)
- [Sequence submissions](#)
- [Genome assembly submissions](#)
- [Submitting environmental sequences](#)
- [Citing ENA data](#)
- [Rest URLs for data retrieval](#)
- [Rest URLs to search ENA](#)

### Latest ENA news

**03 Apr 2017:** [ENA Release 131](#)

Release 131 of ENA's assembled/annotated sequences now available

16:09  
20-Apr-17

## Interface dari European Nucleotide Archive

Terhubung dengan database lain seperti NCBI dan DDBJ

ENA relatif mudah diakses dengan pengelompokan data yang baik

www.ebi.ac.uk/ena/data/ X www.ebi.ac.uk/ena/data/ X De Novo Assembly and X www.ebi.ac.uk/ena/data/ X European Nucleotide Arc X Galaxy X

www.ebi.ac.uk/ena

Services Research Training About us

EMBL-EBI

ENA  
European Nucleotide Archive

Search

Advanced  
Sequence

Home Search & Browse Submit & Update Software About ENA Support

## European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation. [More about ENA](#)

Access to ENA data is provided through the browser, through search tools, large scale file download and through the API.

### Text Search

Search

Examples: [BN000065](#), [histone](#)

Advanced search

### Sequence Search

Enter or paste a nucleotide sequence or accession number

### Popular

- Submit and update
- Sequence submissions
- Genome assembly submissions
- Submitting environmental sequences
- Citing ENA data
- Rest URLs for data retrieval
- Rest URLs to search ENA

### Latest ENA news

**03 Apr 2017:** [ENA Release 131](#)

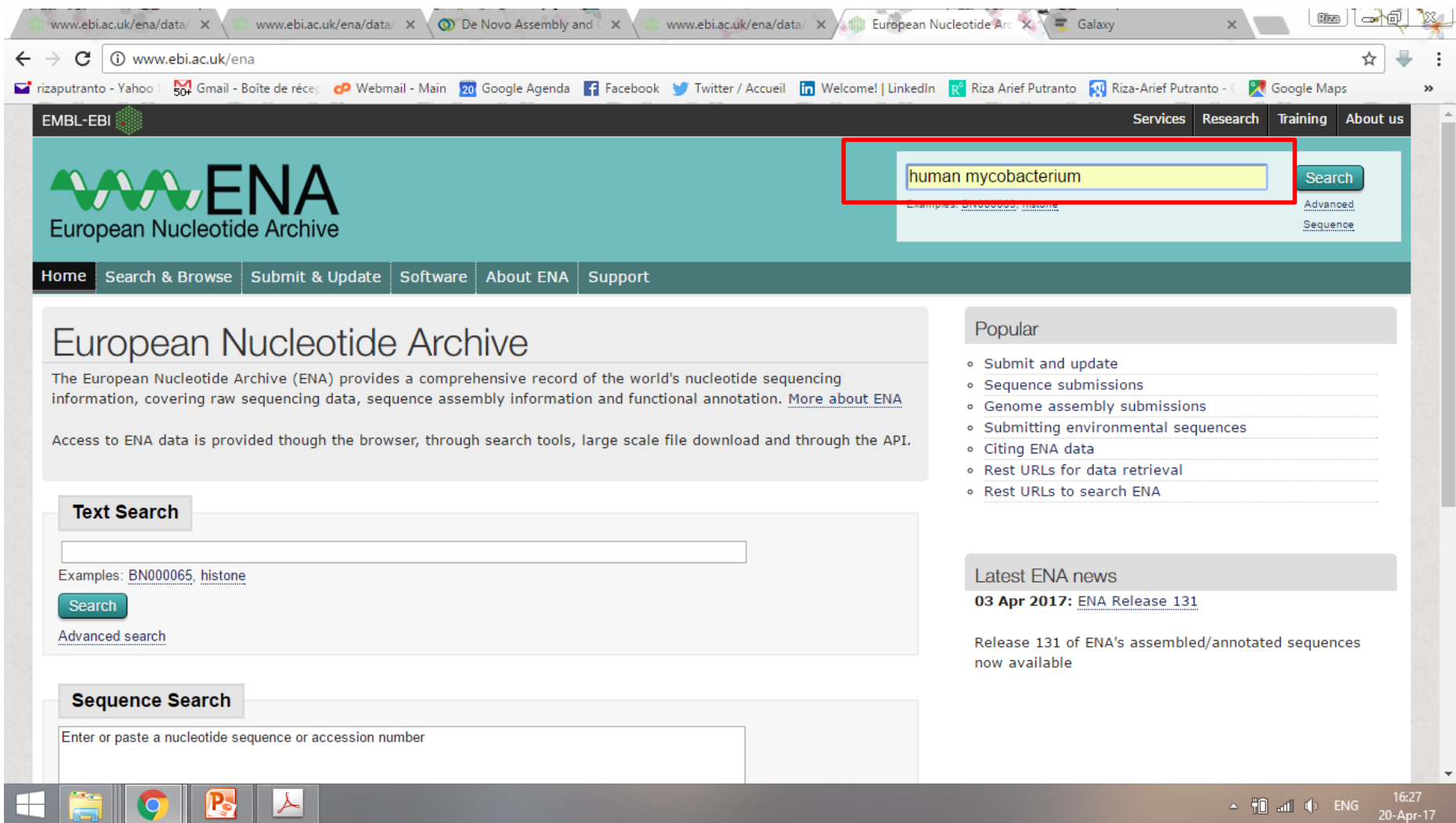
Release 131 of ENA's assembled/annotated sequences now available

www.ebi.ac.uk/ena/browse/data-retrieval-rest

16:06  
20-Apr-17

## Step 1.

Gunakan kata kunci untuk mencari database transkriptomik dan genomik  
Gunakan kata kunci **“Human Mycobacterium”**



**Step 2.** Kata kunci “**Human Mycobacterium**”  
Ketikkan kata kunci ke dalam **Search box**

www.ebi.ac.uk/ena/data/ X www.ebi.ac.uk/ena/data/ X De Novo Assembly and X www.ebi.ac.uk/ena/data/ X www.ebi.ac.uk/ena/data/ X Galaxy

www.ebi.ac.uk/ena/data/search?query=human+mycobacterium

rizaputranto - Yahoo! 50+ Gmail - Boîte de réception Webmail - Main 20 Google Agenda Facebook Twitter / Accueil Welcome! | LinkedIn Riza Arief Putranto Riza-Arief Putranto - Google Maps

### Search results for *human mycobacterium*

Show more data from EMBL-EBI

- Assembly**  
Assembly (70)
- Sequence**  
Sequence (Release) (5,347)
- Contig set**  
Genome assembly contig set (44)
- Coding**  
Coding (Release) (251,666)
- Non-coding**  
Non-coding (Release) (3,014)
- Read**  
Experiment (8)
- Study**  
Study (61)  
Study (Sequence) (150)
- Sample**  
Sample (439)

**Assembly (70 results found)**  
GCA\_000026685.1 ASM2668v1 assembly for Mycobacterium leprae Br4923  
View all 70 results

**Sequence (Release) (5,347 results found)**  
AF009359 Mycobacterium tuberculosis orfH-200 gene, partial cds.  
View all 5,347 results

**Genome assembly contig set (44 results found)**  
JLSX01000000 Mycobacterium tuberculosis TKK\_04\_0029, WGS project JLSX01000000 data  
View all 44 results

**Coding (Release) (251,666 results found)**  
AAK98161 Human immunodeficiency virus 1 partial envelope glycoprotein  
View all 251,666 results

**Non-coding (Release) (3,014 results found)**  
MLIL01000001.1:642491..642567:tRNA Mycobacterium chelonae tRNA-Leu  
View all 3,014 results

**Experiment (8 results found)**

**Step 3.** Kata kunci “Human Mycobacterium”  
Klik data **Genomic assembly contig set**

www.ebi.ac.uk/ena/data/view/JLSX01000000

## WGS Sequence Set: JLSX01000000

Mycobacterium tuberculosis TKK\_04\_0029, WGS project JLSX01000000 data

View: [TEXT](#) [XML](#) Download: [XML](#) [TEXT](#)

Lineage	Molecule type	Topology	Data class	Taxonomic Division
<a href="#">Mycobacterium tuberculosis</a> <a href="#">TKK_04_0029</a>	genomic DNA	linear	SET	PRO

Number of sequences: 25 Set Version: 1

Keywords: WGS.

Lineage: [Bacteria](#), [Actinobacteria](#), [Corynebacteriales](#), [Mycobacteriaceae](#), [Mycobacterium](#), [Mycobacterium tuberculosis complex](#)

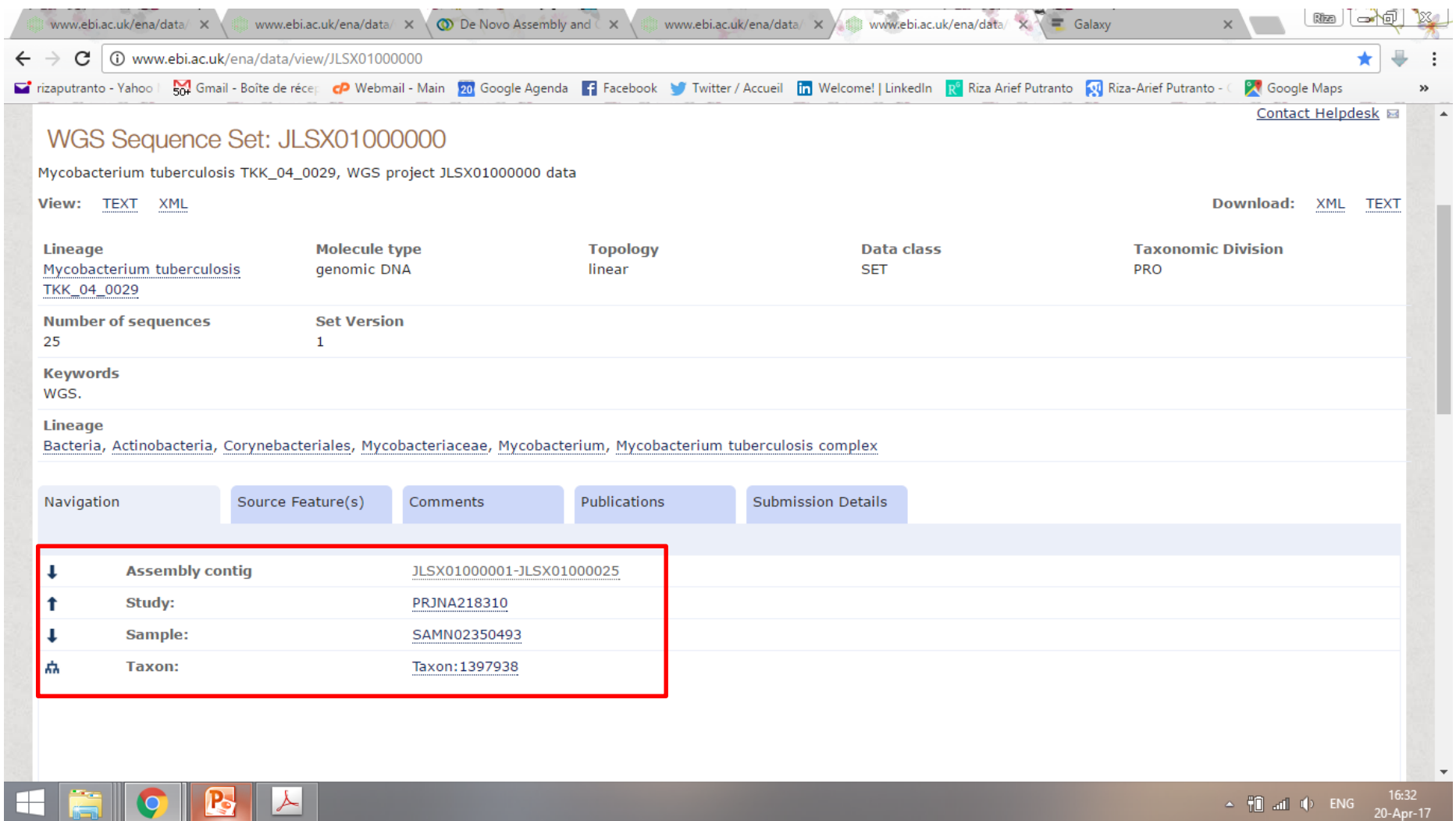
Navigation | Source Feature(s) | Comments | Publications | Submission Details

[1] **The Genome Sequence of Mycobacterium tuberculosis TKK\_04\_0029**  
Earl A.M., Cohen K., Pym A., Bishai W., Maharaj K., Desjardins C., Abeel T., Young S., Zeng Q., Gargeya S., Abouelleil A., Alvarado L., Chapman S.B., Gainer-Dewar J., Goldberg J., Griggs A., Gujja S., Hansen M., Howarth C., Imamovic A., Larimer J., Murphy C., Naylor J., Pearson M., Poon T.W., Priest M., Roberts A., Saif S., Shea T., Sykes S., Wortman J., Nusbaum C., Birren B.

### Step 4. Kata kunci “Human Mycobacterium”

Publikasi dari data tersebut dapat dicek pada laman **Publications**  
Data tersebut belum pernah dipublikasi dalam jurnal resmi

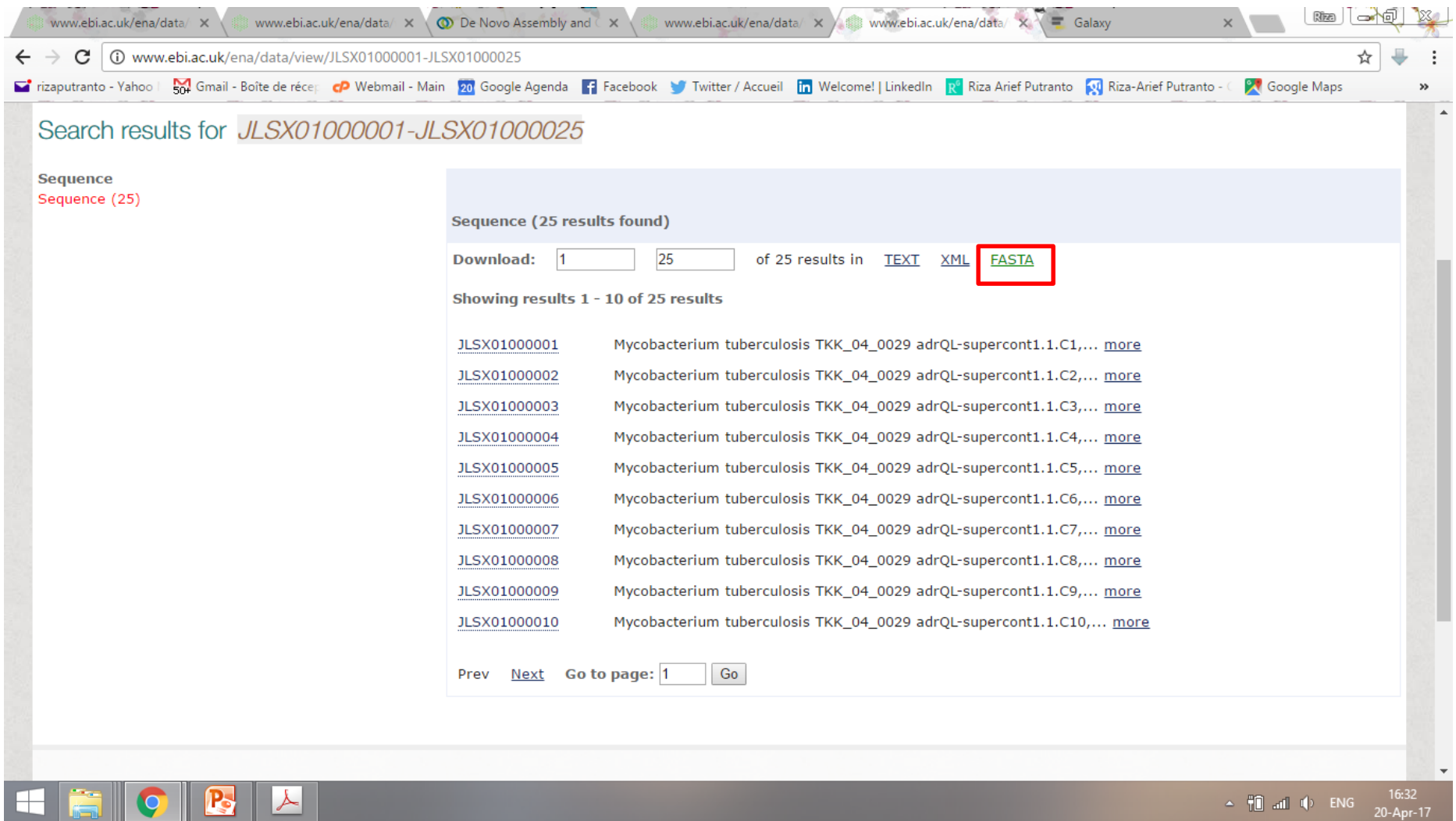




## Step 5. Kata kunci “Human Mycobacterium”

Data genomik dapat diakses di laman **Navigation**

Klik kode sekuen **JLSX01000001** untuk mengunduh data



## Step 6. Kata kunci “Human Mycobacterium”

Data sekuen transkrip mRNA dapat diunduh dalam bentuk **.fasta** dengan langsung klik **FASTA**

## 2. Studi komparatif sekuen pembandingan dan target menggunakan Galaxy

galaxy.southgreen.fr/galaxy/

Galaxy

Analyze Data Workflow Shared Data Visualization Help User

Using 94%

Tools

search tools

Get Data

Send Data

BASIC TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

SEQUENCE ANALYSIS

Gene/Protein prediction

EMBOSS

Operate on Genomic Intervals

Fetch Sequences

Genomics

Fetch Alignments

Extract Features

NGS ANALYSIS

NGS: Quality Control

NGS : Mapping

NGS: GATK Tools

NGS: GATK2 Tools

NGS: SAM/BAM Manipulations

History

search datasets

Pelatihan Bioinformatika 2017

5 shown

639.57 KB

5: tblastn 11 documents from (b) Arabidopsis COBRA aa.fasta vs GBRJ01000001-GBRJ01000476.fasta

4: tblastn 11 documents from (b) Arabidopsis COBRA aa.fasta vs GBRJ01000001-GBRJ01000476.fasta

3: 11 documents from (b) Arabidopsis COBRA aa.fasta

2: 11 documents from (a) Arabidopsis COBRA cDNA.fasta

1: GBRJ01000001-GBRJ01000476.fasta

South Green bioinformatics platform

Welcome to GALAXY

Our pre-configured and validated workflows

Chromosome reconstruction

Scaffrehunter tools assemble scaffold pseudomolecules using markers genotyped in a population (Martin et al, 2016)

Input: Fastq + FASTA

Access workflow

## Step 1.

Masuk ke *interface Galaxy* tools melalui <http://galaxy.southgreen.fr/galaxy/>

galaxy.southgreen.fr/galaxy/

Tools

search tools

Get Data

Send Data

BASIC TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

SEQUENCE ANALYSIS

Gene/Protein prediction

EMBOSS

Operate on Genomic Intervals

Fetch Sequences

Genomics

Fetch Alignments

Extract Features

NGS ANALYSIS

NGS: Quality Control

NGS : Mapping

NGS: GATK Tools

NGS: GATK2 Tools

NGS: SAM/BAM Manipulations

NGS: RNASeq

South Green bioinformatics platform

Welcome to GALAXY

Our pre-configured and validated workflows

NGS analyses

SNP calling

SNP analysis

GWAS

GWAS

SNIPlay3 GWAS workflow: Tassel-based GWAS workflow (GLM model) including population structure and correction for structure (Dereeper et al, 2015)

Input: VCF + Phenotypic tabulated file

Access workflow

History

search datasets

Unnamed history

0 b

This history is empty. You can load your own data or get data from an external source

galaxy.southgreen.fr/galaxy/user/create

16:13 21-Apr-17

## Step 2.

Register untuk membuat akun baru di laman **Galaxy**

The screenshot shows the Galaxy web interface. The main content area is titled "Create account" and contains the following fields:

- Email address:**
- Password:**
- Confirm password:**
- Public name:**

Below the fields, there is a note: "Your public name is an identifier that will be used to generate addresses for information you share publicly. Public names must be at least three characters in length and contain only lower-case letters, numbers, and the '-' character." A "Submit" button is located at the bottom of the form.

The left sidebar contains a "Tools" section with a search bar and a list of tool categories: "Get Data", "Send Data", "BASIC TOOLS", "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "Convert Formats", "SEQUENCE ANALYSIS", "Gene/Protein prediction", "EMBOSS", "Operate on Genomic Intervals", "Fetch Sequences", "Genomics", "Fetch Alignments", "Extract Features", "NGS ANALYSIS", "NGS: Quality Control", "NGS : Mapping", "NGS: GATK Tools", "NGS: GATK2 Tools", "NGS: SAM/BAM Manipulations", and "NGS: RNASer".

The right sidebar contains a "History" section with a search bar and a message: "This history is empty. You can load your own data or get data from an external source".

The top navigation bar includes "Analyze Data", "Workflow", "Shared Data", "Visualization", "Help", and "User". The top status bar shows "Using 0 bytes".

### Step 3.

Isikan kolom yang diminta dan **submit**

Sekali register, akun dapat digunakan sampai kapanpun, termasuk data registry

galaxy.southgreen.fr/galaxy/

Galaxy

Analyze Data Workflow Shared Data Visualization Help User Using 0%

Tools

search tools

Get Data Send Data

BASIC TOOLS

Text Manipulation Filter and Sort Join, Subtract and Group Convert Formats

SEQUENCE ANALYSIS

Gene/Protein prediction EMBOSS Operate on Genomic Intervals Fetch Sequences Genomics Fetch Alignments Extract Features

NGS ANALYSIS

NGS: Quality Control NGS: Mapping NGS: GATK Tools NGS: GATK2 Tools NGS: SAM/BAM Manipulations

cirad IRD Institut de recherche pour le développement Bioersity INRA SCIENCE & INNOVATION SupAgro

# South Green

bioinformatics platform

Welcome to GALAXY

## Our pre-configured and validated workflows

NGS analyses SNP calling SNP analysis GWAS Structural variations

### Structural variations

Scaffremodeler can be used to detect large structural variations between a reference sequence and a resequenced genome (Martin et al, 2016)

Input: Fastq + FASTA

Access workflow

History

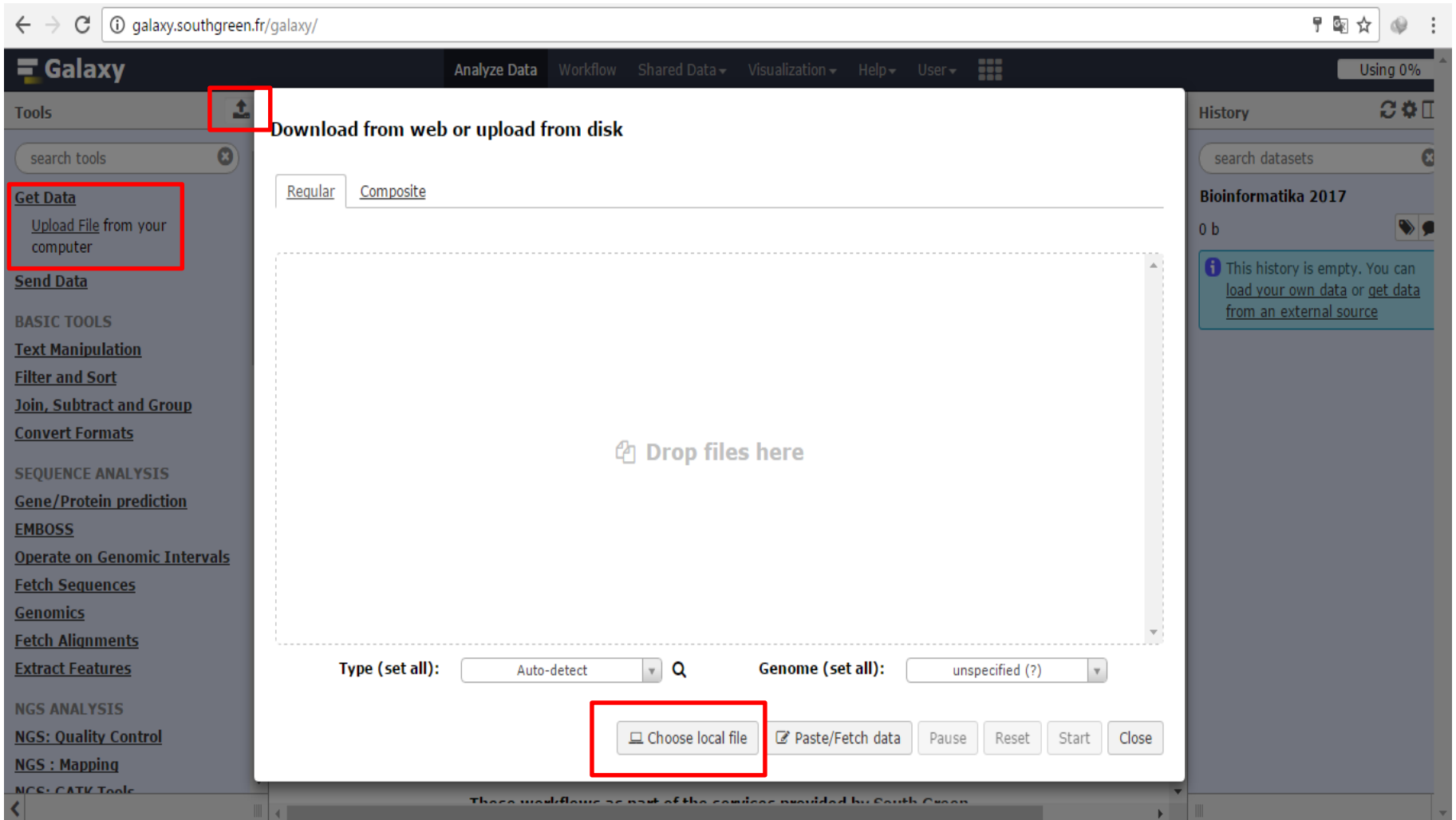
search datasets

Bioinformatika 2017

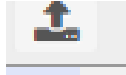
This history is empty. You can load your own data or get data from an external source

## Step 4.

*Rename history* dengan nama yang diinginkan



## Step 5.

Klik **Get Data (upload file)** atau tanda  di bagian kiri atas  
Lalu klik **Choose local file**



galaxy.southgreen.fr/galaxy/

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 0%

Tools search tools

Get Data Upload File from your computer

Send Data

BASIC TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

SEQUENCE ANALYSIS

Gene/Protein prediction

EMBOSS

Operate on Genomic Intervals

Fetch Sequences

Genomics

Fetch Alignments

Extract Features

NGS ANALYSIS

NGS: Quality Control

NGS : Mapping

NGS: CATK Tools

History search datasets

Bioinformatika 2017



















0 b



This history is empty. You can [load your own data](#) or [get data from an external source](#)



### Download from web or upload from disk

Regular Composite

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

Name	Size	Type	Genome	Settings	Status
 Oryza-officinalis-transcripts-GBRJ01000001-GBRJ01000476.fasta	598.5 KB	Auto-dete...  	unspecified (?) 		0% 
 Mycobacterium-tuberculosis-TKK_04_0029-JLSX01000001-JLSX01000025.fasta	4.3 MB	Auto-dete...  	unspecified (?) 		0% 
 11 Arabidopsis COBRA aa.fasta	7 KB	Auto-dete...  	unspecified (?) 		0% 

Type (set all):   Genome (set all):  

 Choose local file  Paste/Fetch data

## Step 6.

Upload data-data sekuen terpilih klik **Start**

galaxy.southgreen.fr/galaxy/

Galaxy

Analyze Data Workflow Shared Data Visualization Help User

Tools

search tools

Get Data

Upload File from your computer

Send Data

BASIC TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

SEQUENCE ANALYSIS

Gene/Protein prediction

EMBOSS

Operate on Genomic Intervals

Fetch Sequences

Genomics

Fetch Alignments

Extract Features

NGS ANALYSIS

NGS: Quality Control

NGS: Mapping

NGS: CATK Tools

Using 3%

History

search datasets

Bioinformatika 2017

3 shown

4.88 MB




3: 11 Arabidopsis COBRA aa.fasta

2: Mycobacterium-tuberculosis-TKK\_04\_002 9-JLSX01000001-JLSX01000025.fasta



1: Oryza-officinalis-transcripts-GBRJ01000001-GBRJ01000476.fasta

### Download from web or upload from disk

Regular Composite

Name	Size	Type	Genome	Settings	Status
 Oryza-officinalis-transcripts-GBRJ01000001-GBRJ01000476.fasta	598.5 KB	Auto-dete... Q	unspecified (?)	⚙️	100% ✓
 Mycobacterium-tuberculosis-TKK_04_0029-JLSX01000001-JLSX01000025.fasta	4.3 MB	Auto-dete... Q	unspecified (?)	⚙️	100% ✓
 11 Arabidopsis COBRA aa.fasta	7 KB	Auto-dete... Q	unspecified (?)	⚙️	100% ✓

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

 Choose local file  Paste/Fetch data Pause Reset Start **Close**

## Step 7.

Setelah sekuen 100 % terupload klik **Close**

The screenshot shows the Galaxy web interface. The main panel displays a list of genomic data entries, each with a search bar and a list of tools. The History panel on the right shows a list of datasets. A red box highlights the eye icon for the dataset '3: 11 Arabidopsis COBRA aa.fasta', indicating that it is the target of the next step.

Tools: search tools

Get Data  
Upload File from your computer

Send Data

BASIC TOOLS  
Text Manipulation  
Filter and Sort  
Join, Subtract and Group  
Convert Formats

SEQUENCE ANALYSIS  
Gene/Protein prediction  
EMBOSS  
Operate on Genomic Intervals  
Fetch Sequences  
Genomics  
Fetch Alignments  
Extract Features

NGS ANALYSIS  
NGS: Quality Control  
NGS: Mapping  
NGS: CATK Tools

History  
search datasets

Bioinformatika 2017  
3 shown  
4.88 MB

3: 11 Arabidopsis COBRA aa.fasta  
2: Mycobacterium-tuberculosis-TKK 04 002  
9-JLSX01000001-JLSX0100002  
5.fasta  
1: Oryza-officinalis-transcripts-GBR01000001-GBR01000476.fasta

Tampilkan semua

## Step 8.

Klik tanda “Mata” di sebelah nama file untuk pengecekan sekuen sudah terupload

galaxy.southgreen.fr/galaxy/

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 3%

Tools

**NCBI BLAST+**

- NCBI BLAST+ makeprofiledb Make profile database
- NCBI BLAST+ makeblastdb Make BLAST database
- NCBI BLAST+ database info Show BLAST database information from blastdbcmd
- NCBI BLAST+ blastx Search protein database with translated nucleotide query sequence(s)
- NCBI BLAST+ dustmasker masks low complexity regions
- NCBI BLAST+ blastn Search nucleotide database with nucleotide query sequence(s)
- BLAST XML to tabular Convert BLAST XML output to tabular
- NCBI BLAST+ tblastn Search translated nucleotide database with protein query sequence(s)
- NCBI BLAST+ tblastx Search translated nucleotide database with translated nucleotide

**NCBI BLAST+ tblastn** Search translated nucleotide database with protein query sequence(s) (Galaxy Version 0.1.04) Options

**Protein query sequence(s)**  
3: 11 Arabidopsis COBRA aa.fasta

**Subject database/sequences**  
FASTA file from your history (see warning note below)

**Nucleotide FASTA file to use as database**  
1: Oryza-officinalis-transcripts-GBRJ01000001-GBRJ01000476.fasta

**Type of BLAST**

tblastn - Traditional TBLASTN to compare protein query to translated nucleotide database  
 tblastn-fast - Use longer words for seeding, faster but less accurate

**Set expectation value cutoff**  
0.01

**Output format**  
Tabular (extended 25 columns)

**Advanced Options**  
Hide Advanced Options

Execute

**Note.** Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.

History

search datasets

**Bioinformatika 2017**  
3 shown  
4.88 MB

- 3: 11 Arabidopsis COBRA aa.fasta
- 2: Mycobacterium-tuberculosis-TKK\_04\_002\_9-JLSX01000001-JLSX01000025.fasta
- 1: Oryza-officinalis-transcripts-GBRJ01000001-GBRJ01000476.fasta

## Step 9.

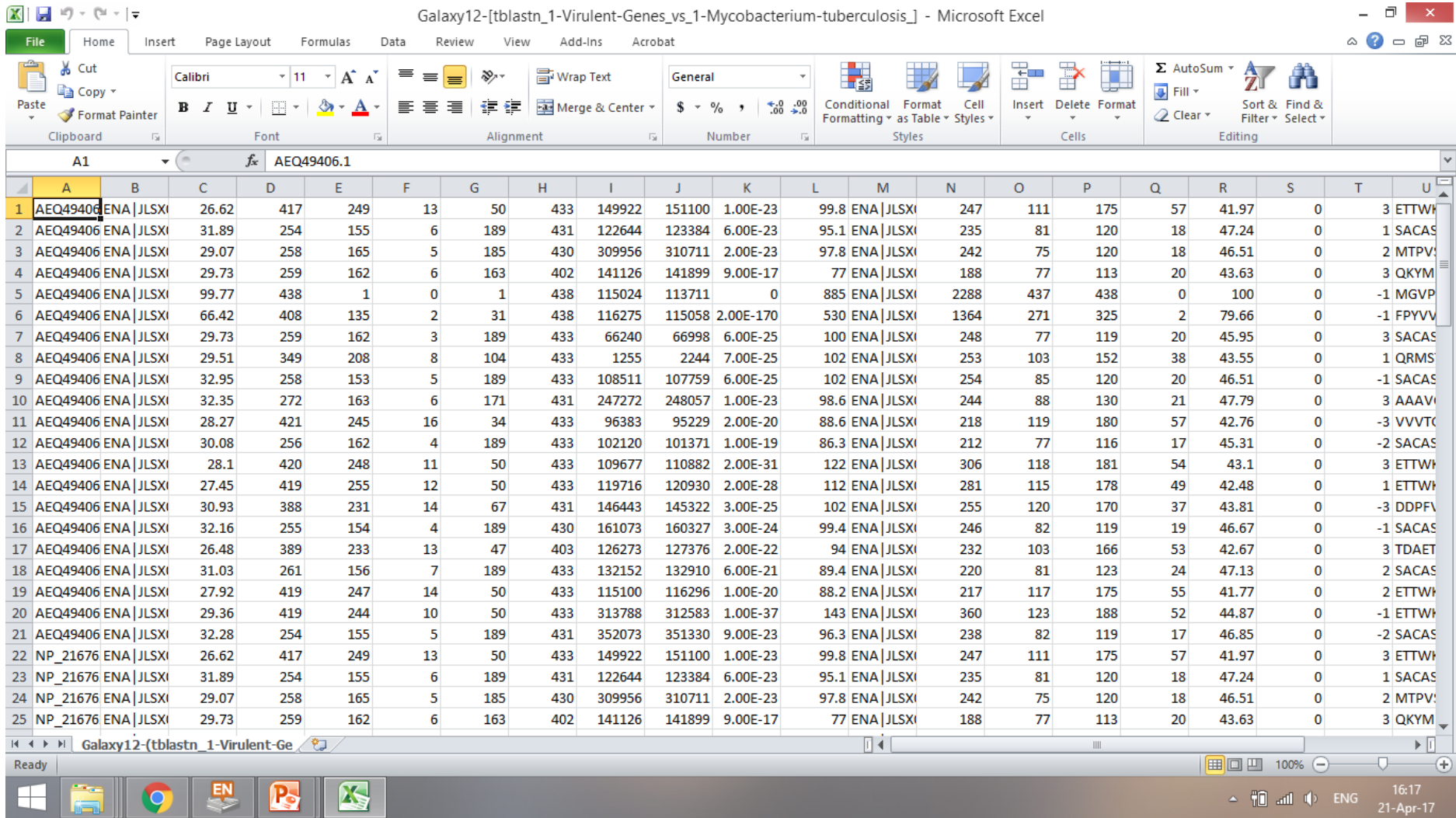
Cari NCBI BLAST+ lalu klik NCBI BLAST+tblastn

The screenshot shows the Galaxy web interface. A green notification box in the center states: "1 job has been successfully added to the queue - resulting in the following datasets: 4: tblastn 11 Arabidopsis COBRA aa.fasta vs Oryza-officinalis-transcripts-GBRJ01000001-GBRJ01000476.fasta". Below this, it says: "You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered." The History panel on the right shows a list of datasets. The top dataset is highlighted in green and has a red box around its save icon (a floppy disk). The dataset name is "4: tblastn 11 Arabidopsis COBRA aa.fasta vs Oryza-officinalis-transcript s-GBRJ01000001-GBRJ01000476.fasta". Below the name, it says "19 lines" and "format: tabular, database: ?". The table below shows columns 1 and 2 with rows of sequence identifiers and accession numbers.

1	2
AT3G02210.1	ENA GBRJ01000002 GBRJ01000
AT3G02210.1	ENA GBRJ01000168 GBRJ01000
AT3G16860.1	ENA GBRJ01000125 GBRJ01000
AT3G20580.1	ENA GBRJ01000404 GBRJ01000
AT3G29810.1	ENA GBRJ01000009 GBRJ01000

## Step 10.

Klik tanda **Save**  untuk menyimpan file hasil tblastn dalam tabular Excel



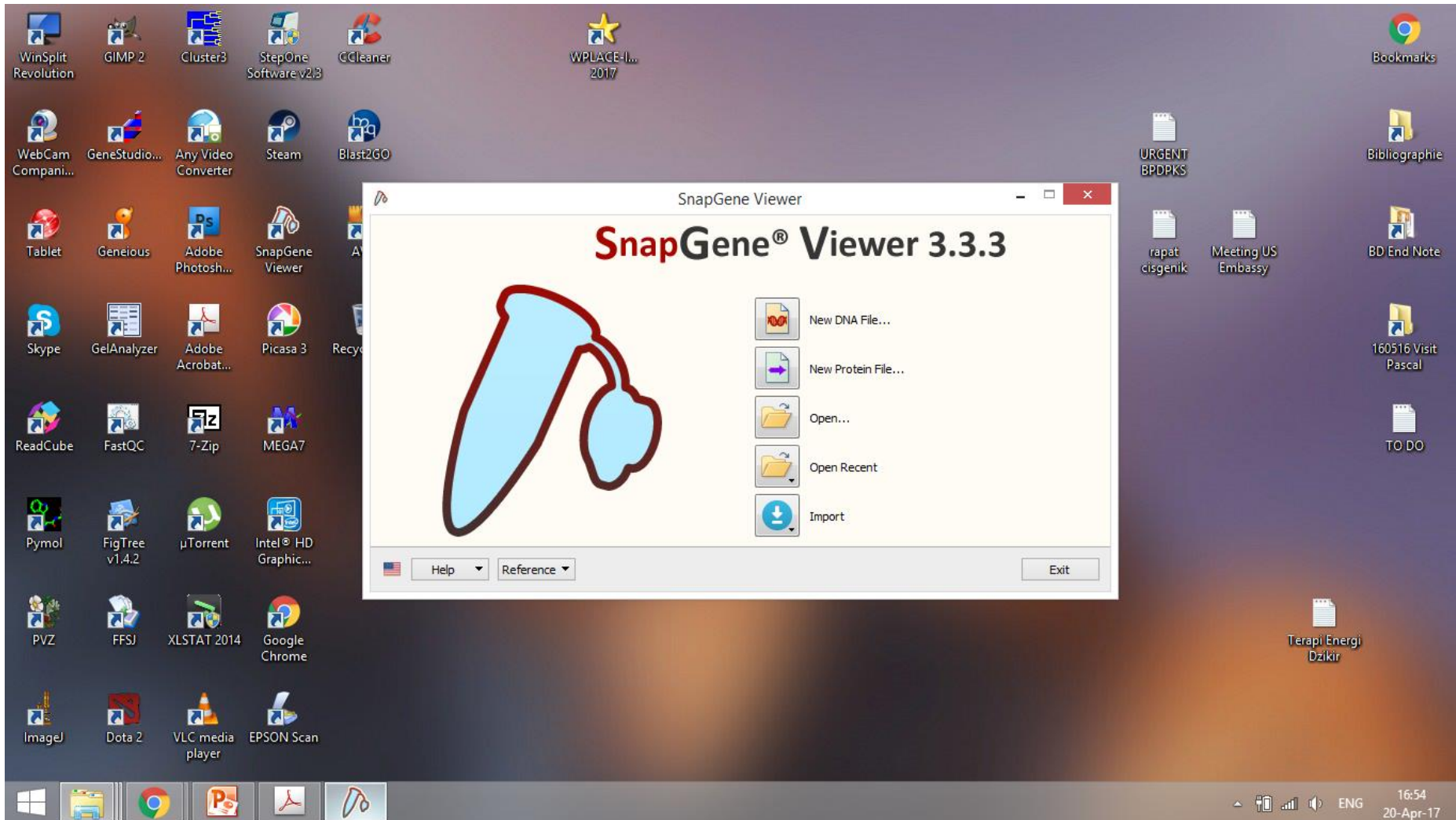
### Step 11.

Tampilan tabel bentuk **tabular**

Pindahkan atau simpan sebagai file **Excel Workbook**

**Penting:** file dalam bentuk tabular tidak *editable*

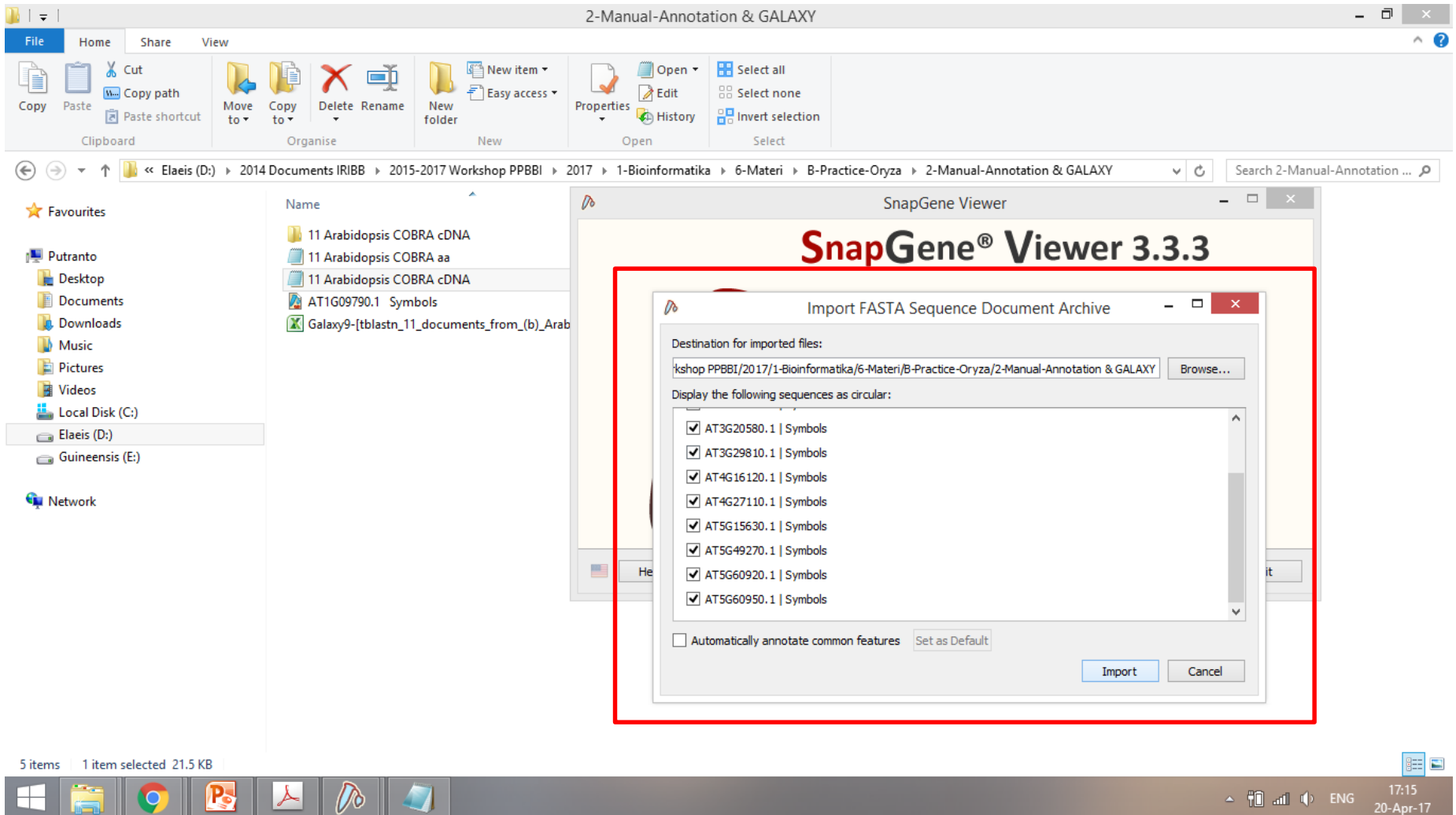
### **3. Anotasi sekuen nukleotida dengan menggunakan SnapGene**



## Step 1.

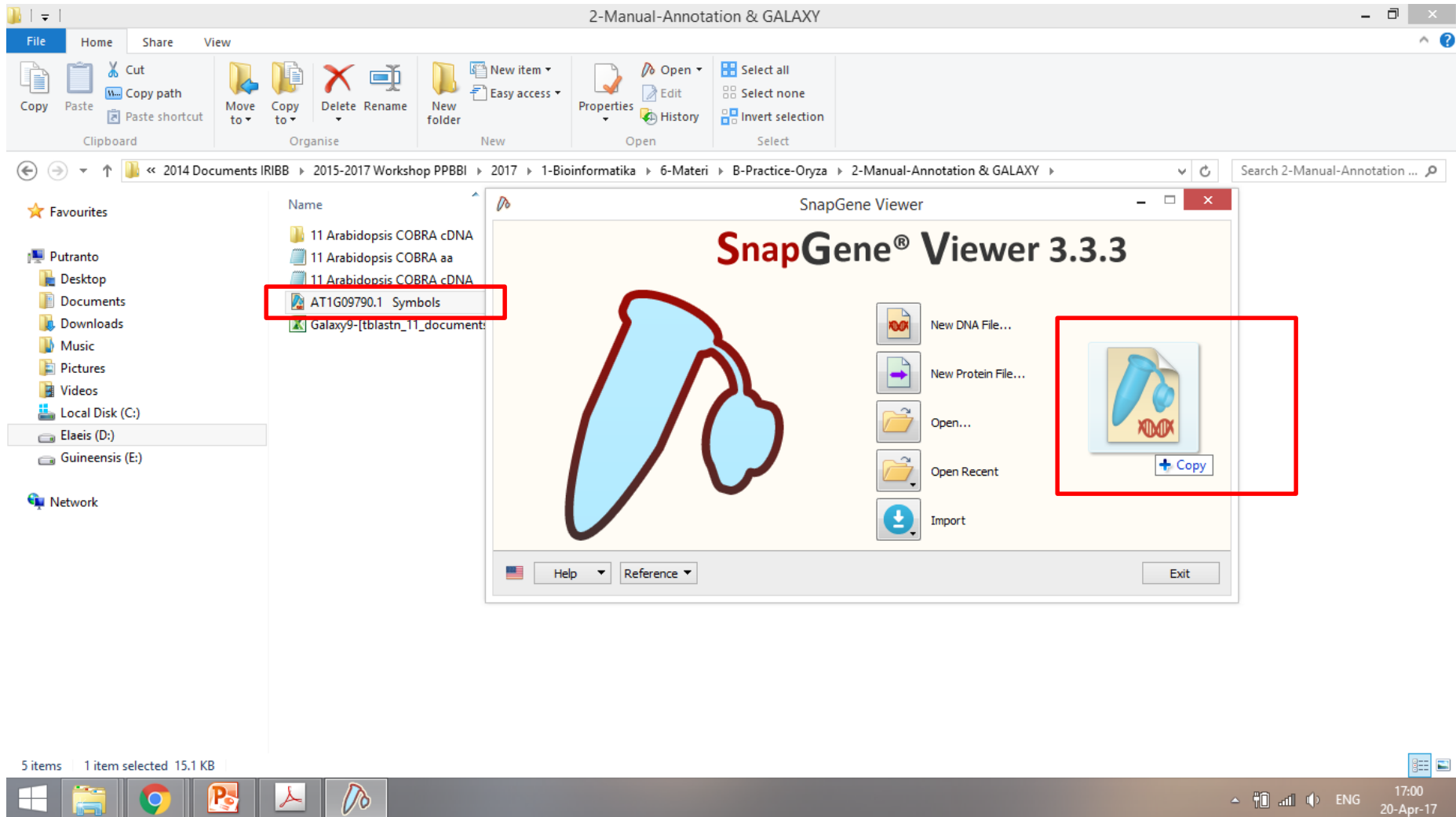
Buka program **SnapGene** dengan halaman interface seperti diatas





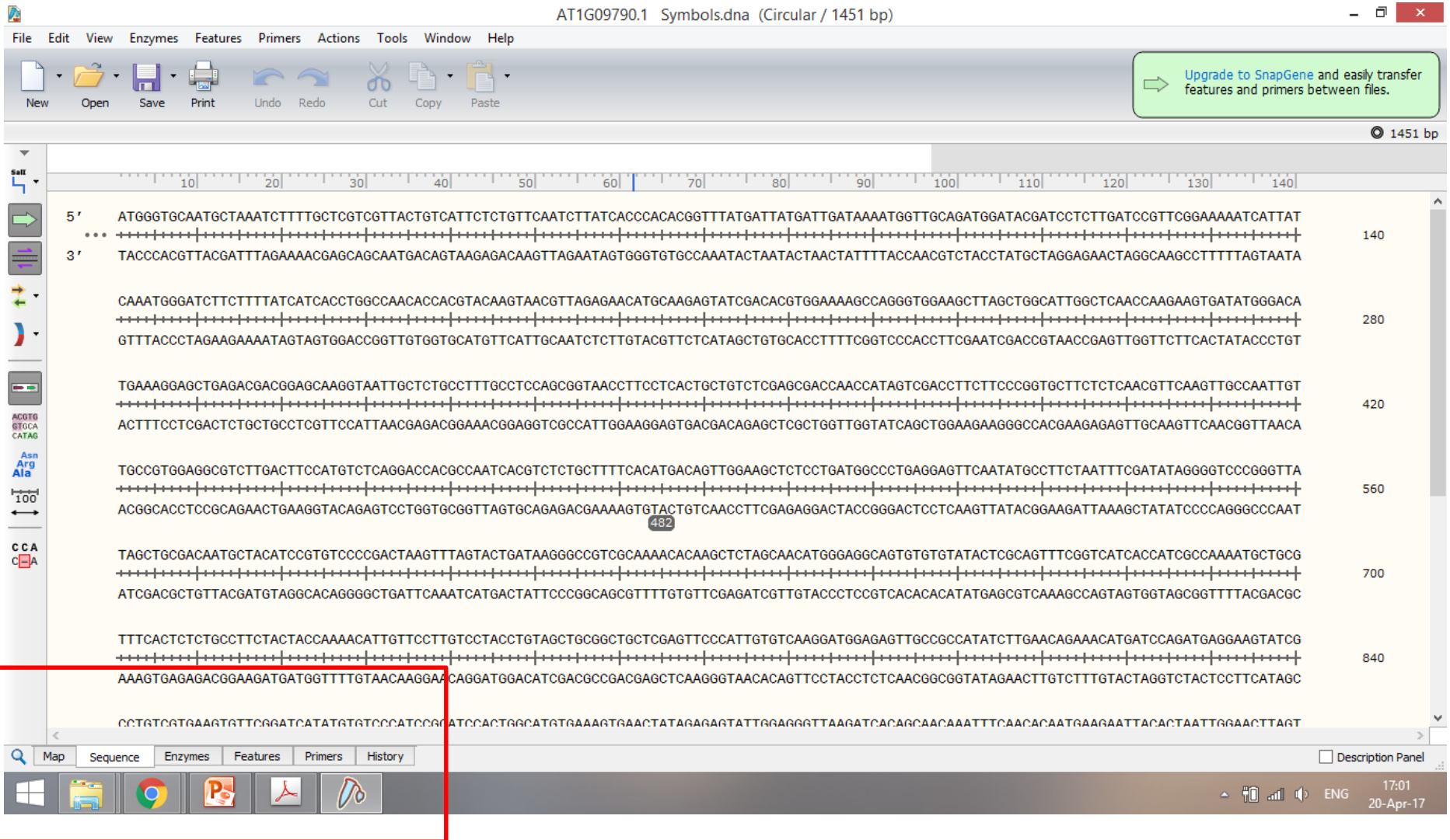
## Step 2.

Untuk merubah file **.fasta** menjadi file SnapGene, unggah file **.fasta** dan klik **Import**



### Step 3.

Buka file SnapGene dengan **drag** atau **open file**



## Step 4.

Fitur **Sequence** dapat diakses di bagian bawah dari program SnapGene  
Sekuen query ditampilkan dalam urutan 5-3' dan 3-5'

AT1G09790.1 Symbols.dna (Circular / 1451 bp)

File Edit View Enzymes Features Primers Actions Tools Window Help

New Open Save Print Undo Redo Cut Copy Paste

Upgrade to SnapGene and easily transfer features and primers between files.

1451 bp

5' ATGGGTGCAATGCTAAATCTTTTGCTCGTCGTTACTGTCATTCTCTGTTCAATCTTATCACCCACACGGTTTATGATTATGATTGATAAAATGGTTGCAGATGGATACGATCCTCTTGATCCGTTCCGAAAAATCATTAT  
3' TACCCACGTTACGATTTAGAAAAACGAGCAGCAATGACAGTAAGAGACAAGTTAGAATAGTGGGTGTGCCAAATACTAATACTAACTATTTTACCAACGCTACCTATGCTAGGAGAAGTGGCAAGCCTTTTATAGTAATA  
M G A M L N L L L V V T V I L C S I L S P T R F M I M I D K M V A D G Y D P L D P F G K I I I  
M P A I S F R K S T T V T M R Q E I K D G V R N I I I I S L I T A S P Y S G R S G N P F I M I  
CAAATGGGATCTTCTTTTATCATCACCTGGCCAACACCACGTACAAGTAACGTTAGAGAACATGCAAGAGTATCGACACGTTGAAAAAGCCAGGGTGAAGCTTAGCTGGCATTGGCTCAACCAAGAAAGTGATATGGGACA  
GTTTACCCTAGAAGAAAATAGTAGTGGACCGGTTGTGGTGCATGTTCAATGCAATCTCTGTACGTTCTCATAGCTGTGCACCTTTTCGGTCCCACCTTCGAATCGACCGTAACCGAGTTGGTTCTTCACTATACCTGT  
K W D L L L S S P G Q H H V Q V T L E N M Q E Y R H V E K P G W K L S W H W L N Q E V I W D  
L H S R R R K D D G P W C W T C T V N S F M C S Y R C T S F G P H F S L Q C Q S L W S T I H S M  
TGAAAGGAGCTGAGACGACGGAGCAAGGTAATTGCTCTGCCTTTGCCTCCAGCGGTAACCTTCTCACTGCTGTCTCGAGCGACCAACCATAGTCGACCTTCTTCCCGGTGCTTCTCTCAACGTTCAAGTTGCCAATTGT  
ACTTTCCTCGACTCTGCTGCCTCGTCCATTAACGAGACGGAACGGAGGTCGCCATTGGAAGGAGTGACGACAGAGCTCGCTGGTGGTATCAGCTGGAAGAAGGCCACGAAGAGAGTTGCAAGTTCAACGGTTAACA  
M K G A E T T E Q G N C S A F A S S G N L P H C C L E R P T I V D L L P G A S L N V Q V A N C  
TGCCGTGGAGGCGTCTTGACTTCCATGTCTCAGGACCACGCCAATCACGTCCTGCTTTTACATGACAGTTGGAAGCTCTCTGATG6CCCTGAGGAGTTCAATATGCCTTCTAATTTGATATAGGGGTCCCGGTTA  
ACGGCACCTCCGAGAACTGAAGGTACAGAGTCTGGTGC6GTTAGTGCAGAGACGAAAAGTGTACTGTCAACCTTCGAGAGGACTACGGGACTCCTCAAGTTATACGGAAGATTAAGCTATATCCCAGGGCCCAAT  
C R G G V L T S M S Q D H A N H V S A F H M T V G S S P D G P E E F N M P S N F D I G V P G Y  
TAGCTGCGACAATGCTACATCCGTGTC6CGACTAAGTTTAGTACTGATAAGGGCCGTCGAAAACACAAGCTCTAGCAACATGGGAGGCAAGTGTGTGTATACTCGCAAGTTTCGGTCATCACCATCGCCAAATGCTGCG

ACGTG GTGCA CATAG

Asn Arg Ala

100

CCA C A

Map Sequence Enzymes Features Primers History

Description Panel

17:04 20-Apr-17

## Step 5.

Klik **Translation** untuk mengakses prediksi urutan residu asam amino dari sekuen nukleotida

Sekuen asam amino forward dalam warna kuning

Sekuen asam amino reverse dalam warna hijau

The screenshot shows the SnapGene software interface. The title bar reads "AT1G09790.1 Symbols.dna (Circular / 1451 bp)". The menu bar includes "File", "Edit", "View", "Enzymes", "Features", "Primers", "Actions", "Tools", "Window", and "Help". The "Features" menu is open, and the "Add Translated Feature..." option is highlighted with a red box. Other options in the menu include "Edit Features...", "Duplicate Feature...", "Remove Features", "Create Feature Segment...", "Delete Feature Segment...", "Merge Feature Segments...", "Add Cleavage Site...", "Remove Cleavage Sites...", "Feature Color...", "Choose Alternative Codons...", "Splice to Remove Introns...", "Show Features Inside the Selection", "Hide Features Inside the Selection", "Sort Feature List...", "Import Features from Another File...", "Export Feature Data...", "Detect Common Features...", "Add to Common Features...", "Browse Common Features", and "Features Tutorial Video". The main window displays a DNA sequence with a protein translation below it. The protein sequence is: I K D G V R N I I I I S L I T A S P Y S G R S G N P F I M I. The sequence is shown in blue and green blocks, with arrows indicating the reading frame. The bottom status bar shows "Map", "Sequence", "Enzymes", "Features", "Primers", "History", and "Description Panel". The system tray at the bottom right shows the date and time: "ENG 17:04 20-Apr-17".

## Step 6.

Lakukan anotasi dengan klik dua kali pada prediksi anotasi CDS yang diinginkan

Klik **Feature**, dan **Add Translated Features**

AT1G09790.1 Symbols.dna (Circular / 1451 bp)

File Edit View Enzymes Features Primers Actions Tools Window Help

New Open Save Print Undo

Selected: ORF (1 .. 1365 = 1365 bp) [45%

Add Feature for AT1G09790.1 Symbols

Label: ORF

Type: CDS

Translate this feature in Sequence view Feature Translation Options...

1365 bp / 1 segment

Segment	Location	Size (bp)	Color
1	ORF	1 .. 1365	

/product =

COBRA-like protein 6 precursor

Prioritize display of this feature in maps

OK Cancel

1451 bp

130 140

TTCGGAAAAATCATTAT 140

AAGCCTTTTATAGTAATA

F G K I I I →

N P F I M I →

AGAAGTGATATGGGACA 280

TCTTCACTATACCCTGT

E V I W D →

S T I H S M →

TTCAAGTTGCCAATTGT 420

AAGTTCAACGGTTAACA

V Q V A N C →

ATAGGGGTCCCGGGTTA 560

TATCCCAGGGCCCAAT

I G V P G Y →

TAGCTGCGACAATGCTACATCCGTGCCCGACTAAGTTTACTACTGATAAGGGCCGTGCGAAAAACACAAGCTCTAGCAACATGGGAGGCAAGTGTGTATACTCGCAGTTTCGGTCATCACCATCGCCAAAATGCTGGC

Map Sequence Enzymes Features Primers History

Description Panel

17:07 20-Apr-17

## Step 7.

Pilih tipe feature: **CDS** seperti ditampilkan diatas

Berikan nama pada feature CDS: **COBRA-like protein 6 precursor**

AT1G09790.1 Symbols.dna\* (Circular / 1451 bp)

File Edit View Enzymes Features Primers Actions Tools Window Help

New Open Save Print Undo Redo Cut Copy Paste

1451 bp

5' ATGGGTGCAATGCTAAATCTTTTCTGCTGCTGTTACTGTCATTCTCTGTTCAATCTTATCACCCACACGGTTTATGATTATGATTGATAAAATGGTTGCAGATGGATACGATCCTCTTGATCCGTTCCGGAAAAATCATTAT  
 ...  
 3' TACCCACGTTACGATTTAGAAAACGAGCAGCAATGACAGTAAGAGACAAGTTAGAATAGTGGGTGTGCCAAATACTAATACTAATACTTTTACCAACGCTACCTATGCTAGGAGAAGCTAGGCAAGCCCTTTTAGTAATA

1 5 10 15 20 25 30 35 40 45  
 M G A M L N L L L V V T V I L C S I L S P T R F M I M I D K M V A D G Y D P L D P F G K I I I  
 ORF  
 M G A M L N L L L V V T V I L C S I L S P T R F M I M I D K M V A D G Y D P L D P F G K I I I  
 M P A I S F R K S T T V T M R Q E I K D G V R N I I I I S L I T A S P Y S G R S G N P F I M I

CAAATGGGATCTTCTTTTATCATCACCTGGCCAACACCACGTACAAGTAACGTTAGAGAACATGCAAGAGTATCGACACGTTGAAAAGCCAGGGTGAAGCTTAGCTGGCATTGGCTCAACCAAGAAGTGATATGGGACA  
 ...  
 GTTTACCCCTAGAAGAAAATAGTAGTGGACCAGTTGTGGTGCATGTTCTTGTACGTTCTCATAGCTGTGCACCTTTTCCGGTCCACCTTGAATCGACCGTAACCGAGTTGGTCTTCACTATACCCCTGT

50 55 60 65 70 75 80 85 90  
 K W D L L L S S P G Q H H V Q V T L E N M Q E Y R H V E K P G W K L S W H W L N Q E V I W D  
 ORF  
 K W D L L L S S P G Q H H V Q V T L E N M Q E Y R H V E K P G W K L S W H W L N Q E V I W D  
 L H S R R K D D G P W C W T C T V N S F M C S Y R C T S F G P H F S L Q C Q S L W S T I H S M

TGAAAGGAGCTGAGACGACGGAGCAAGGTAATTGCTCTGCCTTTGCCTCCAGCGGTAACCTTCTCTACTGCTGTCTCAGCGACCAACCATAGTCGACCTTCTCCCGGTGCTTCTCTCAACGTTCAAGTTGCCAATTGT  
 ...  
 ACTTTCTCGACTCTGCTGCCTCGTTCCATTAACGAGACGGAACGGAGTGCACATTGGAAGGAGTGACGACAGAGCTCGCTGGTTGGTATCAGCTGGAAGAAGGGCCACGAAGAGAGTTGCAAGTTCAACGGTTAACA

95 100 105 110 115 120 125 130 135 140  
 M K G A E T T E Q G N C S A F A S S G N L P H C C L E R P T I V D L L P G A S L N V Q V A N C  
 ORF  
 M K G A E T T E Q G N C S A F A S S G N L P H C C L E R P T I V D L L P G A S L N V Q V A N C

Map Sequence Enzymes Features Primers History Description Panel

17:08 20-Apr-17

## Step 8.

Tampilan hasil anotasi **CDS** di laman sekuen

The screenshot shows a bioinformatics software interface with a DNA sequence of 1451 bp. The sequence is displayed in a multi-line format with a scale at the bottom. A red box highlights the 'Tools' menu, which is open and shows several options. The 'BLAST Selected DNA...' option is highlighted in blue. The sequence is shown in 5' to 3' orientation, with a corresponding amino acid sequence below it. The amino acid sequence is shown in a multi-line format with a scale at the bottom. The amino acid sequence is: M G A M L N L L V ... F M I M I D K M V A D G Y D P L D P F G K I I I ... K W D L L L S S P G Q H H V Q V T L E N M Q E Y R H V E K P G W K L S W H W L N Q E V I W D ... M K G A E T T E Q G N C S A F A S S G N L P H C C L E R P T I V D L L P G A S L N V Q V A N C ... C R G G V L T S M S Q D H A N H V S A F H M T V G S S P D G P E E F N M P S N F D I G V P G Y ...

AT1G09790.1 Symbols.dna (Circular / 1451 bp)

File Edit View Enzymes Features Primers Actions **Tools** Window Help

- BLAST Selected DNA... Ctrl+Shift+B
- BLAST Selected Protein... Ctrl+Alt+B
- BLAST Selected Primers...
- Simulate Agarose Gel Ctrl+Shift+G
- Show DNA Calculations Ctrl+Shift+M
- Align Multiple Sequences... Ctrl+L
- Align Copied Sequence... Ctrl+Alt+L
- Align Full Sequences...
- Letter Codes
- Genetic Code Tables
- Codon Usage Tables

Selected: ORF (1 .. 1365 = 1365 bp) [45% GC] 1451 bp

5' ATGGGTGCAATGCTAAATCTTTTGCTCGTC  
3' TACCCACGTTACGATTTAGAAAACGAGCAC

1 5 10  
M G A M L N L L V

25 30 35 40 45  
F M I M I D K M V A D G Y D P L D P F G K I I I

ORF

50 55 60 65 70 75 80 85 90  
K W D L L L S S P G Q H H V Q V T L E N M Q E Y R H V E K P G W K L S W H W L N Q E V I W D

ORF

95 100 105 110 115 120 125 130 135 140  
M K G A E T T E Q G N C S A F A S S G N L P H C C L E R P T I V D L L P G A S L N V Q V A N C

ORF

145 150 155 160 165 170 175 180 185  
C R G G V L T S M S Q D H A N H V S A F H M T V G S S P D G P E E F N M P S N F D I G V P G Y

ACGTG GTGCA CATAG  
Asn Arg Ala  
100  
C C A  
C A

Map Sequence Enzymes Features Primers History Description Panel

17:25 20-Apr-17

## Step 9.

**Optional:** Jika diperlukan dapat langsung lakukan **BLAST Selected DNA** untuk memastikan nama sekuen dalam database NCBI



AT1G09790.1 Symbols.dna\* (Circular / 1451 bp)

File Edit View Enzymes Features Primers Actions Tools Window Help

New Open Save Print Undo

Selected: 3'-UTR (1366 .. 1451 = 86 bp)

Label: 3'-UTR

Type: 3'-UTR

Translate this feature in Sequence view Feature Translation Options...

86 bp / 1 segment

1366 1451

Split Feature...  
Merge Segments  
Delete Segment

Segment	Location	Size (bp)	Color
1	3'-UTR	1366 .. 1451	

/note =

Prioritize display of this feature in maps

OK

Segment Color

No Color

Orange

More Colors... Cancel

Map Sequence Enzymes Features Primers History

Description Panel

17:11  
20-Apr-17

## Step 10.

Lanjutkan anotasi untuk daerah **3'-UTR**

Warna feature dapat dipilih sesuai keinginan

AT1G09790.1 Symbols.dna\* (Circular / 1451 bp)

File Edit View Enzymes Features Primers Actions Tools Window Help

New Open Save Print Undo Redo Cut Copy Paste

1451 bp

V L H P N L K S V Q Q V F S F N Y K S L T P Y Q N S I N D T G M F W G V Q F Y N D V L L Q E →

GAAAGATTGGGAATGTTCAAACAGAGTTATTGCTGAAGAAGGATATGGGAAATTTCACTTTTCAGAGAAGGTTGGGCTTTCCCAAGGAGAATCTTGTTCAATGGTGATGAATGTGTTATGCCTTCTCCAGATGACTTTCCA  
 CTTTCTAACCCTTACAAGTTGTCTCAATAACGACTTCTTCTATACCCCTTAAAGTGAAGTCTCTTCCAACCCGAAAGGGTTCCTCTTAGAACAAGTTACCACTACTTACACAATACGGAAGAGGCTACTGAAAGGT 1260

375 380 385 390 395 400 405 410 415 420  
 G K I G N V Q T E L L L K K D M G N F T F R E G W A F P R R I L F N G D E C V M P S P D D F P  
 ORF

G K I G N V Q T E L L L K K D M G N F T F R E G W A F P R R I L F N G D E C V M P S P D D F P →

AGGCTGCCAAATCTGCTCATTCTCTTCGTTCTTCTGCTGTTATTAGCTCTGTTTCAGTCGTTTTCTGTTTTCTCCTTCATCATCTTCTTCTACTAGTTTGAGATTGTTTCTTTTCATTTGACCTACATTGTTTC  
 TCCGACGGTTTTAGACGAGTAAGGAGAAGCAGAAGGAGACGACAATAATCGAGACAAAGTCAGCAAAGACAAAAGAGGAAGTAGTAGAAGAAGATGATCAAACACTCTAACAAAGAAAGTAAACTGGATGTAACAAG 1400

425 430 435 440 445 450 455  
 R L P K S A H S S S S S S S A V I S S V S V V F C F L L H H L L L L V  
 ORF

R L P K S A H S S S S S S S A V I S S V S V V F C F L L H H L L L L V →

S Q K L R K R N E G E D D E E V L K →

L N N R E N S G V N N →

3'-UTR

CATGCTTCTTTGACTTTTCCAAGAAACAAAGCCTCTAATAAAGCTTGCATC 3'  
 GTACGAAGAACTGAAAAGGTTCTTTGTTTCGGAGATTATTTCGAACGTAG 5' 1451

3'-UTR

W A E K V K G L F L A E L L A Q M

Map Sequence Enzymes Features Primers History

Description Panel

17:12  
20-Apr-17

## Step 11.

Tampilan hasil anotasi **3'-UTR** di laman sekuen dengan warna berbeda

AT1G09790.1 Symbols.dna (Circular / 1451 bp)

File Edit View Enzymes Features Primers Actions Tools Window Help

New Open Save Print Undo Redo Cut Copy Paste

Selected: ORF (1 .. 1365 = 1365 bp) [45% GC] 1451 bp

Upgrade to SnapGene to simulate restriction cloning, Gibson Assembly®, In-Fusion® cloning, Gateway® cloning, and more.

5' ATGGGGTGAATGCTAAATCTTTGCTCGTCGTTACTGTCATTCTCTGTTCAATCTTATACCCACACACGGTTTATGATTATGATTGATAAAATGGTTGCAGATGGATACGA 110

3' TACCCACGTTACGATTTAGAAAACGAGCAGCAATGACAGTAAGAGACAAGTTAGAATAGTGGGTGTGCCAAATACTAATACTAATTTTTACCAACGTCTACCTATGCT

1 5 10 15 20 25 30 35

M G A M L N L L L V V T V I L C S I L S P T R F M I M I D K M V A D G Y D

ORF

TCCTCTTGATCCGTTTCGAAAAATCATTATCAAATGGGATCTTCTTTTATCATCACCTGGCCAACACCACGTACAAGTAACGTTAGAGAACATGCAAGAGTATCGACACG 220

AGGAGAACTAGGCAAGCCTTTTTAGTAATAGTTACCCTAGAAGAAAATAGTAGTGGACCGGTTGTGGTGCATGTTTCATTGCAATCTCTGTACGTTCTCATAGCTGTGC

40 45 50 55 60 65 70

P L D P F G K I I I K W D L L L S S P G Q H H V Q V T L E N M Q E Y R H

ORF

TGAAAAAGCCAGGGTGAAGCTTAGCTGGCATTGGCTCAACCAAGAAGTGATATGGGACATGAAAGGAGCTGAGACGACGGAGCAAGGTAATTGCTCTGCCTTTGCCTCC 330

ACCTTTTCGGTCCCACCTTCGAATCGACCGTAACCGAGTTGGTTCTTCACTATAACCTGTACTTTCTCGACTCTGCTGCCTCGTTCCATTAACGAGACGGAAACGGAGG

75 80 85 90 95 100 105 110

V E K P G W K L S W H W L N Q E V I W D M K G A E T T E Q G N C S A F A S

ORF

AGCGGTAACCTTCTCACTGCTGTCTCGAGCGACCAACCATAGTCGACCTTCTCCGGTGCTTCTCTCAACGTTCAAGTTGCCAATTGTTGCCGTGGAGGCGTCTTGAC 440

TCGCCATTGGAAGGAGTGACGACAGAGCTCGCTGGTTGGTATCAGCTGGAAGAAGGGCCACGAAGAGAGTTGCAAGTTCAACGGTTAAACAACGGCACCTCCGCAGAAGCTG

115 120 125 130 135 140 145

S G N L P H C C L E R P T I V D L L P G A S L N V Q V A N C C R G G V L T

ORF

Map Sequence Enzymes Features Primers History

Description Panel

10:48 21-Apr-17

## Step 12.

Sekuen protein ORF dapat di ekstrak dengan cara **Copy Protein**

The screenshot shows the SnapGene software interface. The main window displays a DNA sequence for AT1G09790.1 (Symbols.dna, Circular / 1451 bp). The sequence is shown in 5' to 3' orientation. A red box highlights a specific region of the DNA sequence, which is also highlighted in the Notepad window. The Notepad window contains the following text:

```
>COBRA-like protein 6 precursor [Arabidopsis thaliana]
MGAMLNLLVWTVILCSILSPTRFMIMIDKMVADGYDLPDPFGKIIKWDLLSSPGQHVVQVTLNMQEYRHVEKPGWKLSSHHLN
QEVIWDMKGAETTEQGNCSAFASSGNLPHCCLERPTIVDLLPGASLNVQVANCCRGVLTSMQDHANHVSAFHMTVGSSPDGPEEF
NMPSNFDIGVPGYSCDNATSVSPTKFDKGRRKQALATWEAVCVYSQFRSSPSPKCCVLSAFYYQNIIVPCPTCSGCSSSHCVK
DGELPPYLEQKHDPDEEVSPVVKCSDHMCPIRIHWHVKVNYREYWRVKITATNFNTMKNYTNWNLVVLHPNLKSVQVFSFNYKSLT
PYQNSINDTGMFVGQFYNDVLLQEGKIGNVQTELLKKDMGNFTREGWAFPRRILFNGDECVMPSPDDFPRLPKSAHSSSSSSAV
ISSVSVVFCFLHLLLLLV*
```

The Notepad window also shows the amino acid sequence corresponding to the highlighted DNA region: M G A M L N L L V W T V I L C S I L S P T R F M I M I D K M V A D G Y D L P D P F G K I I K W D L L S S P G Q H V V Q V T L N M Q E Y R H V E K P G W K L S S H H L N Q E V I W D M K G A E T T E Q G N C S A F A S S G N L P H C C L E R P T I V D L L P G A S L N V Q V A N C C R G V L T S M Q D H A N H V S A F H M T V G S S P D G P E E F N M P S N F D I G V P G Y S C D N A T S V S P T K F D K G R R K Q A L A T W E A V C V Y S Q F R S S P S P K C C V L S A F Y Y Q N I I V P C P T C S G C S S S H C V K D G E L P P Y L E Q K H D P D E E V S P V V K C S D H M C P I R I H W H V K V N Y R E Y W R V K I T A T N F N T M K N Y T N W N L V V L H P N L K S V Q V F S F N Y K S L T P Y Q N S I N D T G M F V G Q F Y N D V L L Q E G K I G N V Q T E L L K K D M G N F T R E G W A F P R R I L F N G D E C V M P S P D D F P R L P K S A H S S S S S S A V I S S V S V V F C F L H L L L L L V \*.

### Step 13.

Paste sekuen asam amino ke **Notepad**  
Lengkapi data dengan sekuen-sekuen lainnya

The screenshot shows a bioinformatics software interface with a sequence viewer and a 'Save As' dialog box. The sequence viewer displays the following information:

- AT1G09790.1 | Symbols: COBL6 | COBRA-like protein 6 precursor | chr1:3168482-3170819  
REVERSE LENGTH=1451  
MGAMLNLLLVTVILCSILSPTRFMIMIDKMWADGYDPLDPFGKIIKWDLLLSSPGQHHVQVTLNMQEYRHEKPGWKLSSHHLN  
QEVIMDMKGAETTEQGCNCSAFASSGNLPHCCLERPTIVDLLPGASLNVQVANCCRGGVLTSMQDHANHVSFAHMTVGSPPDGPEEF  
NMPNSFDIGVPGYSCDNATSVSPKTFSTDKGRRKQALATWEAVCVYSQFRSSPSPKCCVSLSAFYQNIIVPCPTSCGSCSSSHCVK  
DGELPPYLEQKHPDDEEVSPVVKCSDHMCPIRIHWHVKVNYREYWRVKITATNFNTMKNYTNWNLVVLHPNLKSVQVQVFSFNYKSLT  
PYQNSINDTGMFWGVQFYNDVLLQEGKIGNVQTELLLLKDMGNFTFREGWAFPRRILFNGDECVMPSDDFPRLPKSAHSSSSSSAV  
ISSVSVVFCFLHLLHLLLV\*
- AT3G02210.1 | Symbols: COBL1 | COBRA-like protein 1 precursor | chr3:409201-411052  
REVERSE LENGTH=1684  
MGFFLCSSSSIFFKFGISIIFLVSFGLTPSAAYE  
VIWGMGGQTTEQGDCKFKGTIPHCCKKTPSVVE  
NFTLKAPGPGYTCSPAKIVKPTRFIGTDKRRYDQ  
VDPKGPRIASVIPNGKNAYIPPLVQCTKHMVPV  
PLTPYASINDTGLWGIKFYNDLLMQAGPFGVQ  
LFAAMALLLIVFLHGNL\*
- AT3G16860.1 | Symbols: COBL8 | COBRA-like protein 8 precursor | chr3:168601-168801  
REVERSE LENGTH=2361

The 'Save As' dialog box is open, showing the following information:

- File name: 11 Arabidopsis COBRA aa.fasta
- Save as type: Text Documents (\*.txt)
- Encoding: ANSI

The dialog box is highlighted with a red rectangle.

### Step 13.

Simpan dokumen dalam bentuk **.fasta**

Penting: Untuk menyimpan dalam format fasta, tambahkan **.fasta** di belakang nama file

Sekuen ini akan diperlukan untuk analisis komparatif berikutnya di **Galaxy**