# Introduction of UGENE

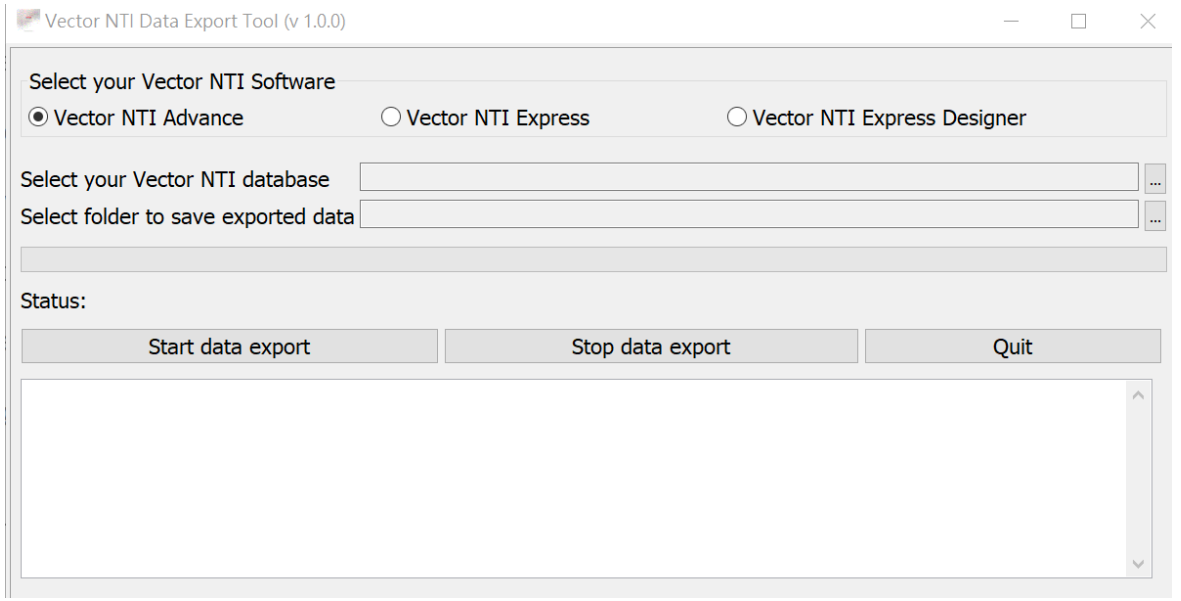
BBSC Bioinformatics Core

陳誼寧

**Yi-Ning Chen**

# Why not Vector NTI

- All support for Vector NTI software will end on **December 31, 2020**. No substitute is planned.
- **Data Export:**
  - Vector NTI Data Export Tool
- Alternative software
  - UGene, Geneous, Benchling, SnapGene

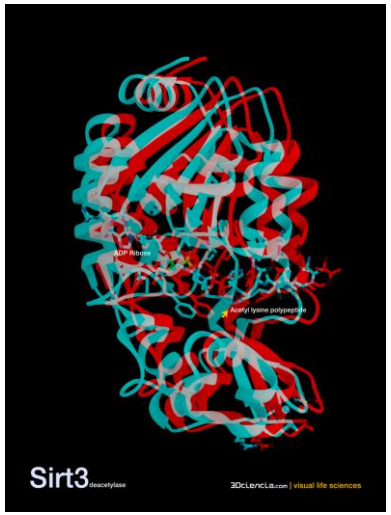


# Key Features of UGENE

- Free cross-platform genome analysis suite
- Creating, editing and annotating nucleic acid and protein sequences
- Plasmid construction and annotation
- Restriction enzyme finder with integrated REBASE restriction enzymes list
- PCR *in silico*
- Multiple sequence alignment:
  - ClustalW, ClustalO, MUSCLE, Kalign, MAFFT, T-Coffee

# Key Features of UGENE

- Open reading frame finder
- Auto Plasmid Annotation :
  - annotate functional elements of the given sequence
    - promoters, terminators, origin of replication, known genes, common primers and other features.
- 3D Structure viewer for PDB and MMDB  files, anaglyph view support



# Key Features of UGENE

- Protein secondary structure prediction with GOR IV and PSIPRED algorithms
- Constructing dotplots for genome sequences
- Local and NCBI BLAST search
- mRNA alignment with Spidey
- Contigs assembly with CAP3

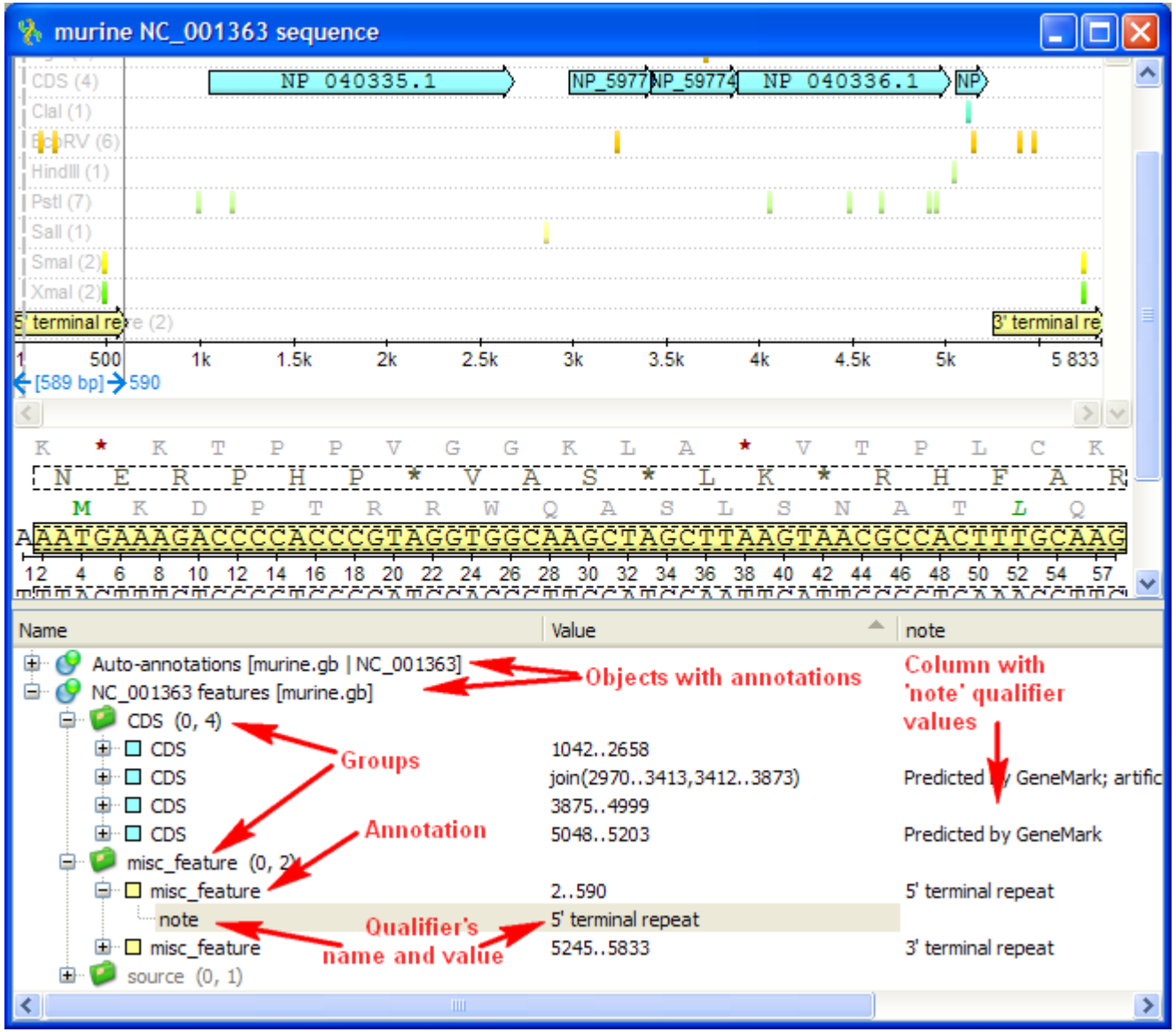
# GenBank Files

- ... LOCUS 3K1V_A 34 bp RNA linear SYN 10-OCT-2012
- ... DEFINITION Chain A, Structure Of A Mutant Class-I Preq1.
- ... ACCESSION 3K1V_A
- ... VERSION 3K1V_A GI:260656459
- ... SOURCE synthetic construct
- ... ORGANISM synthetic construct
- ... other sequences; artificial sequences.
- ... REFERENCE 1 (bases 1 to 34)
- ... AUTHORS Klein,D.J., Edwards,T.E. and Ferre-D'Amare,A.R.
- ... TITLE Cocrystal structure of a class I preQ1 riboswitch
- ... COMMENT SEQRES.
- ... FEATURES Location/Qualifiers
- ... source 1..34
- ... /organism="synthetic construct"
- ... /mol_type="other RNA"
- ... /db_xref="taxon:32630"
- ... misc_binding 1..30
- ... /note="Preq1 riboswitch"
- ... /bound_moiety="preQ1"
- ... ORIGIN
- ... 1 agaggttcta gcacatccct ctataaaaaa ctaa
- ... //

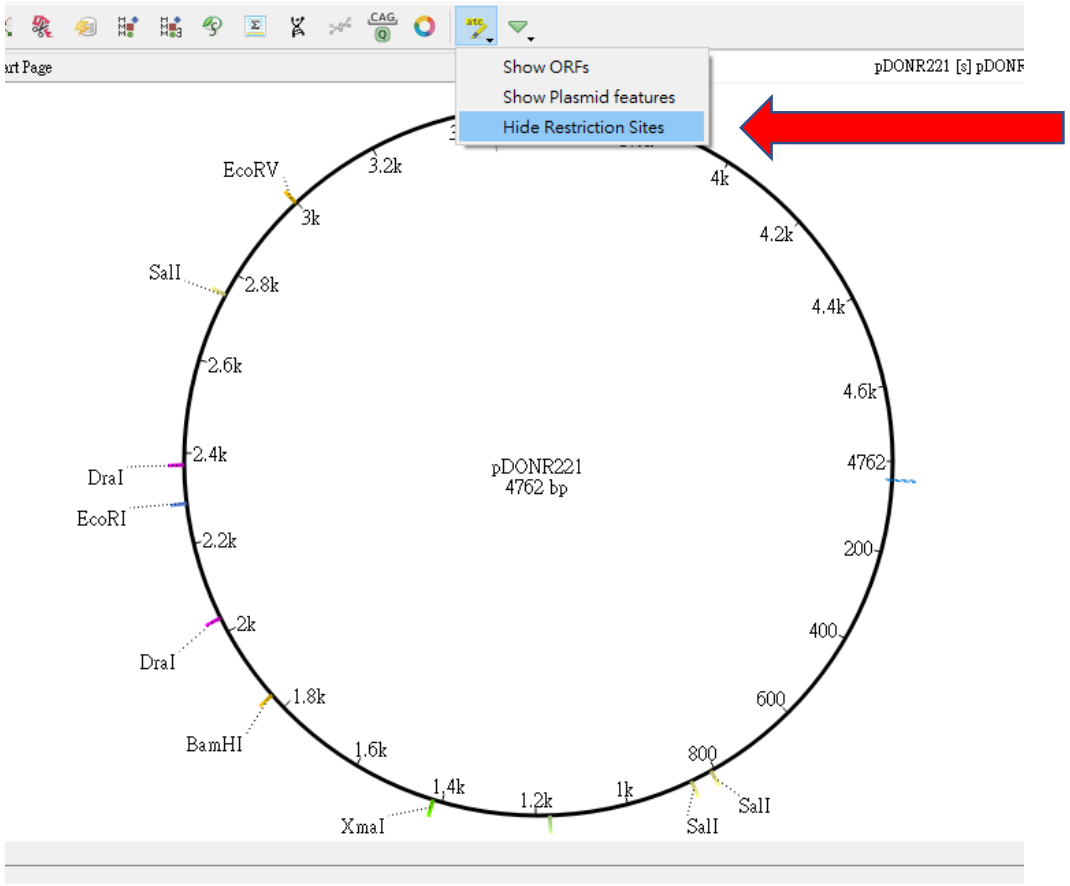
# Fasta Files

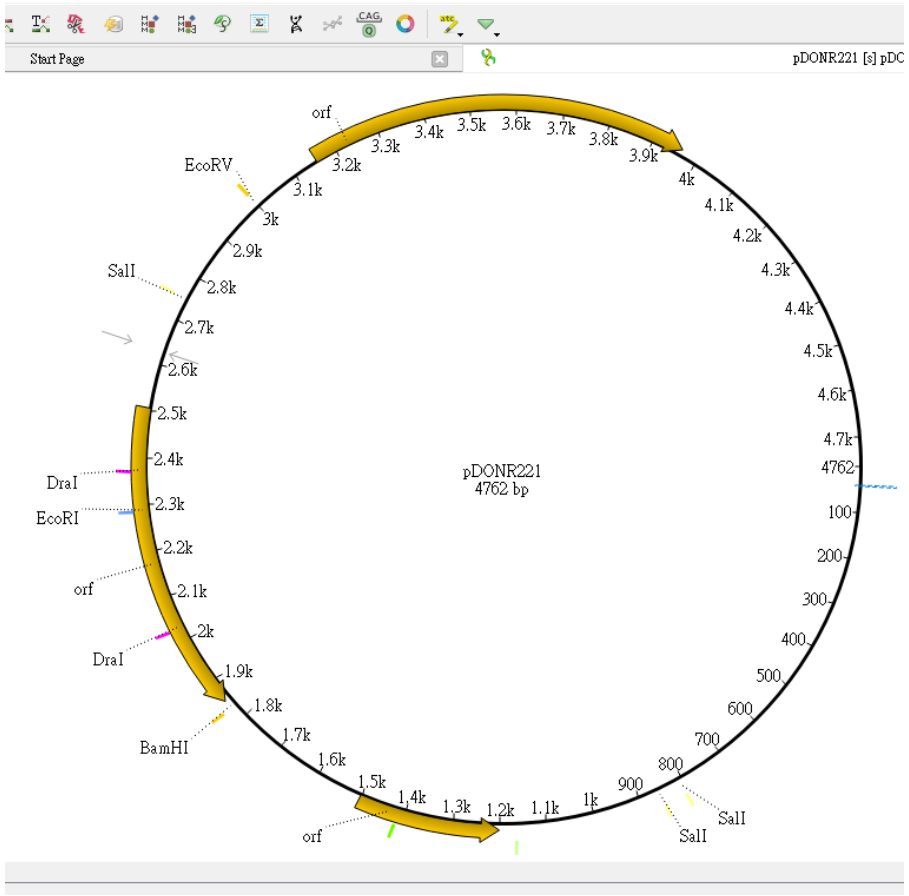
- >seq_name
- atggtgcatgcatgt

# Annotations Editor

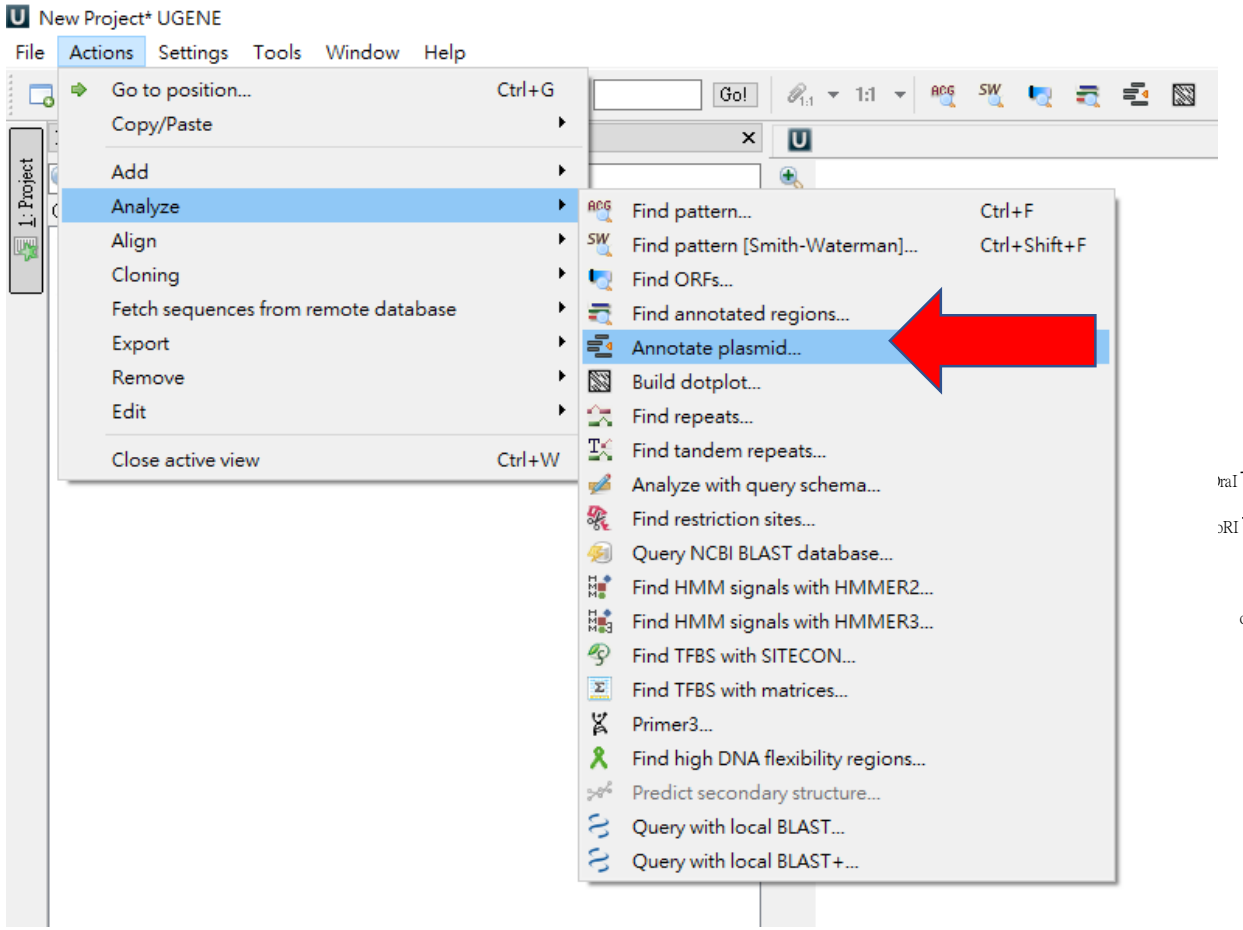


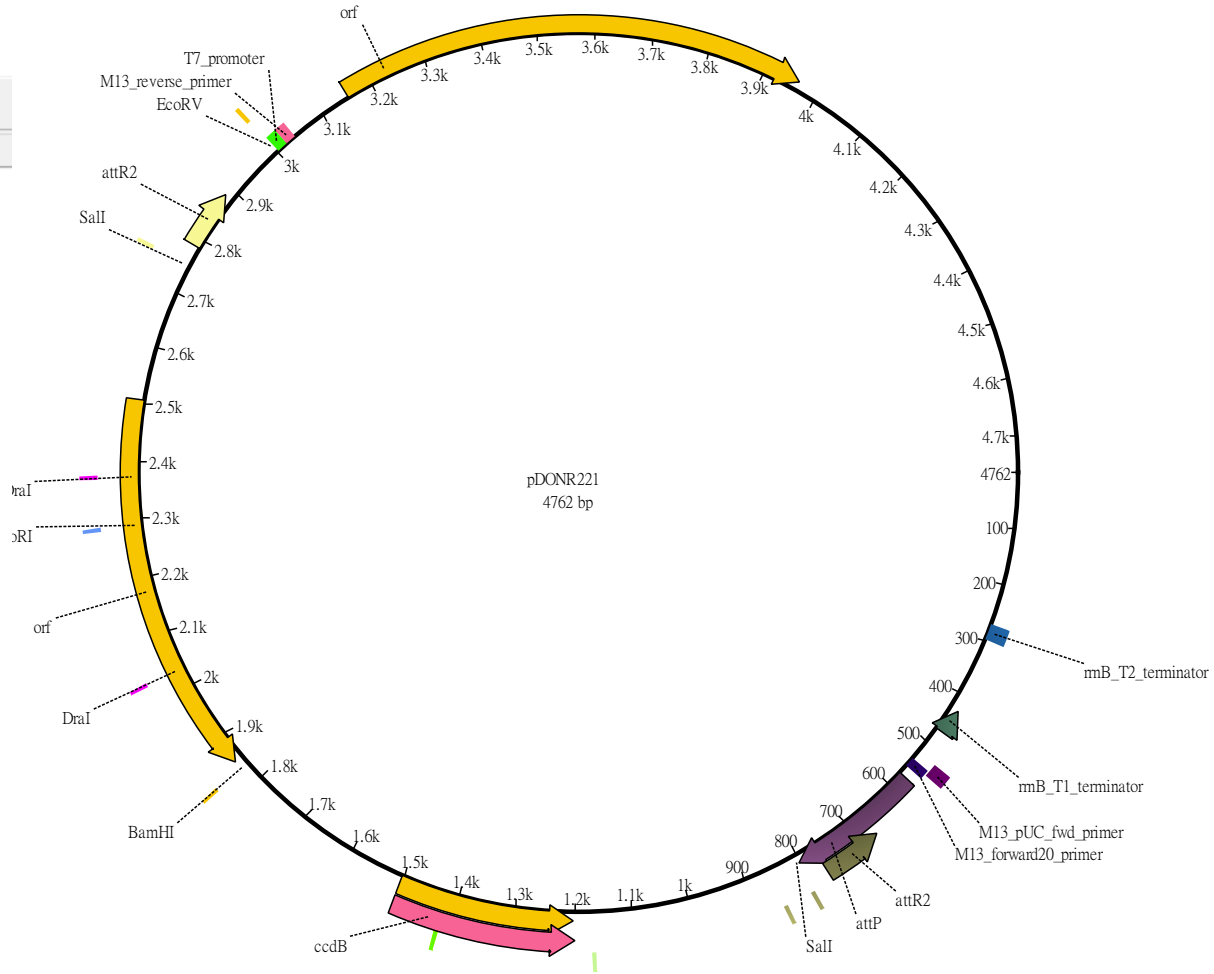
# Plasmid Annotation



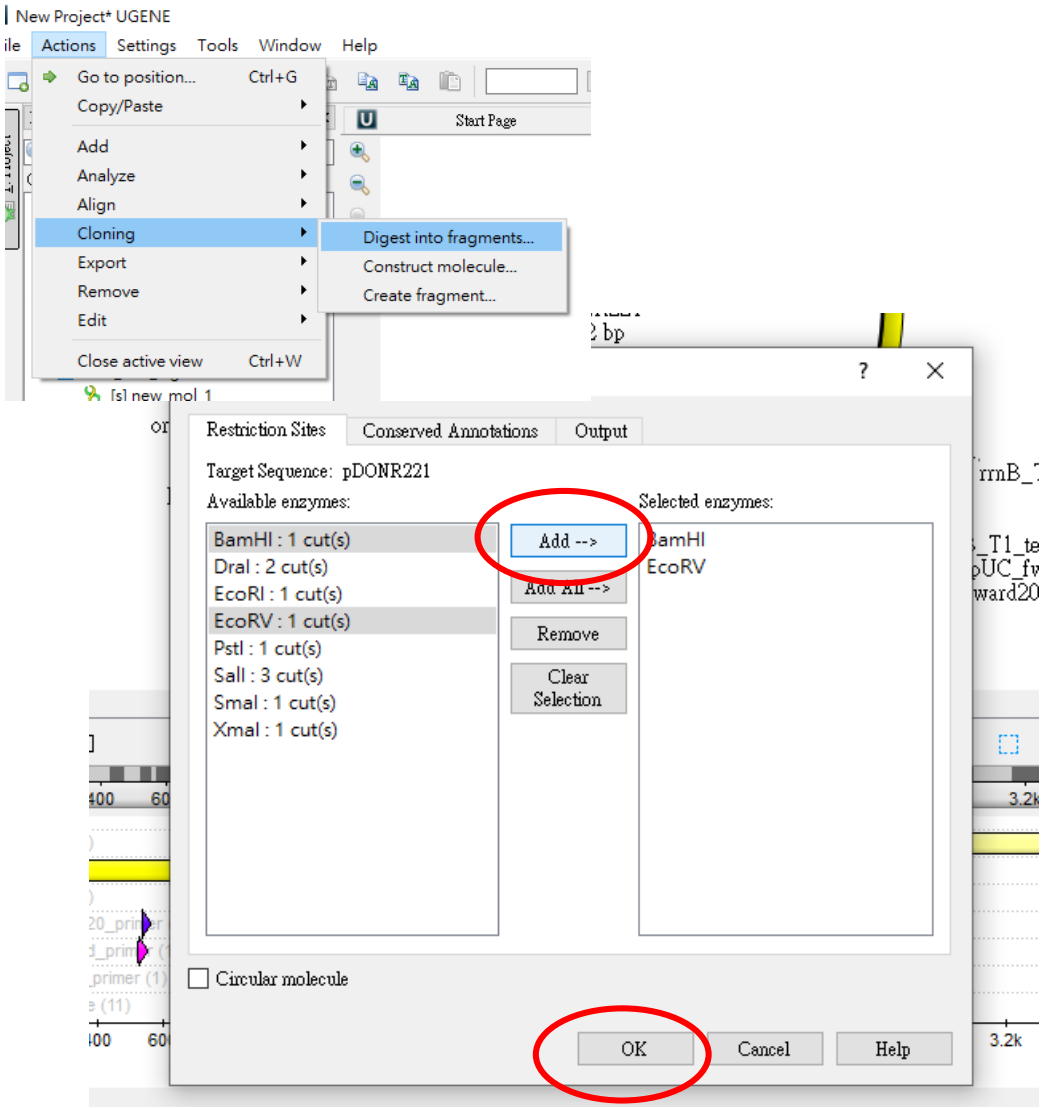



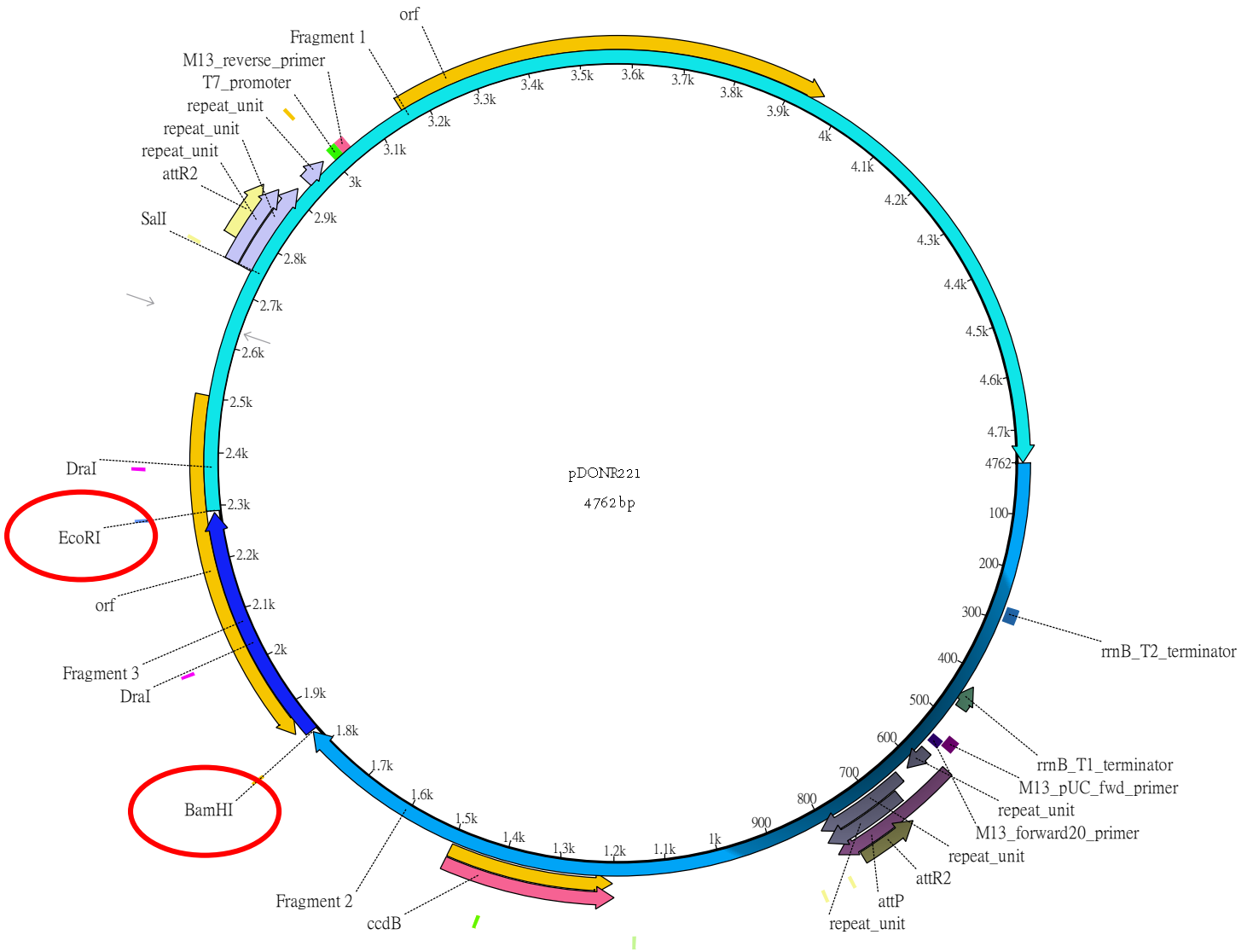
# Plasmid Annotation

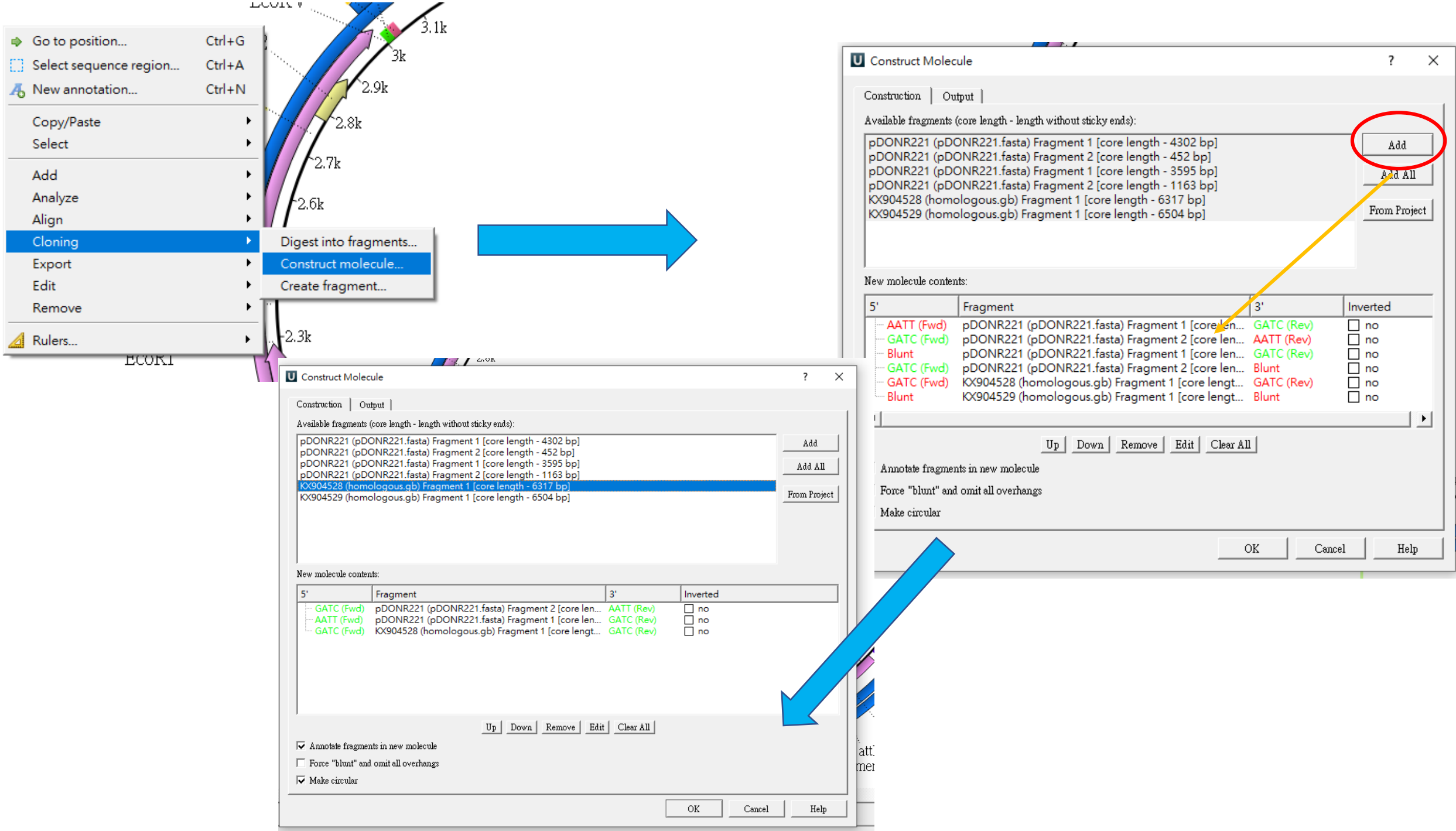


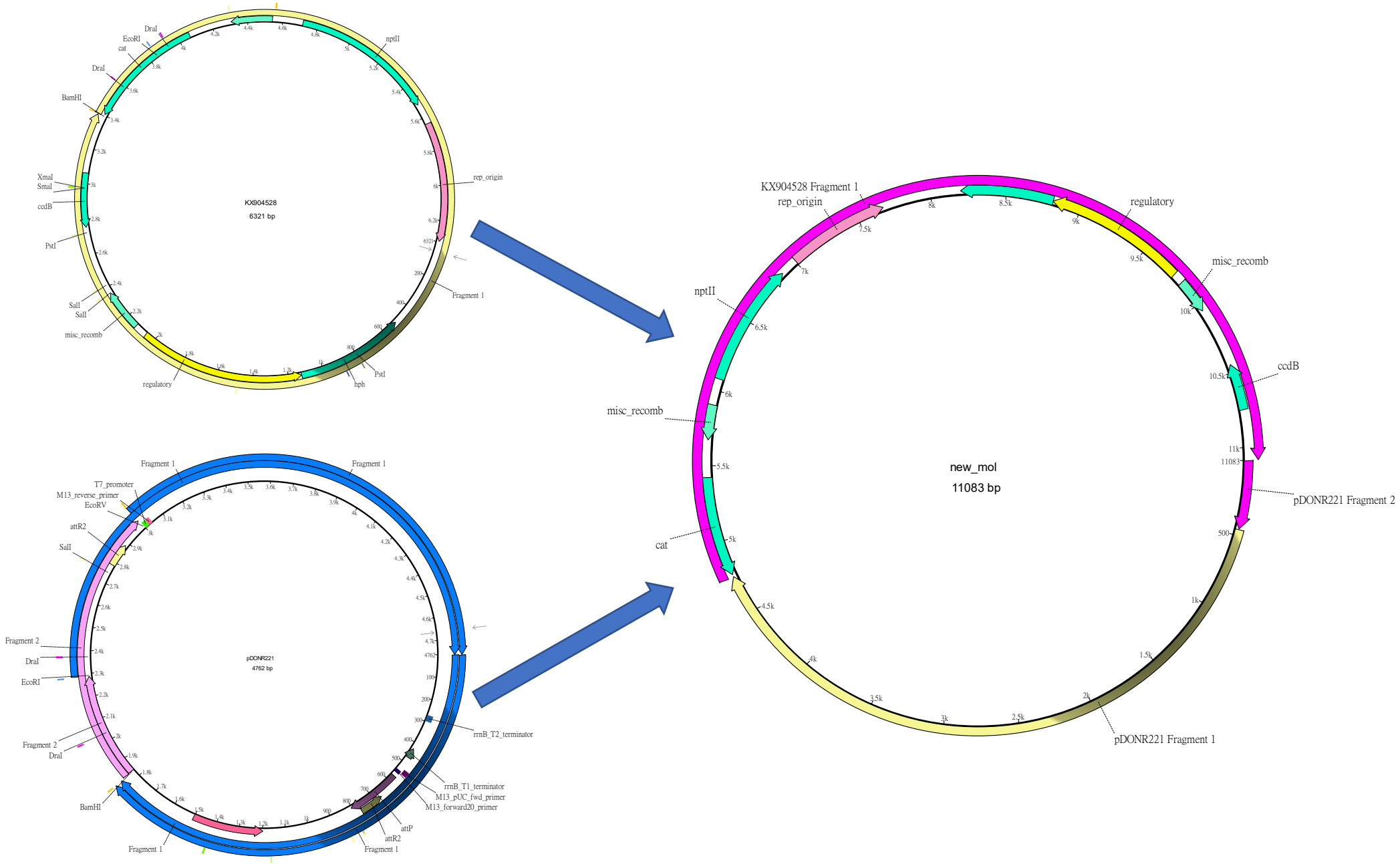



# Cloning


New Project* UGENE
File
Actions
Settings
Tools
Window
Help
Go to position... Ctrl+G
Copy/Paste
Add
Analyze
Align
Cloning
Export
Remove
Edit
Close active view Ctrl+W
Start Page
Digest into fragments...
Construct molecule...
Create fragment...
Restriction Sites
Conserved Annotations
Output
Target Sequence: pDONR221
Available enzymes:
BamHI : 1 cut(s)
DraI : 2 cut(s)
EcoRI : 1 cut(s)
EcoRV : 1 cut(s)
PstI : 1 cut(s)
SalI : 3 cut(s)
SmaI : 1 cut(s)
XmaI : 1 cut(s)
Add -->
Add All -->
Remove
Clear Selection
Selected enzymes:
BamHI
EcoRV
Circular molecule
OK
Cancel
Help


pDONR221
4762 bp
orf
Fragment 1
M13_reverse_primer
T7_promoter
repeat_unit
repeat_unit
repeat_unit
attR2
SalI
DraI
EcoRI
orf
Fragment 3
DraI
BamHI
Fragment 2
ccdB
rrnB_T2_terminator
rrnB_T1_terminator
M13_pUC_fwd_primer
repeat_unit
M13_forward20_primer
repeat_unit
attR2
attP
repeat_unit


Go to position... Ctrl+G
Select sequence region... Ctrl+A
New annotation... Ctrl+N
Copy/Paste
Select
Add
Analyze
Align
Cloning
Export
Edit
Remove
Rulers...
Digest into fragments...
Construct molecule...
Create fragment...
3.1k
3k
2.9k
2.8k
2.7k
2.6k
2.3k
ECORI
Construct Molecule
Construction
Output
Available fragments (core length - length without sticky ends):
pDONR221 (pDONR221.fasta) Fragment 1 [core length - 4302 bp]
pDONR221 (pDONR221.fasta) Fragment 2 [core length - 452 bp]
pDONR221 (pDONR221.fasta) Fragment 1 [core length - 3595 bp]
pDONR221 (pDONR221.fasta) Fragment 2 [core length - 1163 bp]
KX904528 (homologous.gb) Fragment 1 [core length - 6317 bp]
KX904529 (homologous.gb) Fragment 1 [core length - 6504 bp]
Add
Add All
From Project
New molecule contents:
5' Fragment 3' Inverted
AATT (Fwd) pDONR221 (pDONR221.fasta) Fragment 1 [core len... GATC (Rev) no
GATC (Fwd) pDONR221 (pDONR221.fasta) Fragment 2 [core len... AATT (Rev) no
Blunt pDONR221 (pDONR221.fasta) Fragment 1 [core len... GATC (Rev) no
GATC (Fwd) pDONR221 (pDONR221.fasta) Fragment 2 [core len... Blunt no
GATC (Fwd) KX904528 (homologous.gb) Fragment 1 [core lengt... GATC (Rev) no
Blunt KX904529 (homologous.gb) Fragment 1 [core lengt... Blunt no
Up
Down
Remove
Edit
Clear All
Annotate fragments in new molecule
Force "blunt" and omit all overhangs
Make circular
OK
Cancel
Help


KX904528
6321 bp
DraI
EcoRI
cat
DraI
BamHI
XmaI
SmaI
ccdB
PstI
SalI
SalI
misc_recomb
regulatory
hph
PstI
nptII
rep_origin
Fragment 1
pDONR221
4762 bp
Fragment 1
Fragment 1
T7_promoter
M13_reverse_primer
EcoRV
attR2
SalI
Fragment 2
DraI
EcoRI
Fragment 2
DraI
BamHI
Fragment 1
Fragment 1
rrnB_T2_terminator
rrnB_T1_terminator
M13_pUC_fwd_primer
M13_forward20_primer
attP
attR2
new_mol
11083 bp
KX904528 Fragment 1
rep_origin
regulatory
misc_recomb
nptII
misc_recomb
ccdB
cat
pDONR221 Fragment 2
pDONR221 Fragment 1

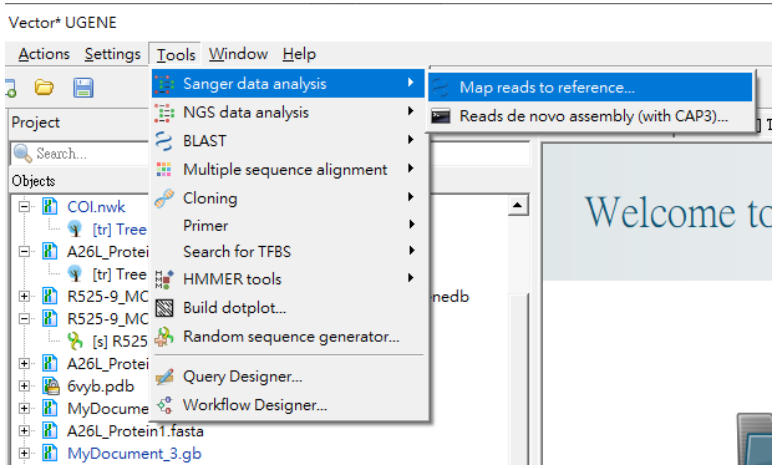
# Chromatogram Viewer

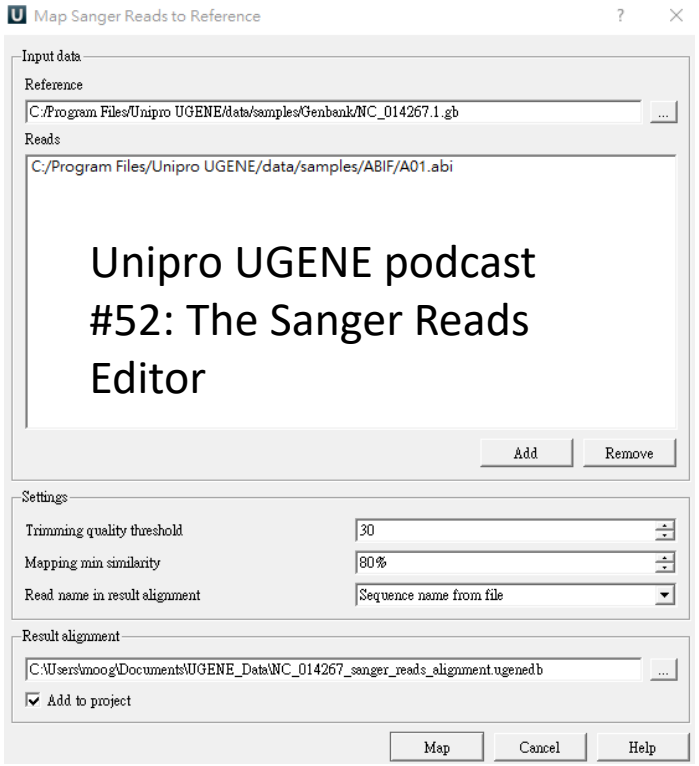
# Contigs assembly with CAP3

- CAP3 (CONTIG ASSEMBLY PROGRAM Version 3) is a sequence assembly program for small-scale assembly with or without quality values.
- Open *Tools ▸ Sanger data analysis* submenu of the main menu.

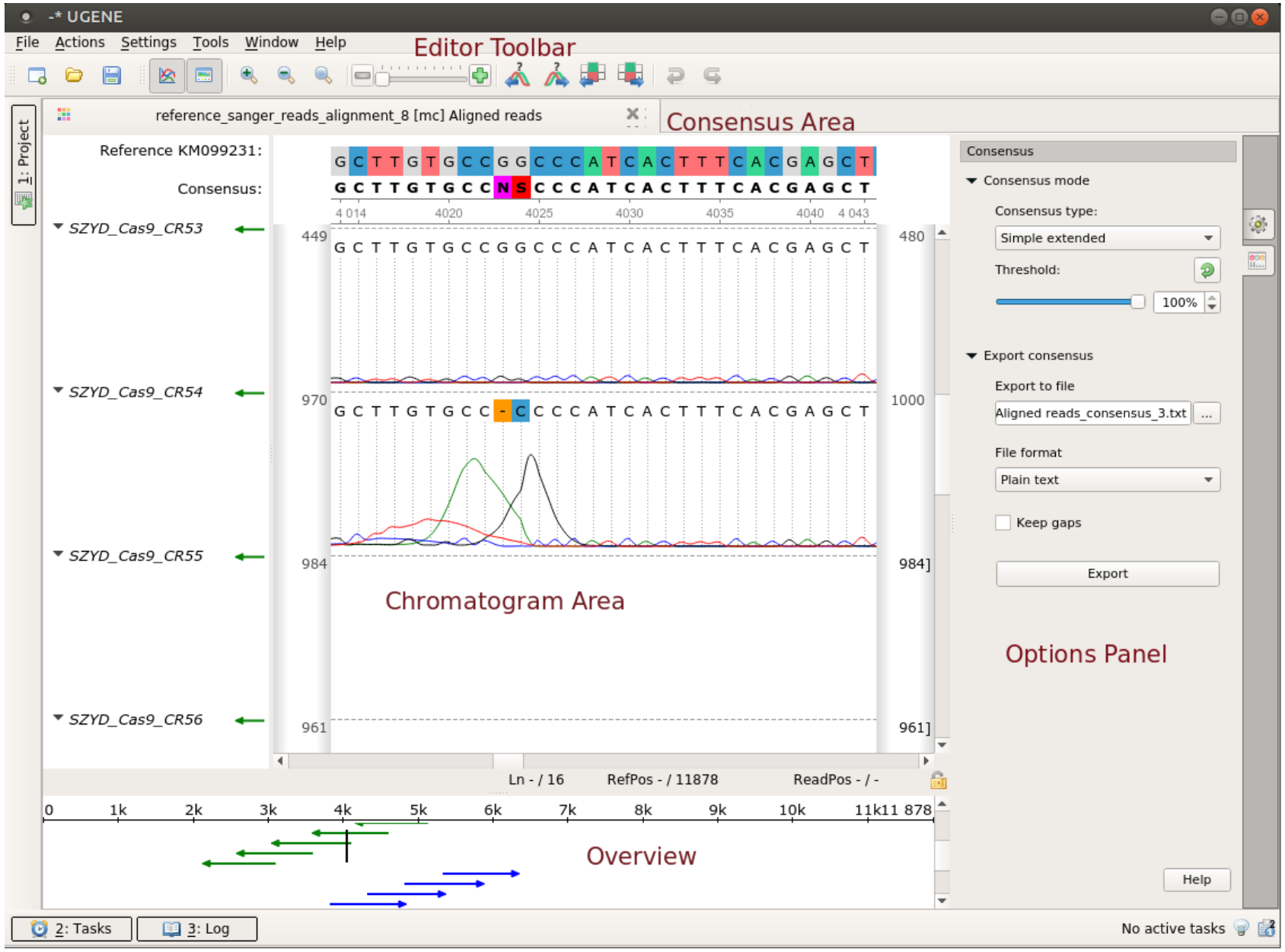
- *Input files* are files with a long DNA reads in FASTA, FASTQ, SCF or ABI formats.
- *CAP3* produces assembly results in the ACE file format (".ace"). The file contains one or several contigs assembled from the input reads.

# Sanger Reads Editor Components




Unipro UGENE podcast #52: The Sanger Reads Editor



# Local and  BLAST search

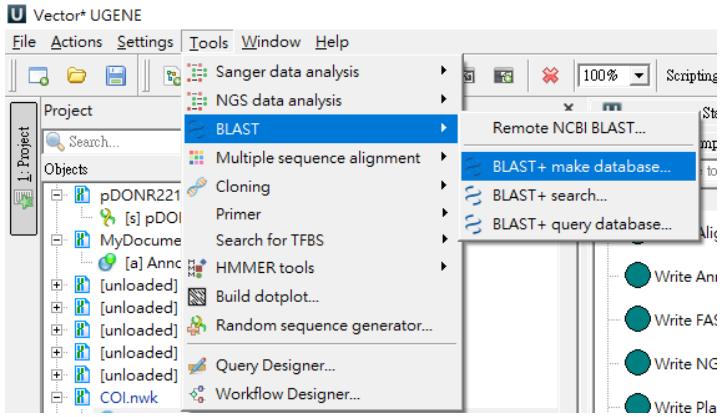
## BLAST Protocols

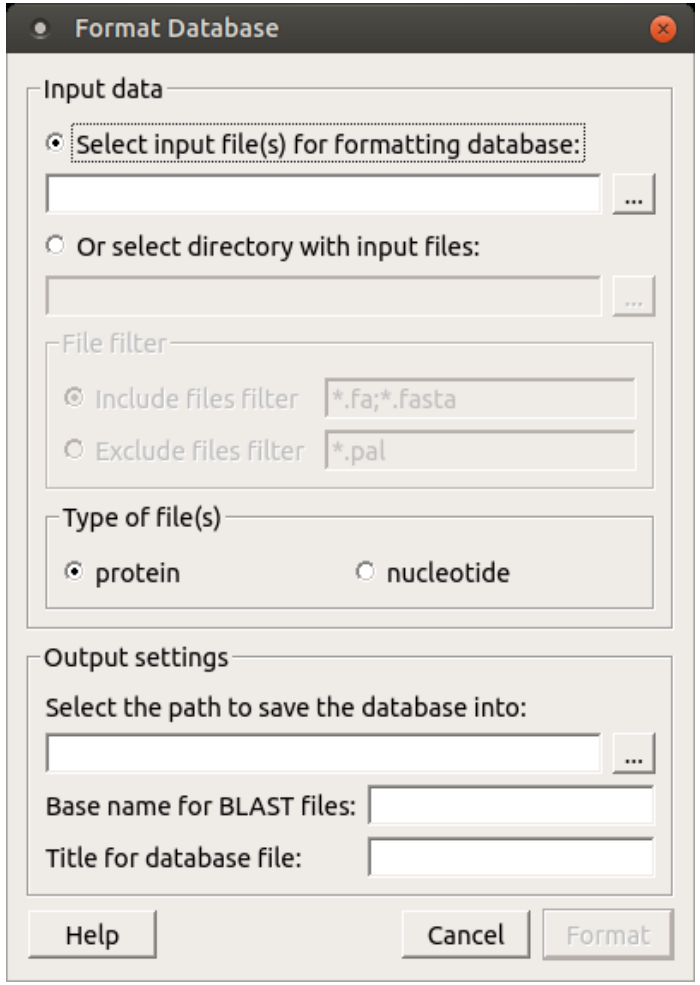
- The most common BLAST search includes five protocols:

| Program | Database | Query |
|---|---|---|
| BLASTN | Nucleotide | Nucleotide |
| BLASTP | Protein | Protein |
| BLASTX | Protein | Nt. → Protein |
| TBLASTN | Nt. → Protein | Protein |
| TBLASTX | Nt. → Protein | Nt. → Protein |

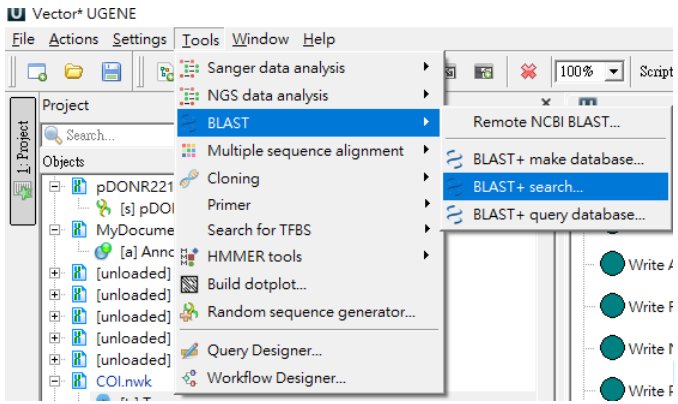
# Creating Database

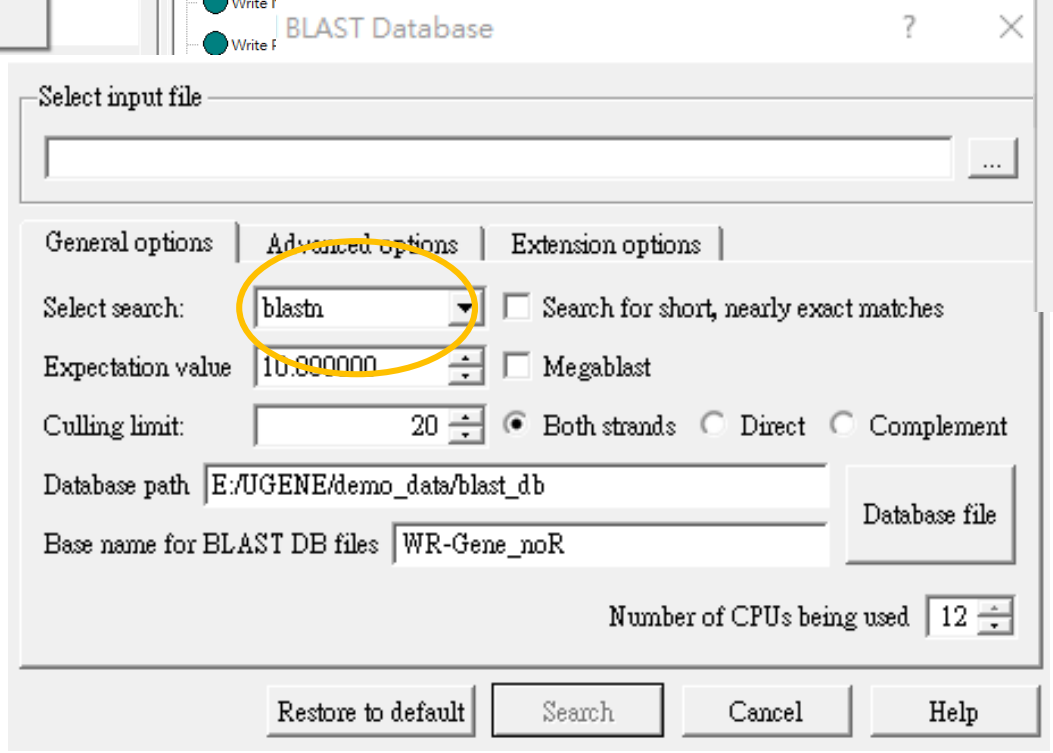
- using *BLAST+* open *Tools* ▸ *BLAST* ▸ *BLAST+ make DB*.
- You can choose either *protein* or *nucleotide* type of the files.
- Then you must select the path to save the database file and specify a *Base name for BLAST files* and a *Title for database file*.

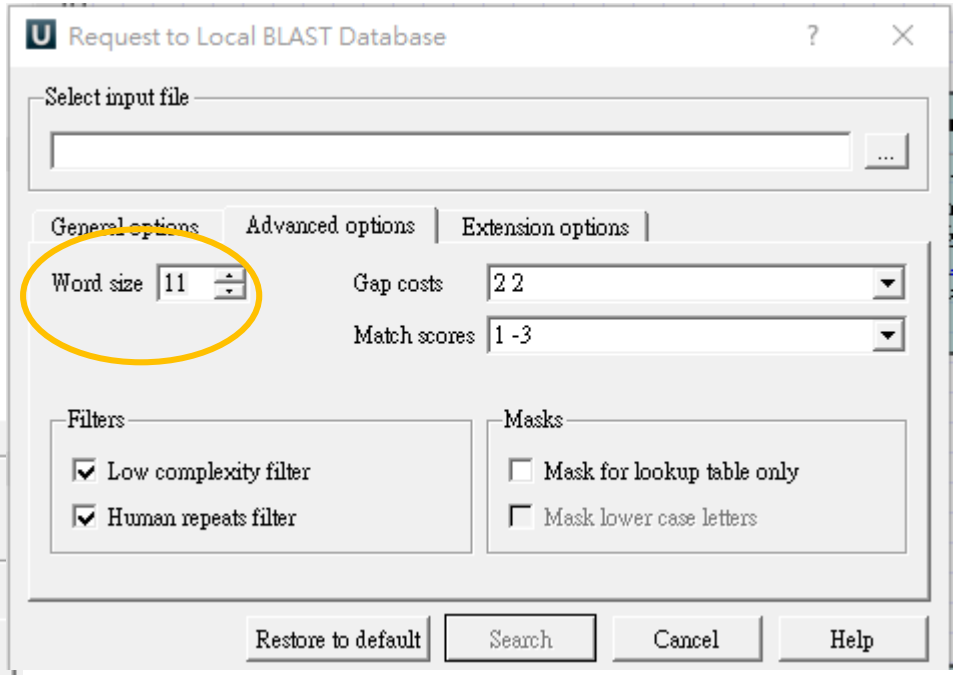


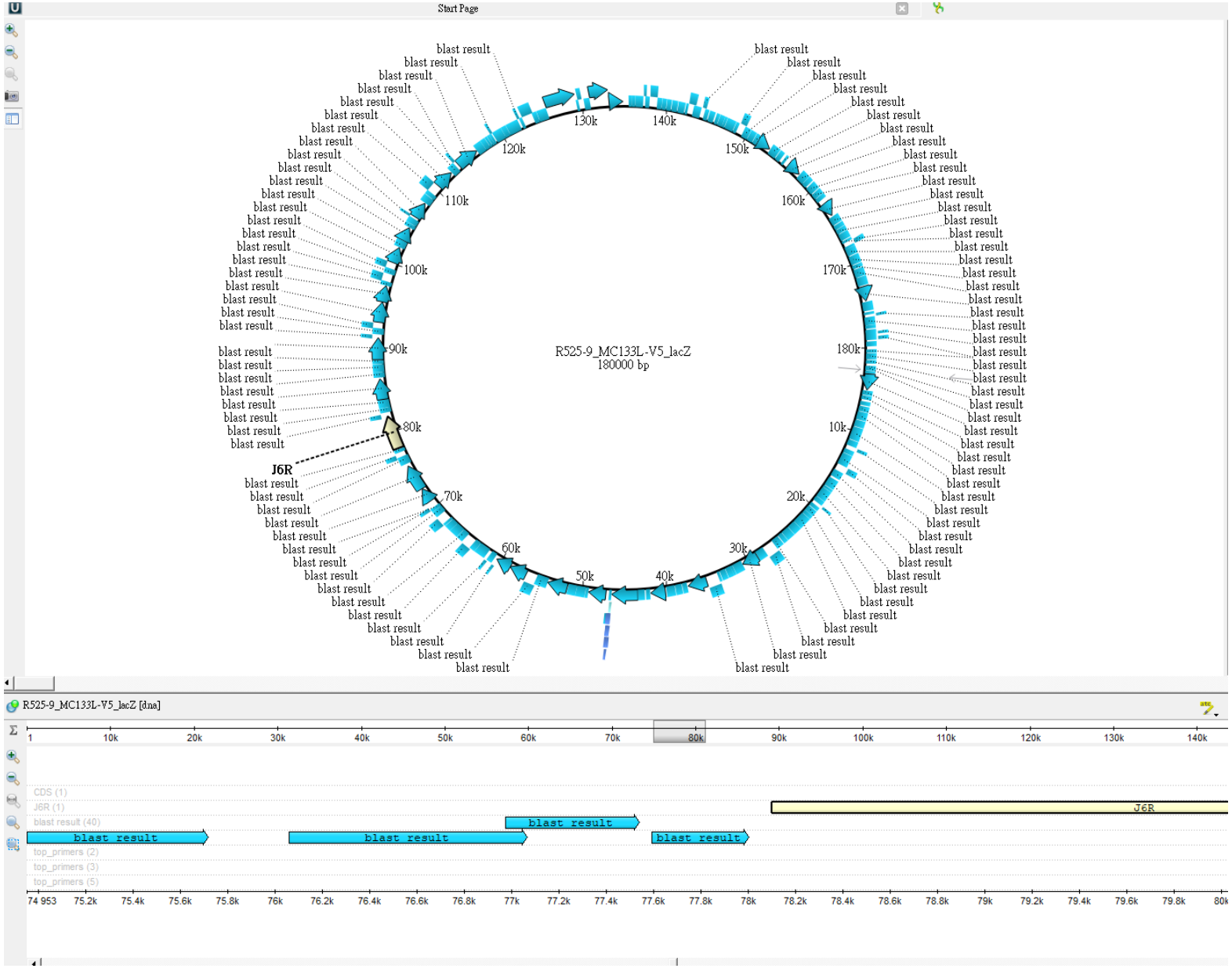
# BLAST search



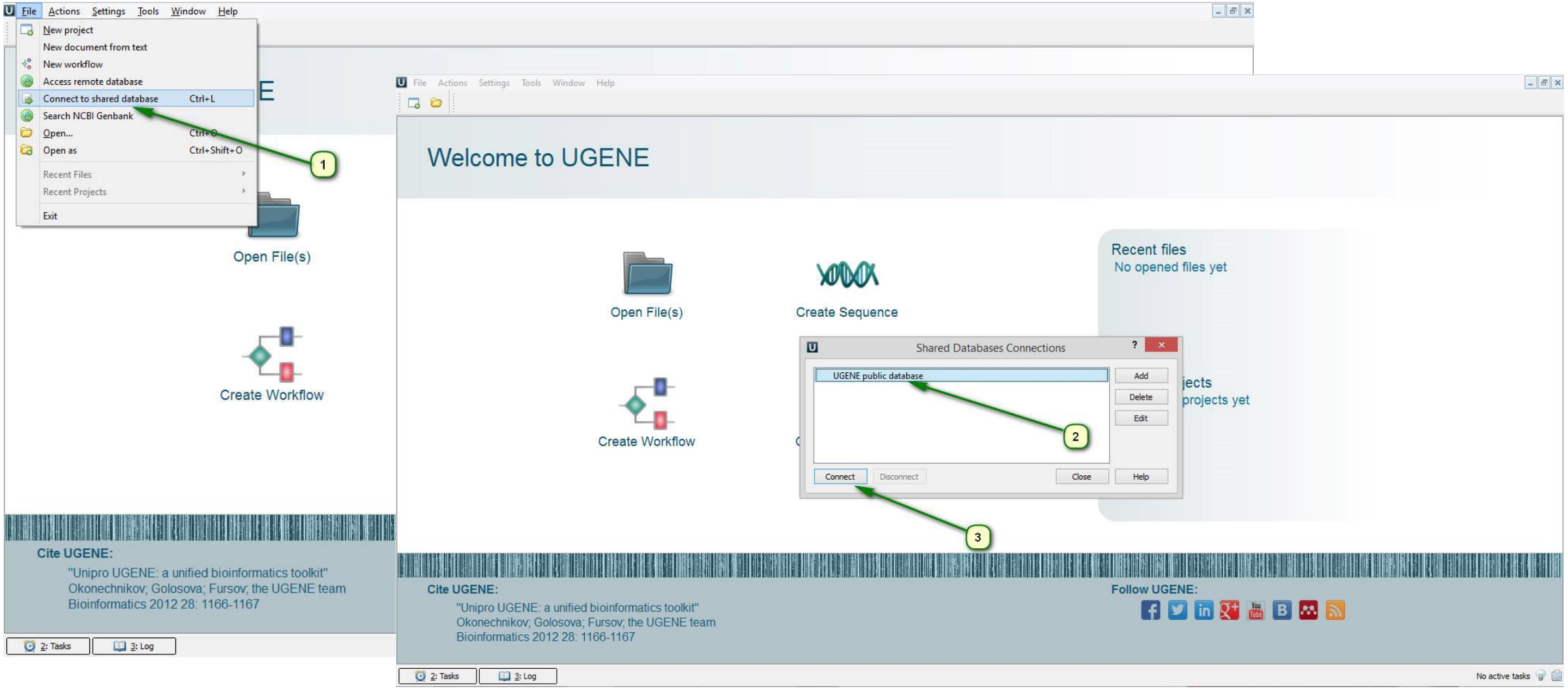



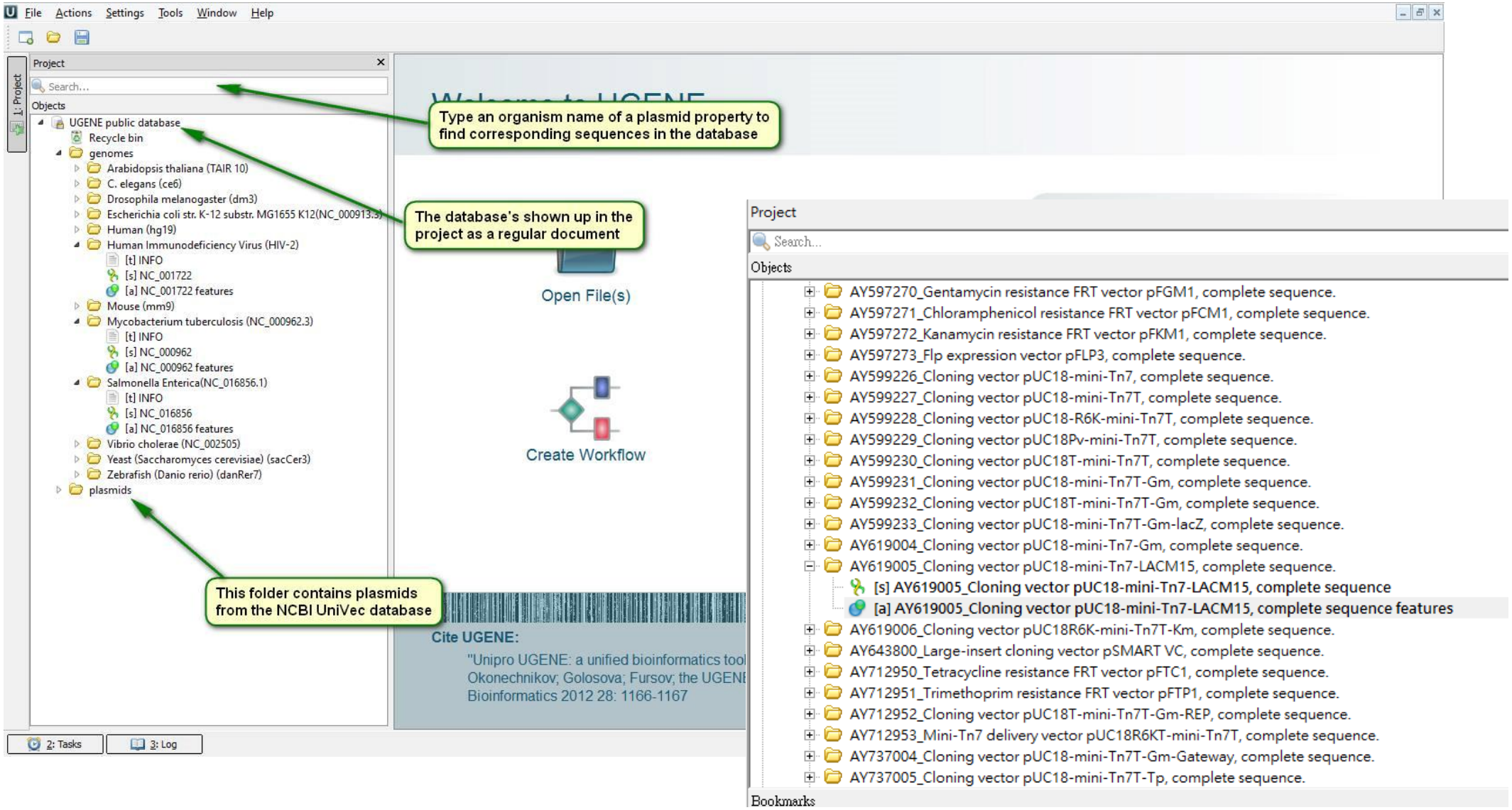
# BLAST search result



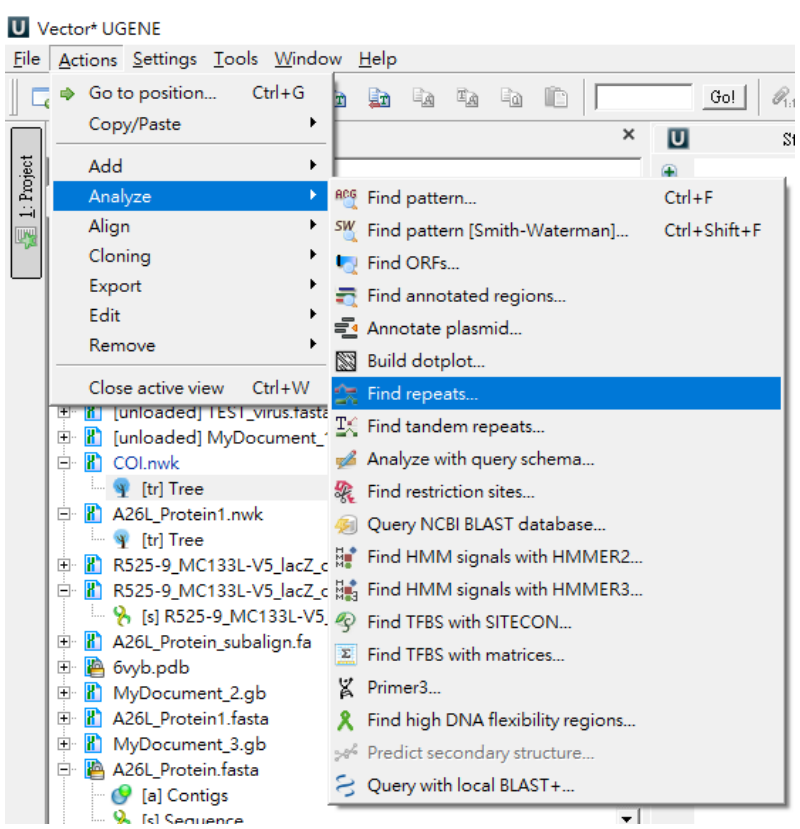
# Making Request to **Public Databases** Database

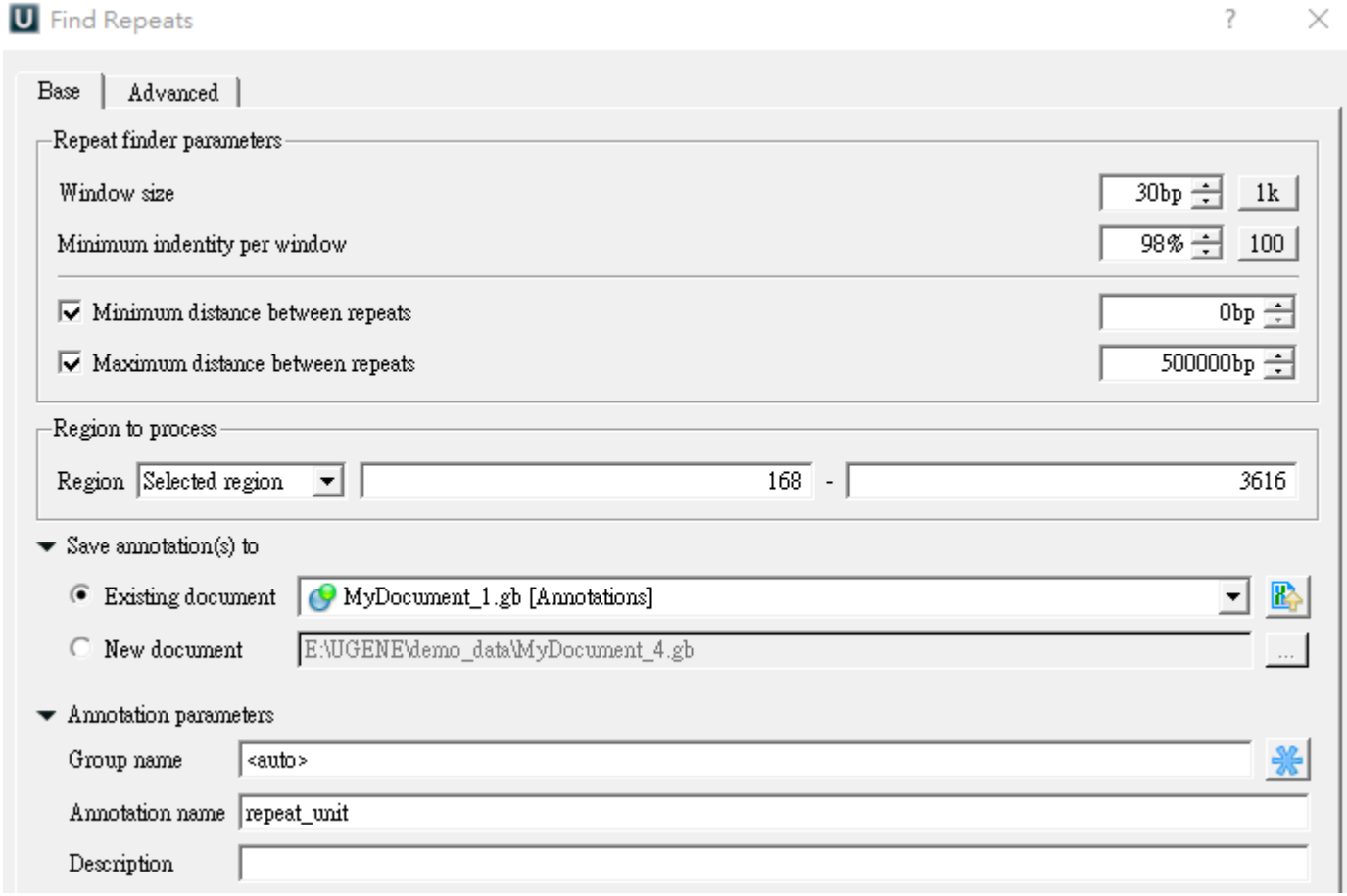


# Making Request to **Public Databases** Database

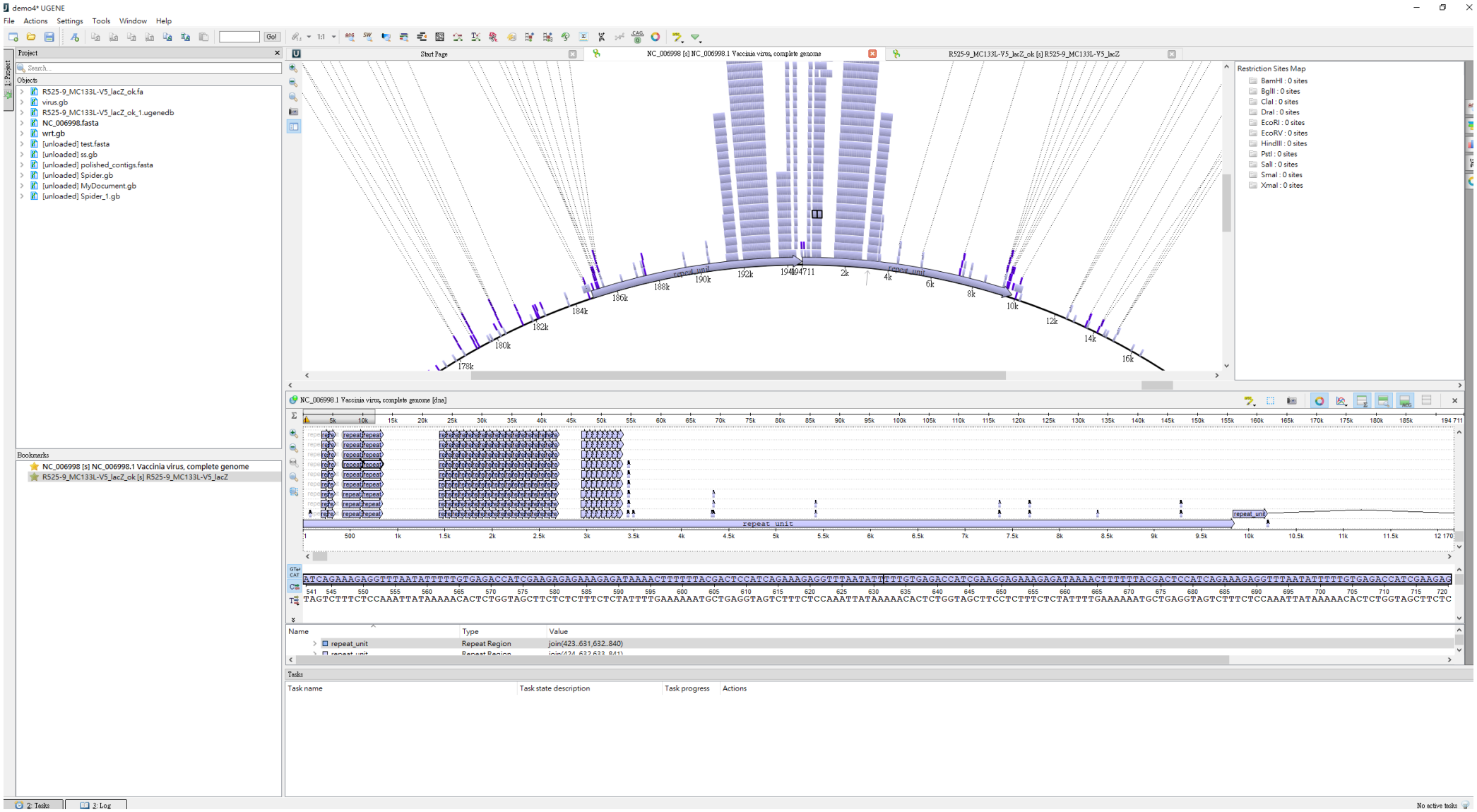


# Search for Sequence Repeats

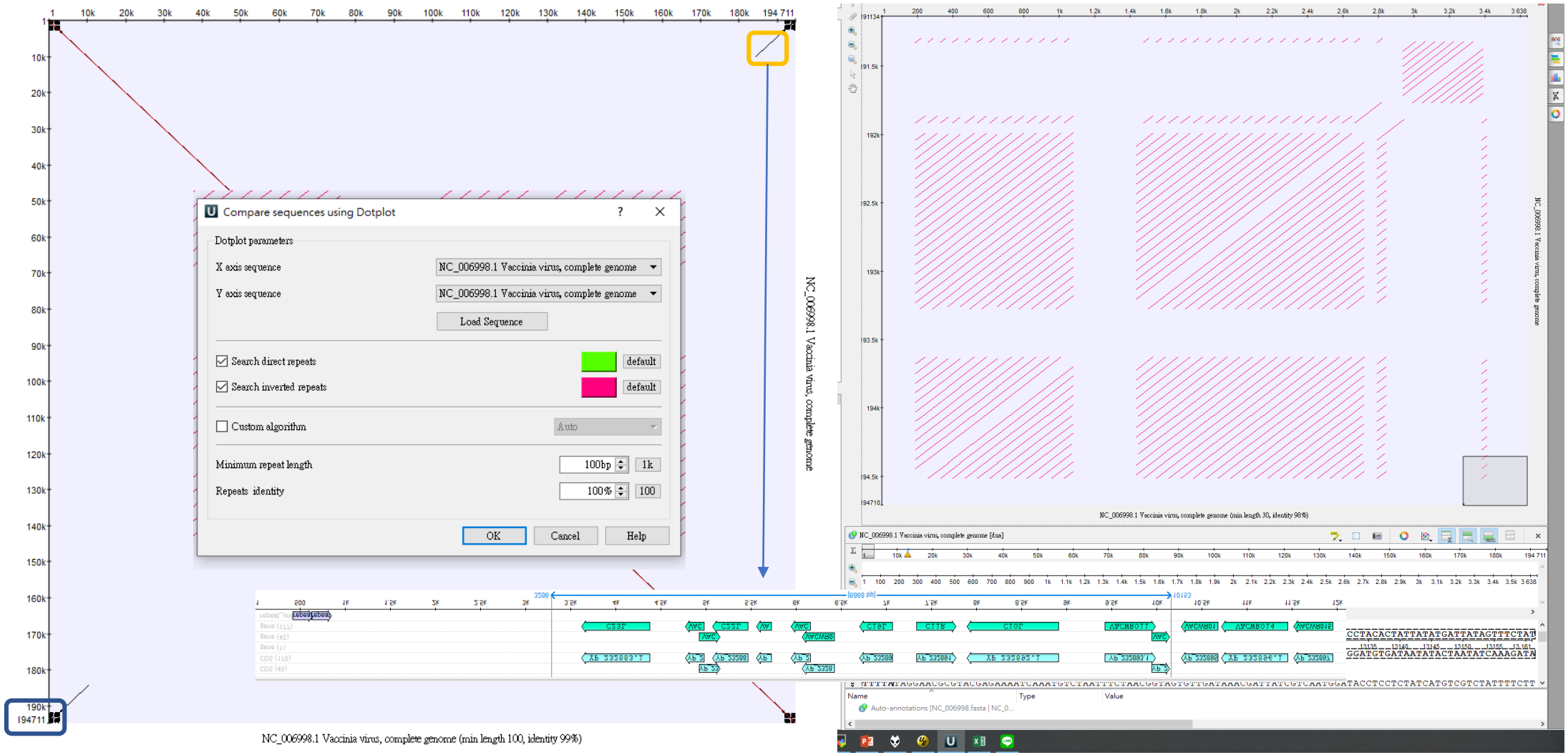


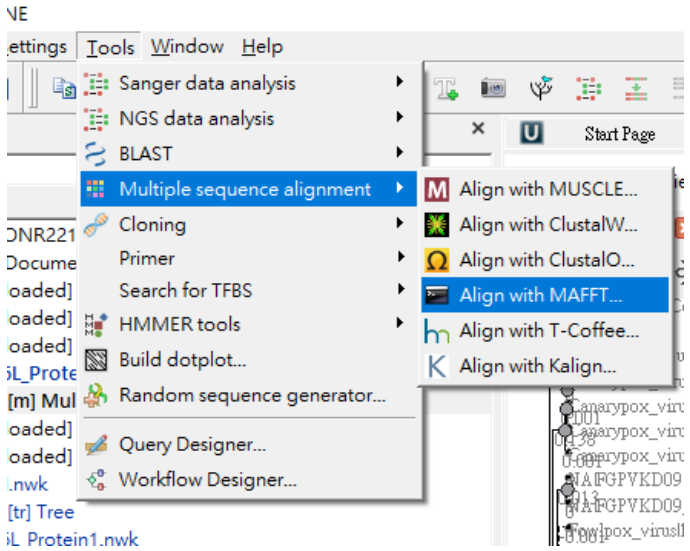
# Search for Sequence Repeats



# Search for Sequence Repeats using Dotplot

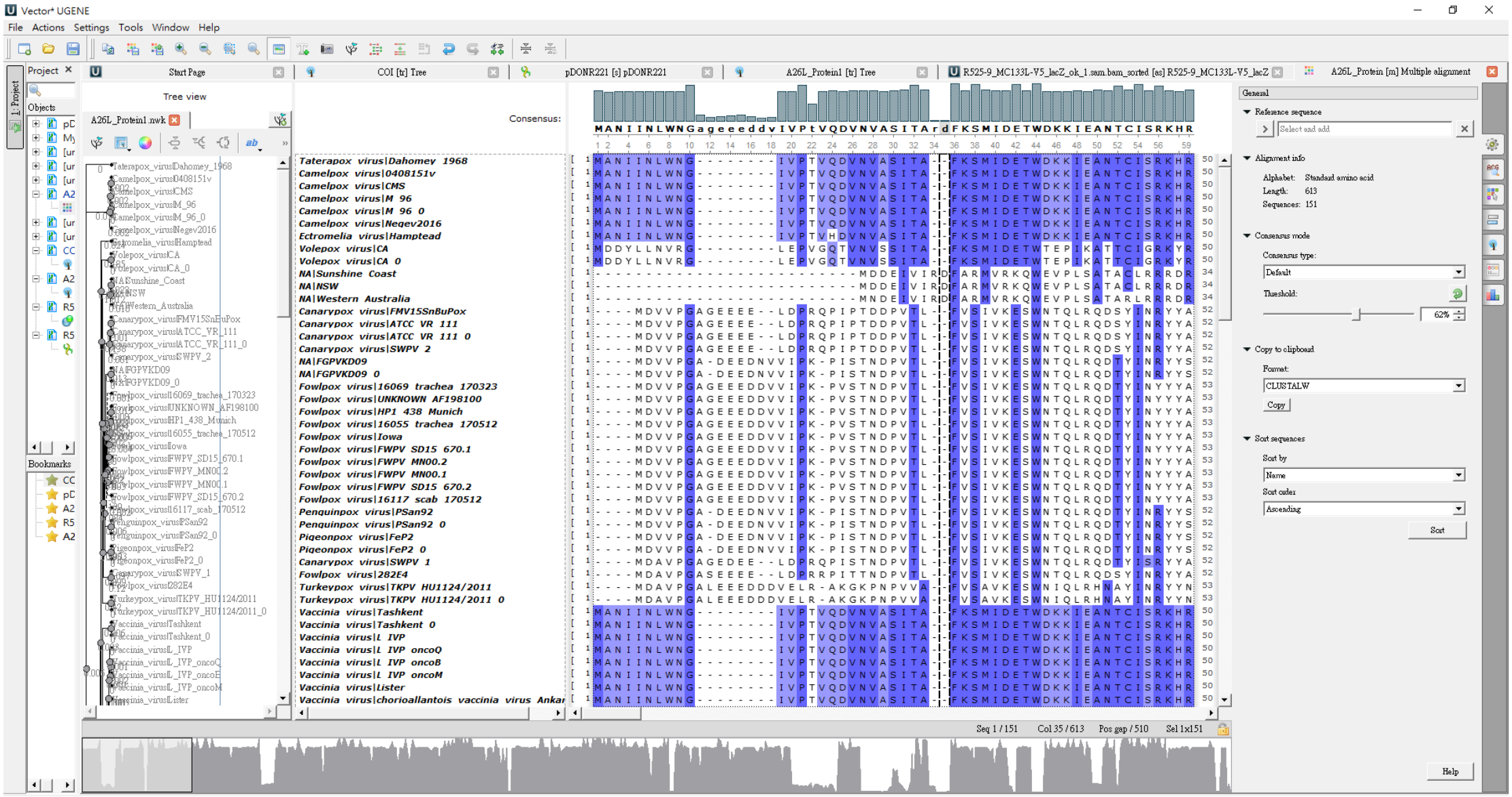


# Multiple sequence alignment

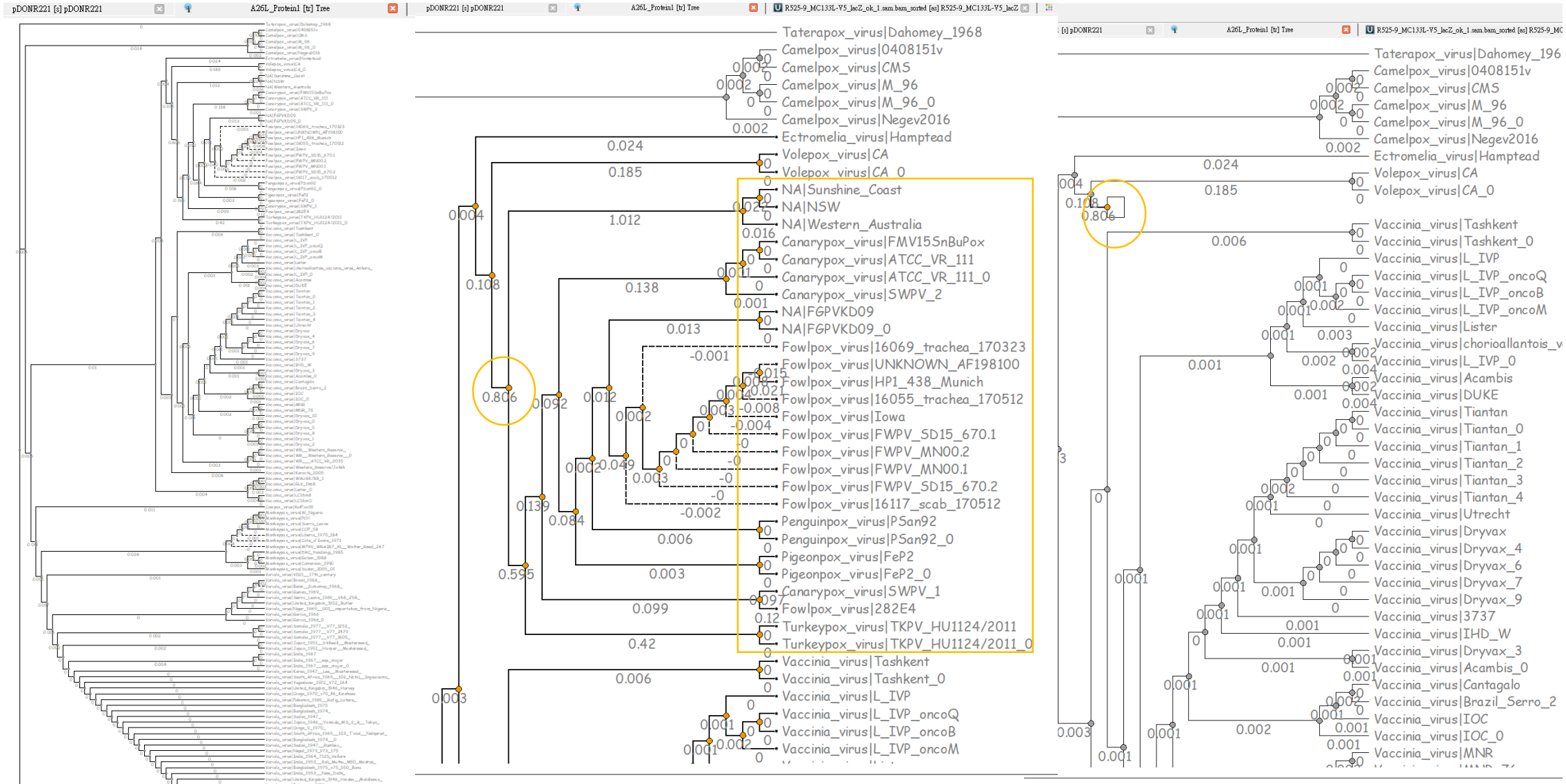



- Clustal is a widely used multiple sequence alignment program. It is used for both nucleotide and protein sequences.
- Kalign is a fast and accurate multiple sequence package designed to align large numbers of protein sequences.
- MAFFT  is used for both nucleotide and protein sequences.

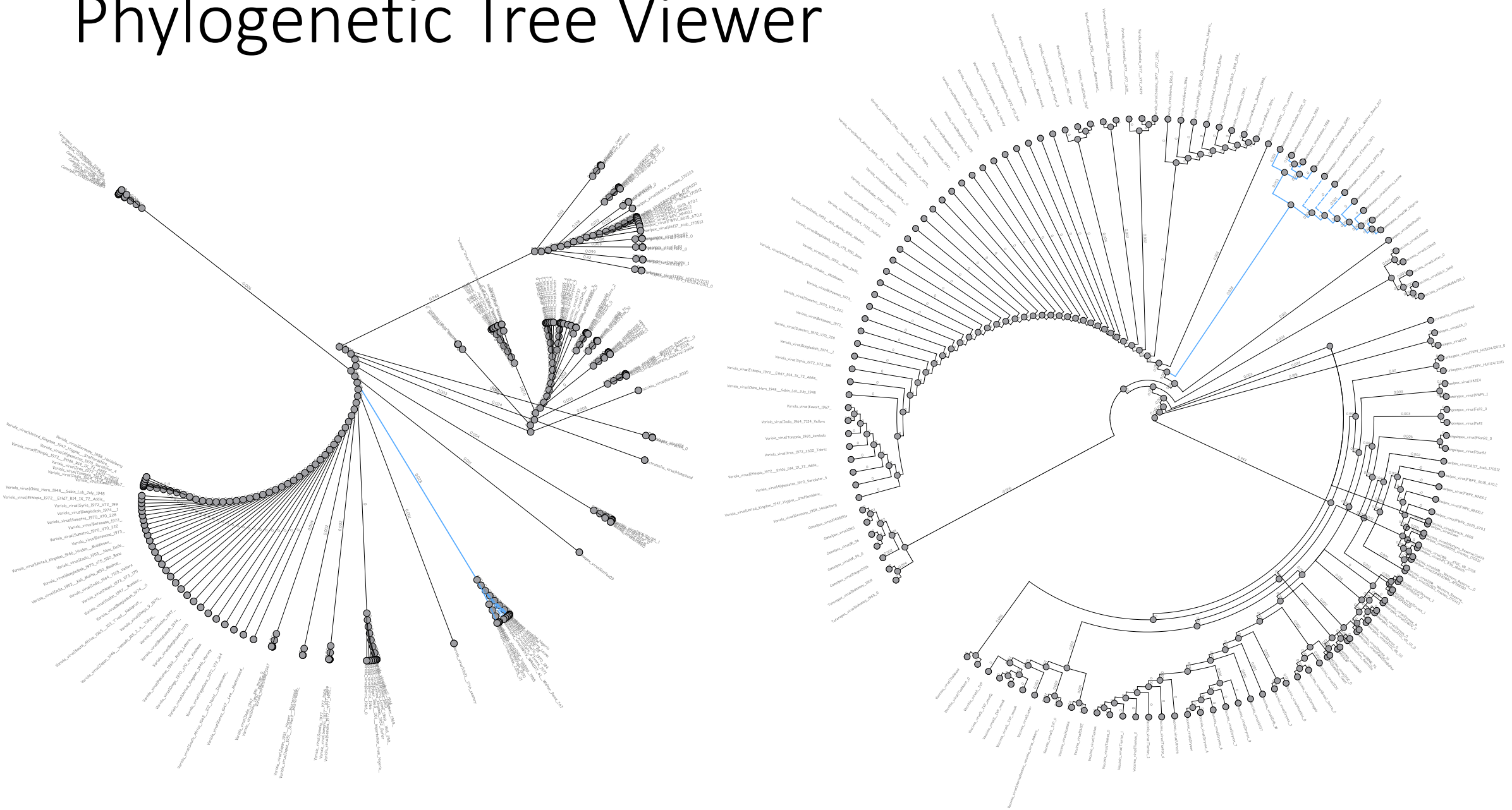
# Alignment Editor



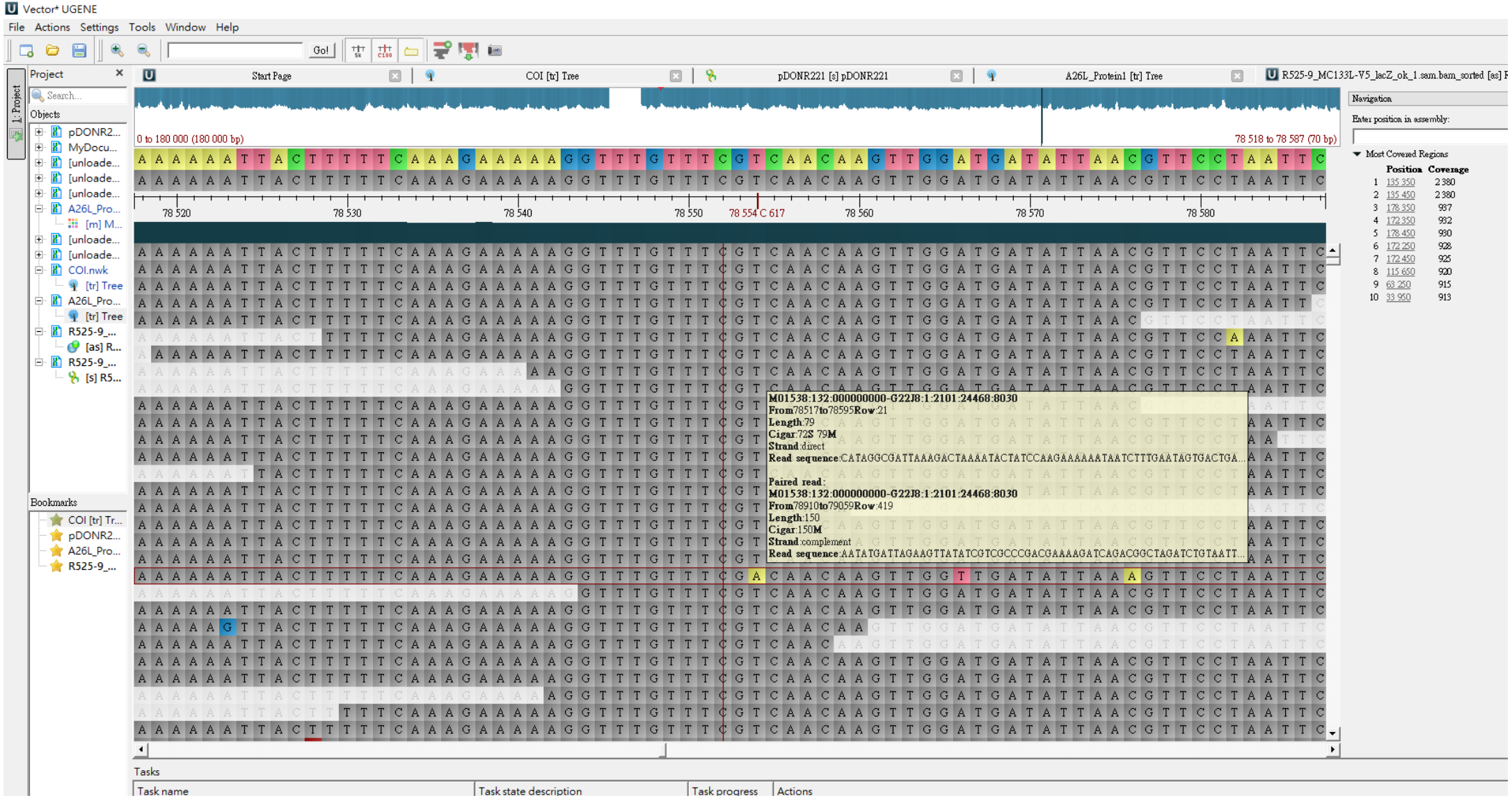
# Phylogenetic Tree Viewer



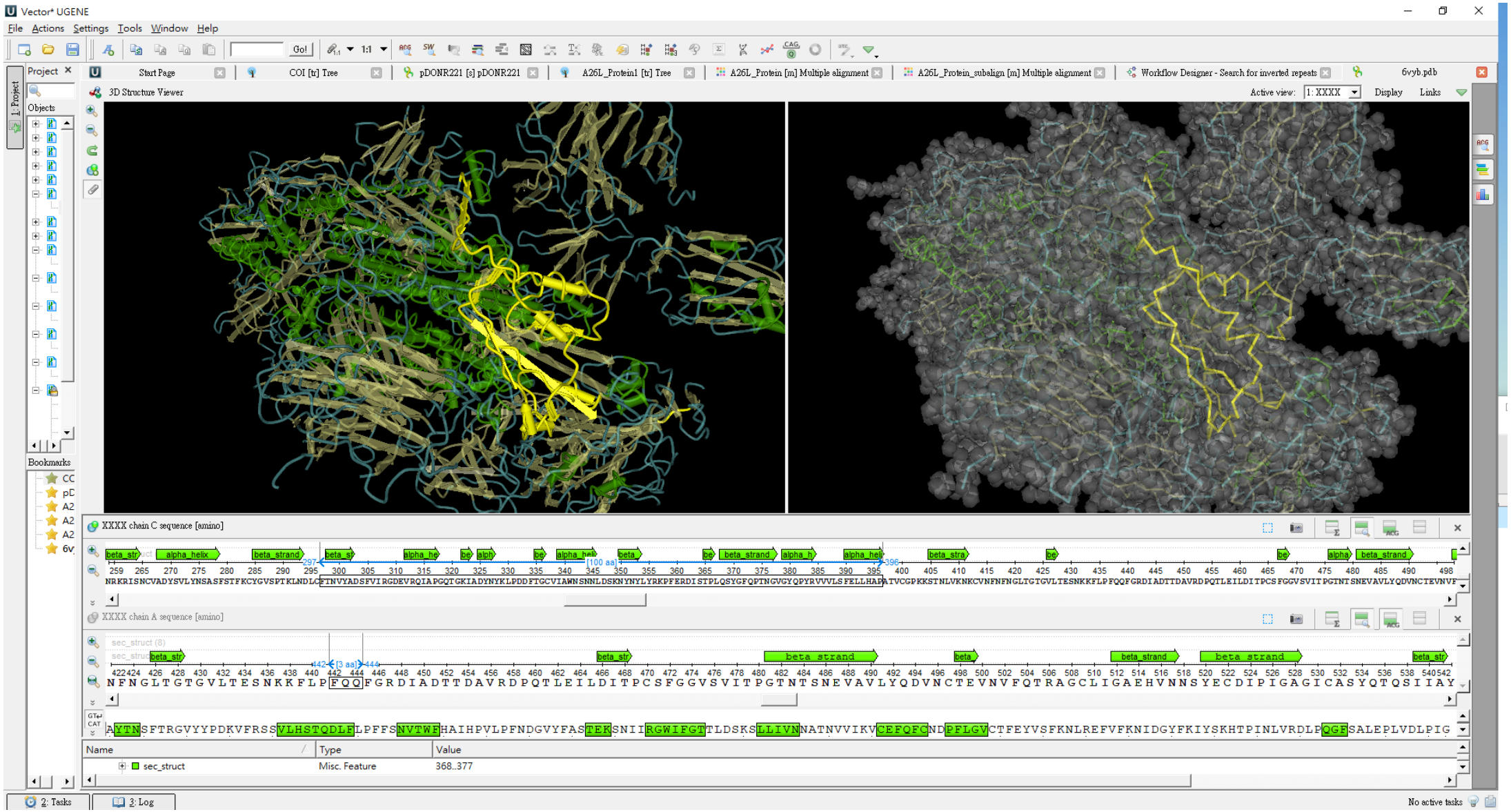
# Phylogenetic Tree Viewer

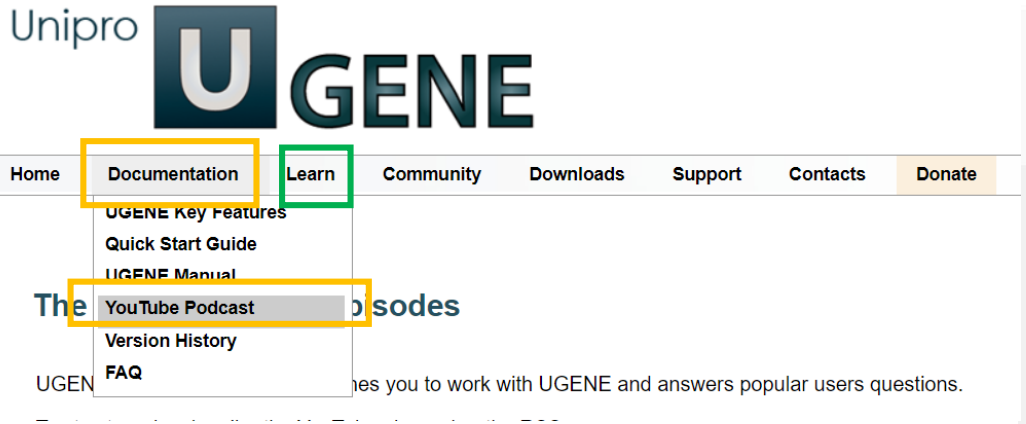
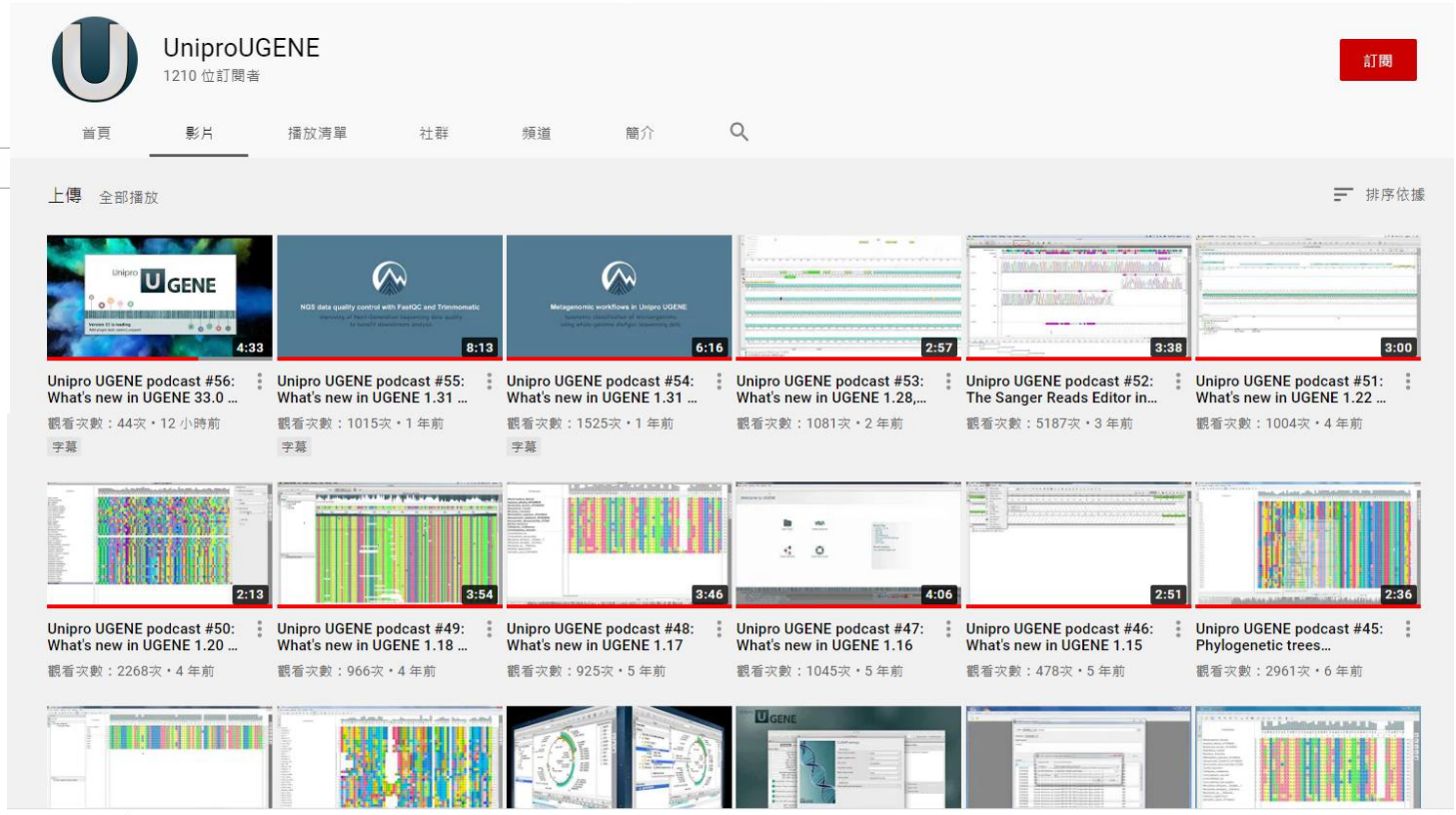


# Assembly Browser



# 3D Structure Viewer



# Tutorials

- https://www.youtube.com/channel/UCPYGoqiRgG4iMU-NgnCGxEg