



Introduction of UGENE

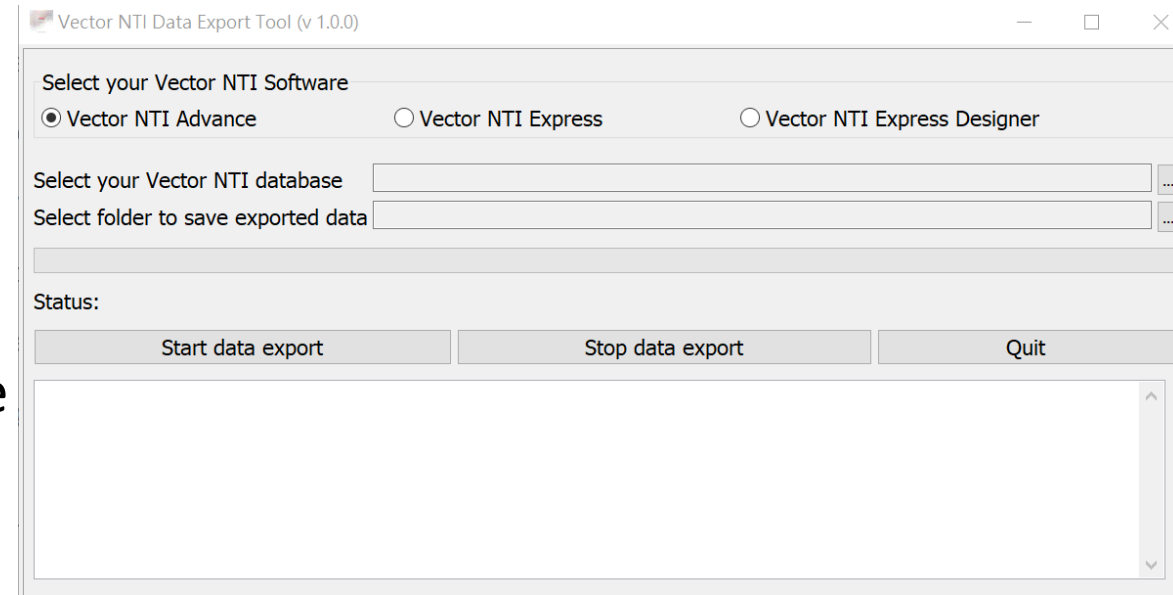
BBSC Bioinformatics Core

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

The screenshot displays the Annotations Editor for the murine NC_001363 sequence. The interface is divided into several sections:

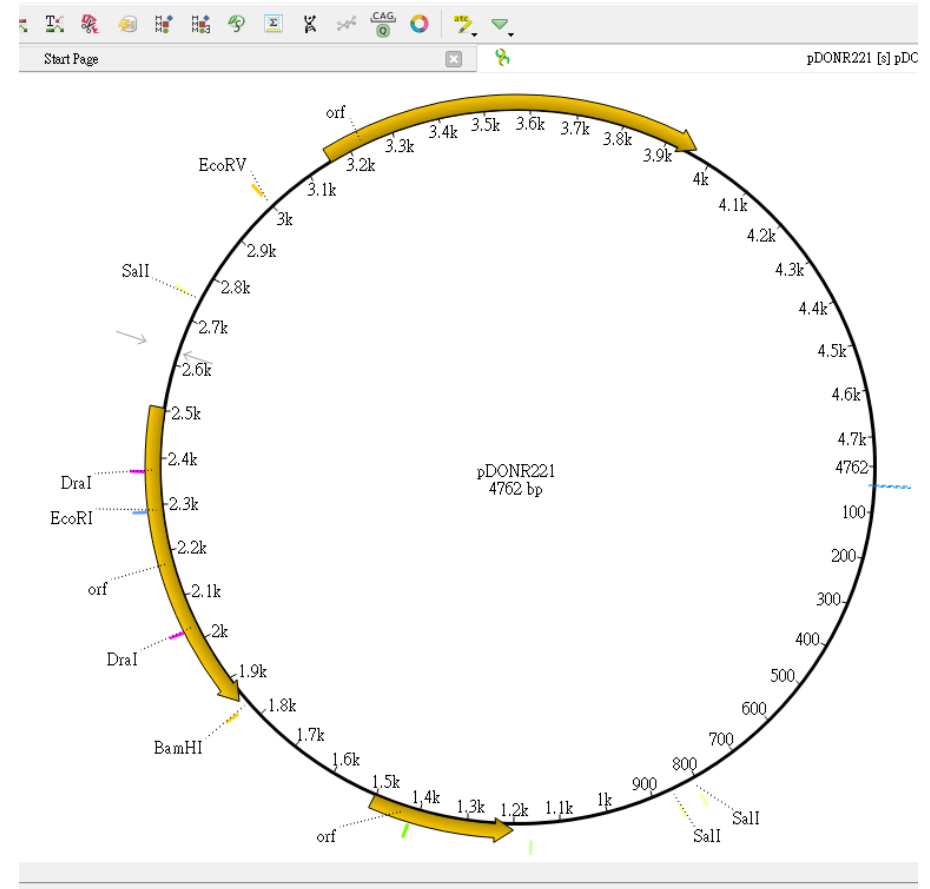
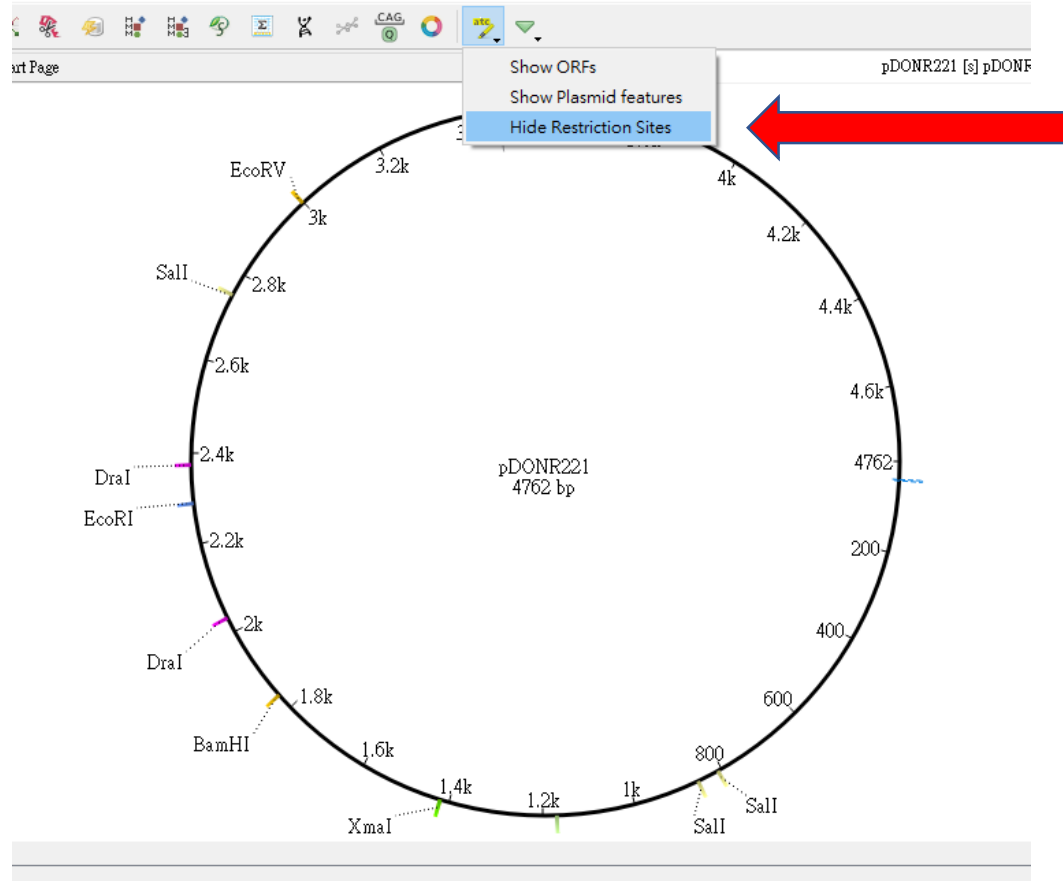
- Genomic Map:** Shows various features along the sequence, including CDS (4), Clal (1), RV (6), HindIII (1), PstI (7), Sall (1), Smal (2), and XmaI (2). The sequence length is 5,833 bp.
- Protein Sequence:** Displays the amino acid sequence: K * K T P P V G G K L A * V T P L C K. A dashed box highlights a segment: N E R P H P * V A S * L K * R H F A R.
- DNA Sequence:** Shows the corresponding nucleotide sequence: A A T G A A A G A C C C C A C C C G T A G G T G G C A A G C T A G C T T A A G T A A C G C C A C T T T G C A A G.
- Annotations Table:** A table listing features and their values. Red arrows point to specific elements in the table and the genomic map.

Name	Value	note
Auto-annotations [murine.gb NC_001363]		
NC_001363 features [murine.gb]		
CDS (0, 4)		
CDS	1042..2658	
CDS	join(2970..3413,3412..3873)	Predicted by GeneMark; artifact
CDS	3875..4999	
CDS	5048..5203	Predicted by GeneMark
misc_feature (0, 2)		
misc_feature	2..590	5' terminal repeat
note	5' terminal repeat	
misc_feature	5245..5833	3' terminal repeat
source (0, 1)		

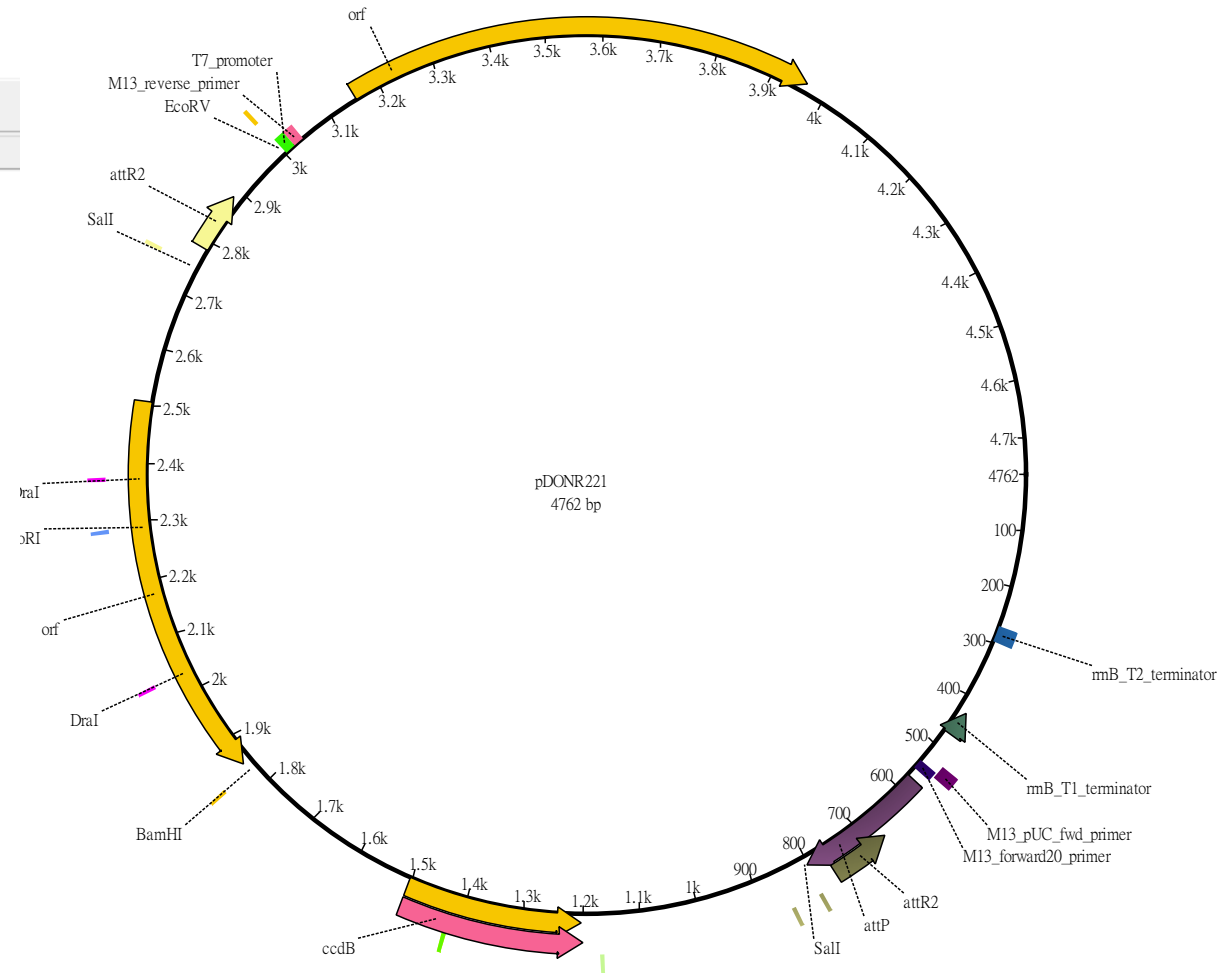
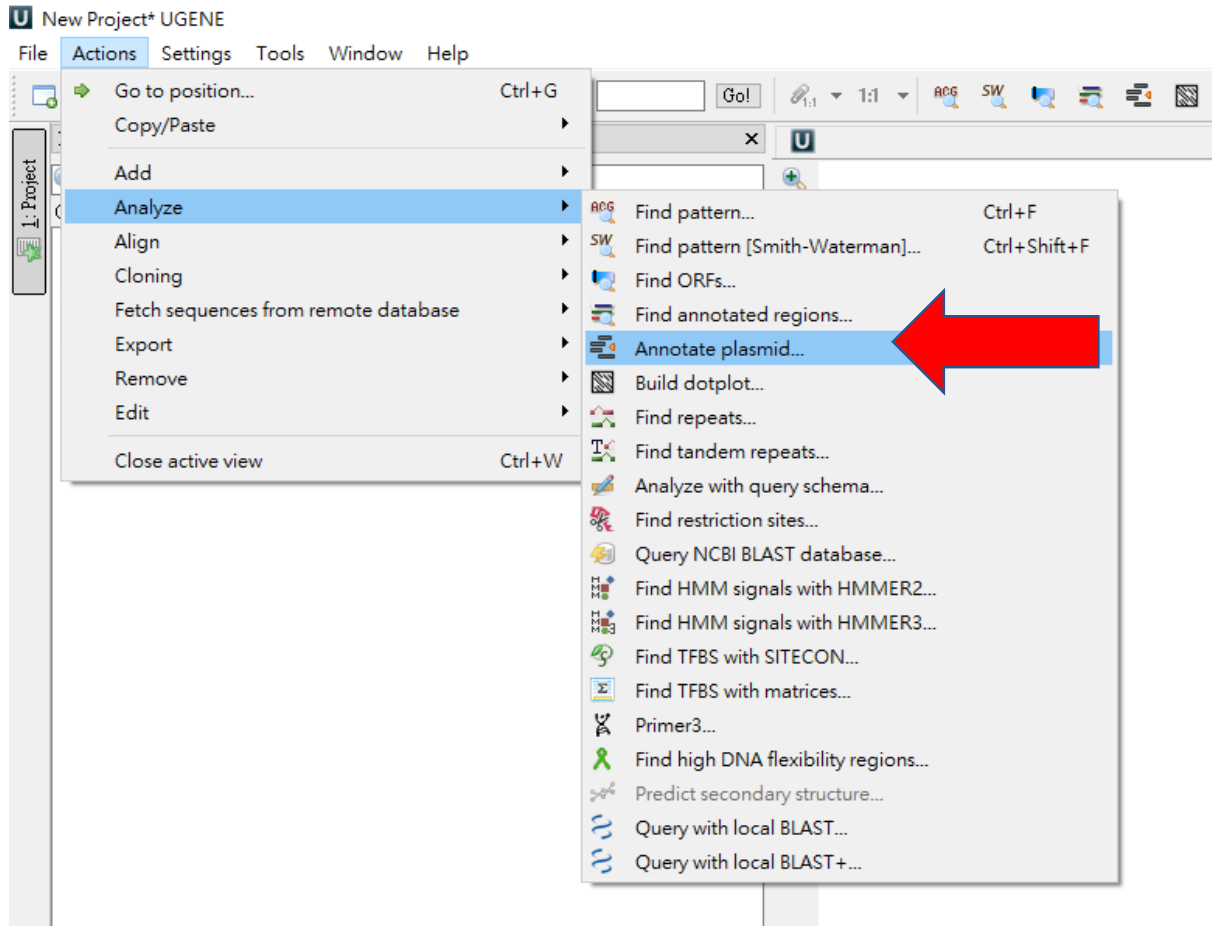
Annotations Editor Annotations

- Objects with annotations:** Points to the CDS (0, 4) group in the tree view.
- Groups:** Points to the CDS (0, 4) group in the tree view.
- Annotation:** Points to a specific CDS entry in the tree view.
- Qualifier's name and value:** Points to the 'note' column in the table, specifically to the value '5' terminal repeat'.
- Column with 'note' qualifier values:** Points to the 'note' column in the table.

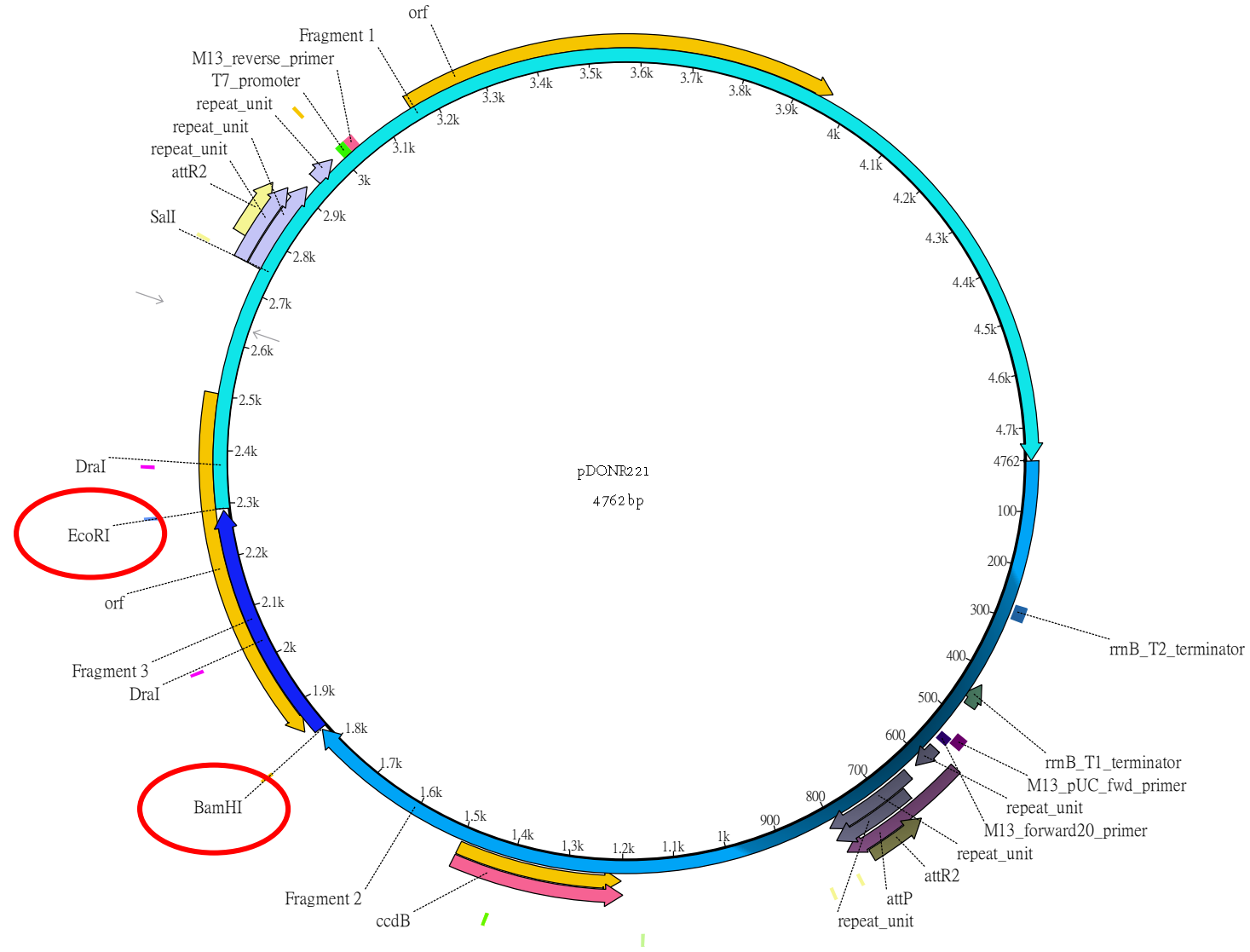
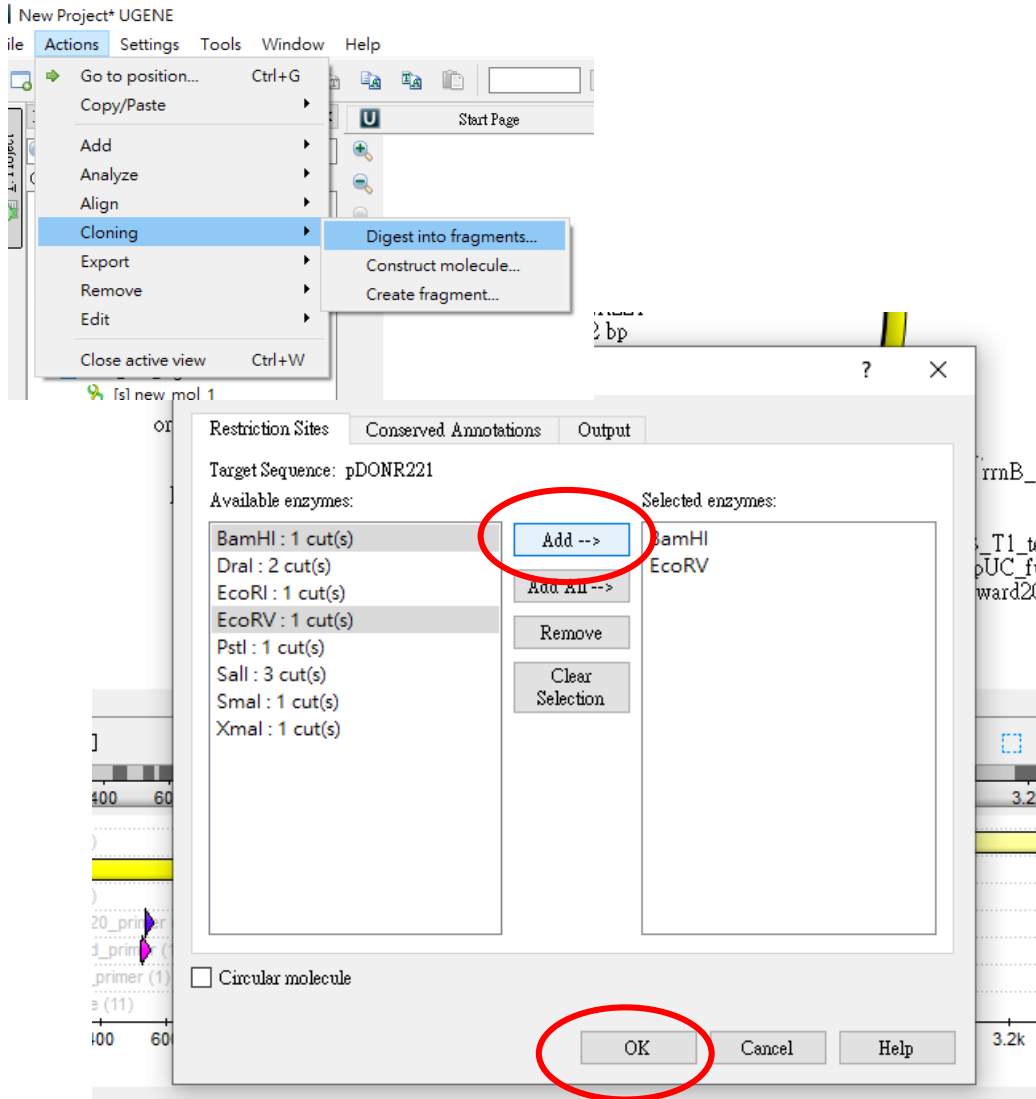
Plasmid Annotation



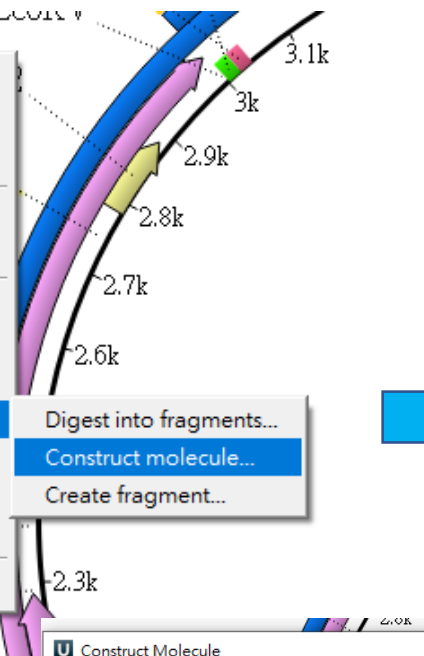
Plasmid Annotation



Cloning



- Go to position... Ctrl+G
- Select sequence region... Ctrl+A
- New annotation... Ctrl+N
- Copy/Paste
- Select
- Add
- Analyze
- Align
- Cloning**
 - Digest into fragments...
 - Construct molecule...**
 - Create fragment...
- Export
- Edit
- Remove
- Rulers...



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]

New molecule contents:

5'	Fragment	3'	Inverted
GATC (Fwd)	pDONR221 (pDONR221.fasta) Fragment 2 [core len...	AATT (Rev)	<input type="checkbox"/> no
AATT (Fwd)	pDONR221 (pDONR221.fasta) Fragment 1 [core len...	GATC (Rev)	<input type="checkbox"/> no
GATC (Fwd)	KX904528 (homologous.gb) Fragment 1 [core lengt...	GATC (Rev)	<input type="checkbox"/> no

Up Down Remove Edit Clear All

Annotate fragments in new molecule
 Force "blunt" and omit all overhangs
 Make circular

OK Cancel Help

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]

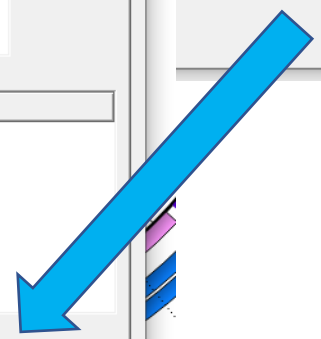
New molecule contents:

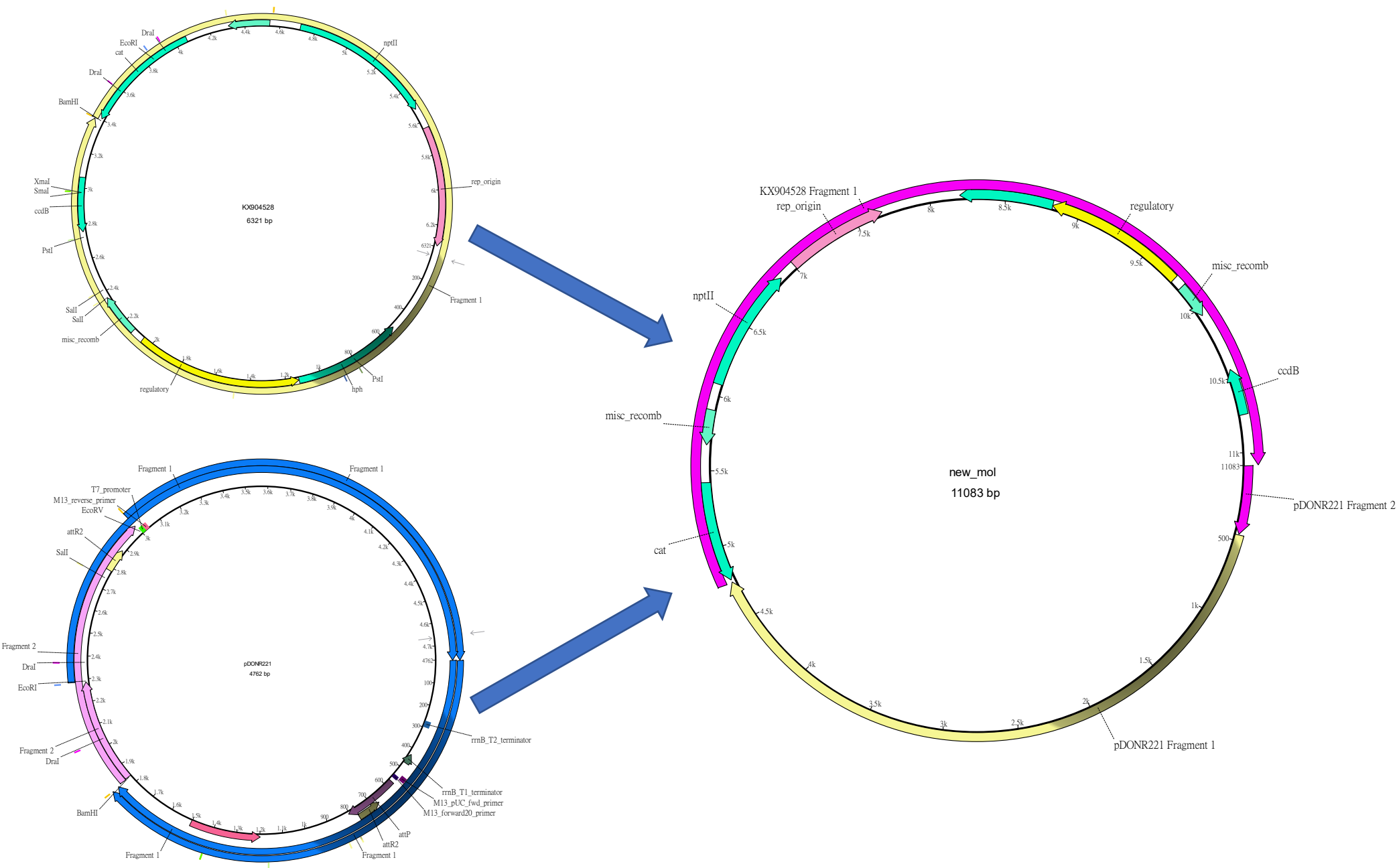
5'	Fragment	3'	Inverted
AATT (Fwd)	pDONR221 (pDONR221.fasta) Fragment 1 [core len...	GATC (Rev)	<input type="checkbox"/> no
GATC (Fwd)	pDONR221 (pDONR221.fasta) Fragment 2 [core len...	AATT (Rev)	<input type="checkbox"/> no
Blunt	pDONR221 (pDONR221.fasta) Fragment 1 [core len...	GATC (Rev)	<input type="checkbox"/> no
GATC (Fwd)	pDONR221 (pDONR221.fasta) Fragment 2 [core len...	Blunt	<input type="checkbox"/> no
GATC (Fwd)	KX904528 (homologous.gb) Fragment 1 [core lengt...	GATC (Rev)	<input type="checkbox"/> no
Blunt	KX904529 (homologous.gb) Fragment 1 [core lengt...	Blunt	<input type="checkbox"/> no

Up Down Remove Edit Clear All

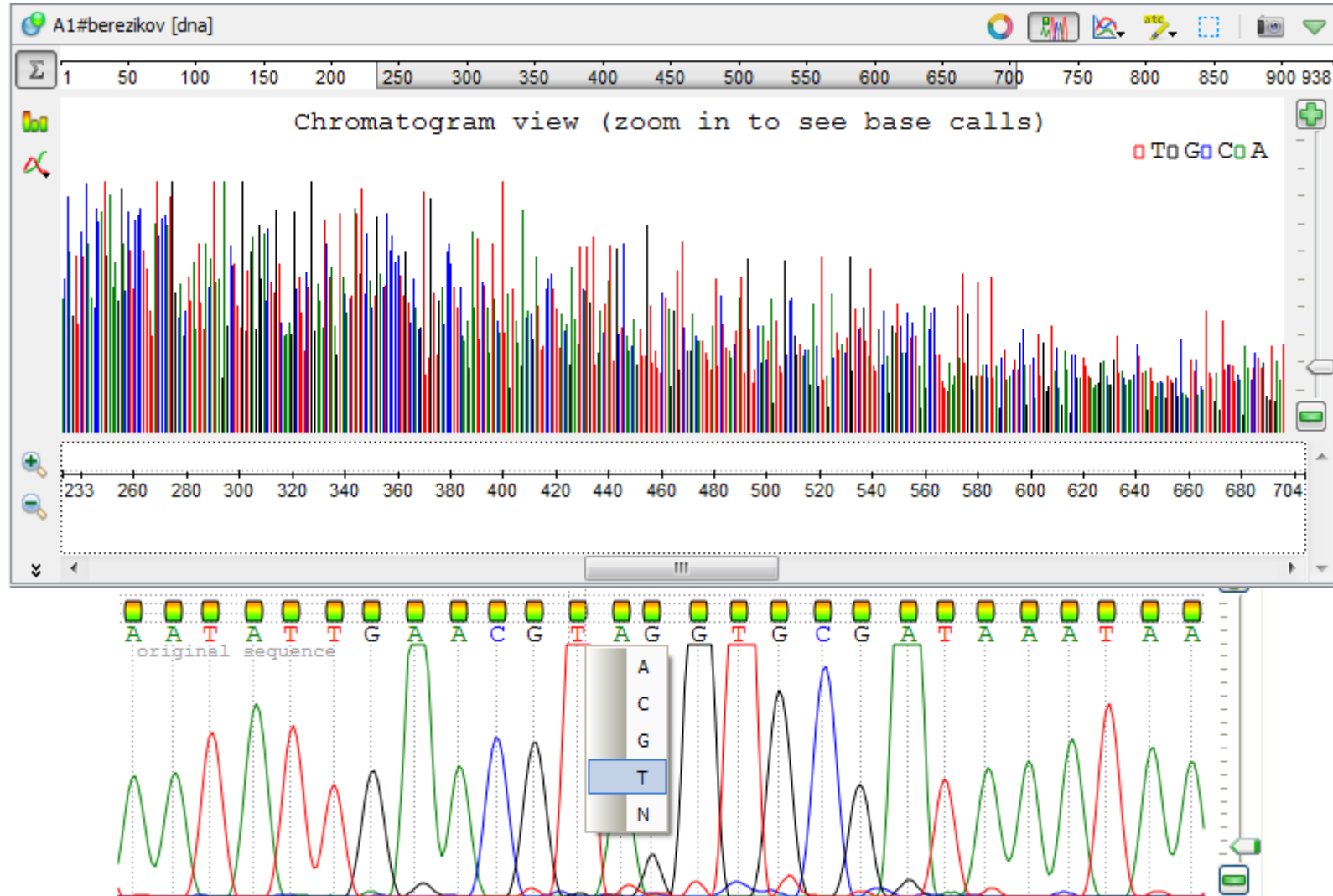
Annotate fragments in new molecule
 Force "blunt" and omit all overhangs
 Make circular

OK Cancel Help





Chromatogram Viewer



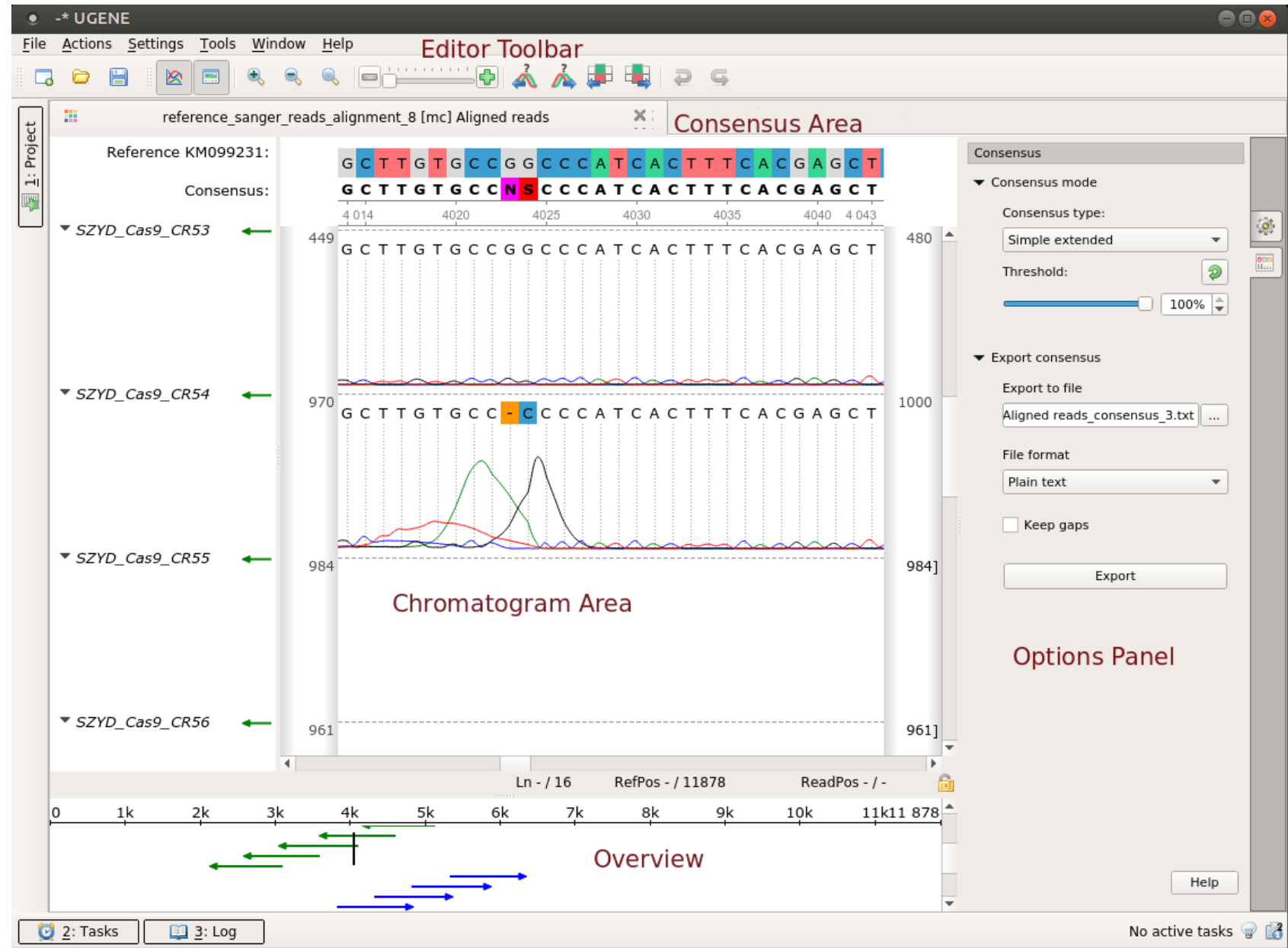
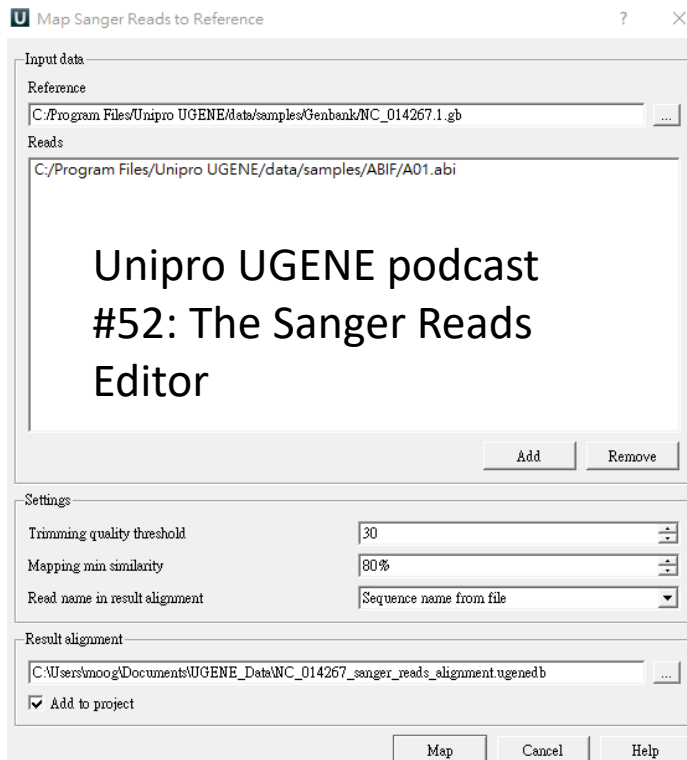
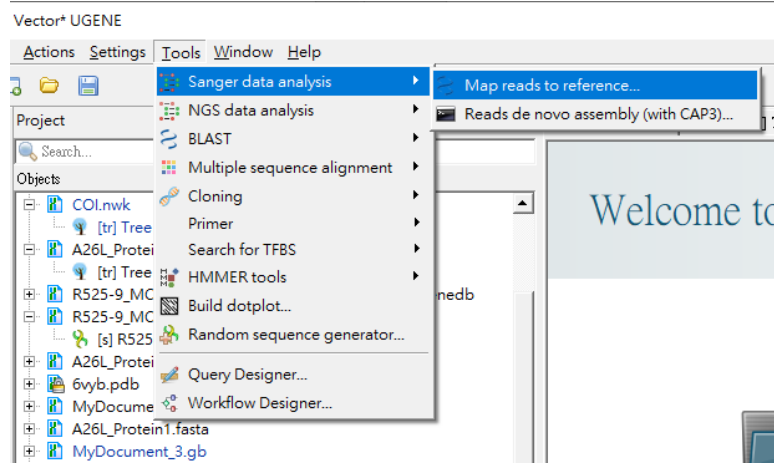
Contigs assembly with CAP3

- 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



Local and BLAST search

BLAST Protocols

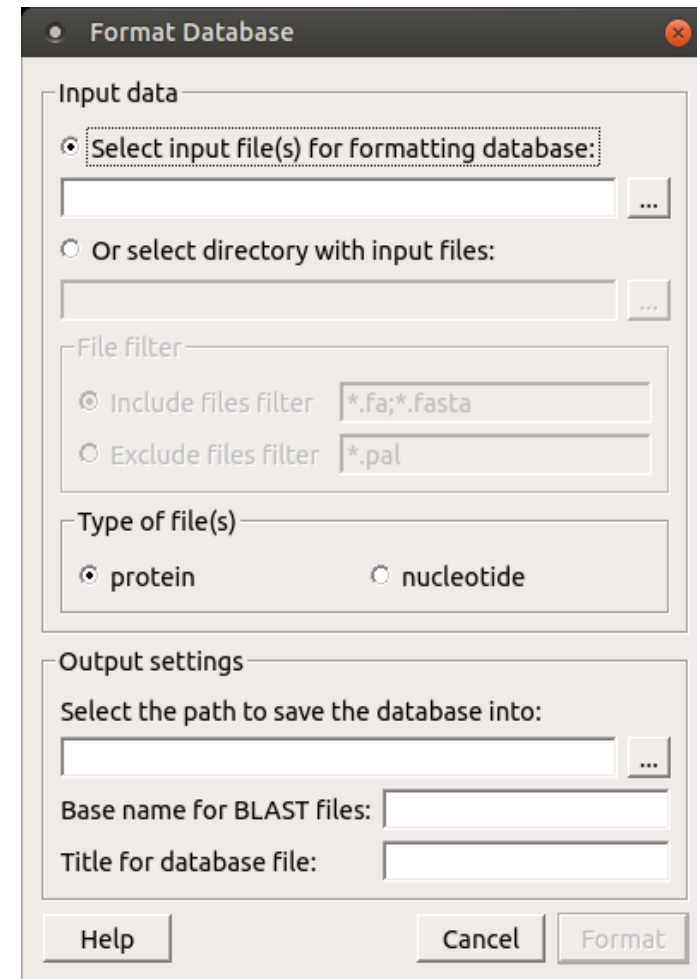
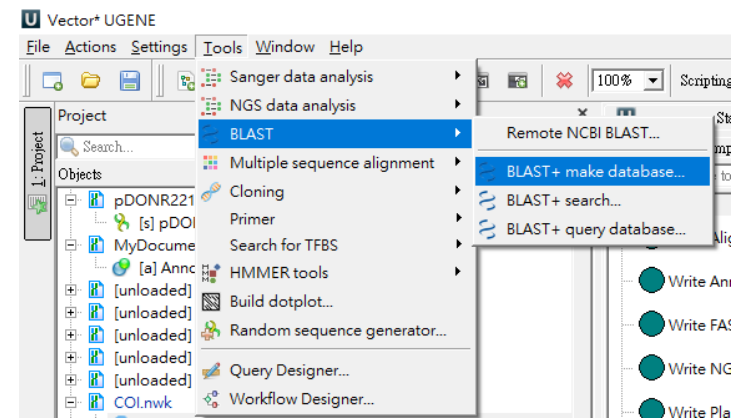
- ◆ The most common BLAST search includes **five protocols:**

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

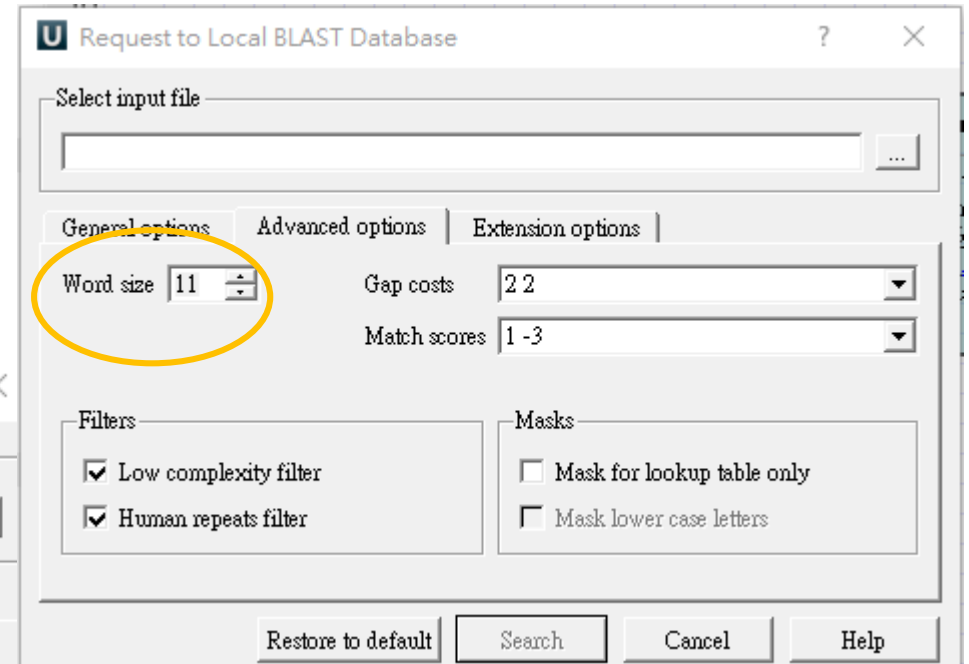
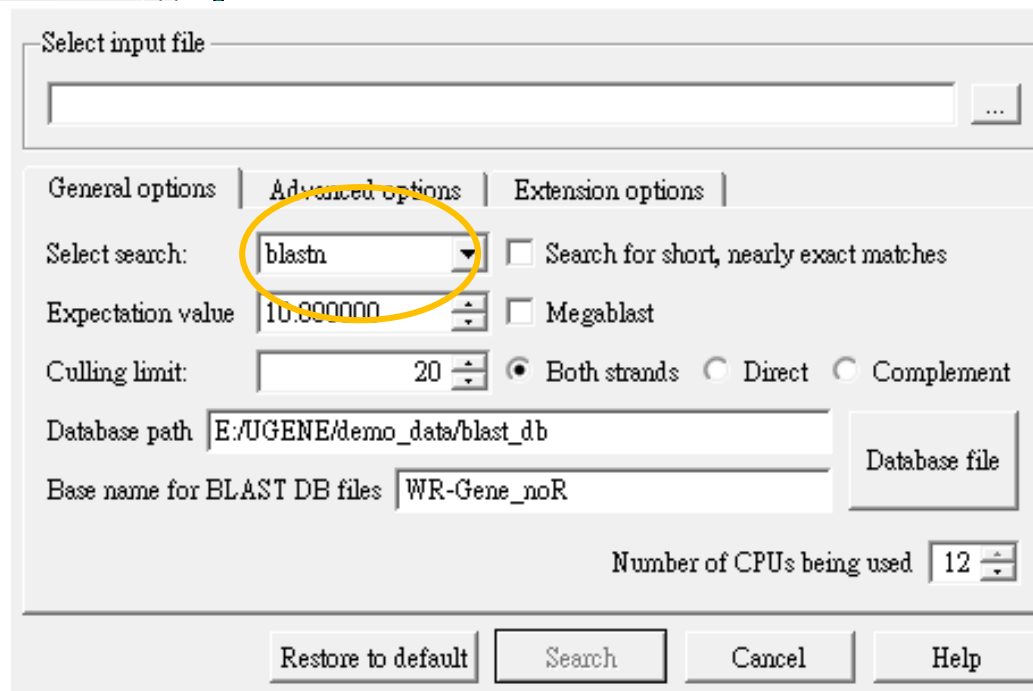
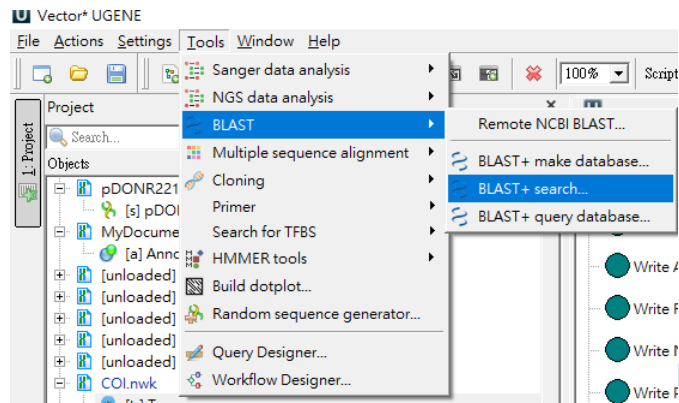


Creating Database

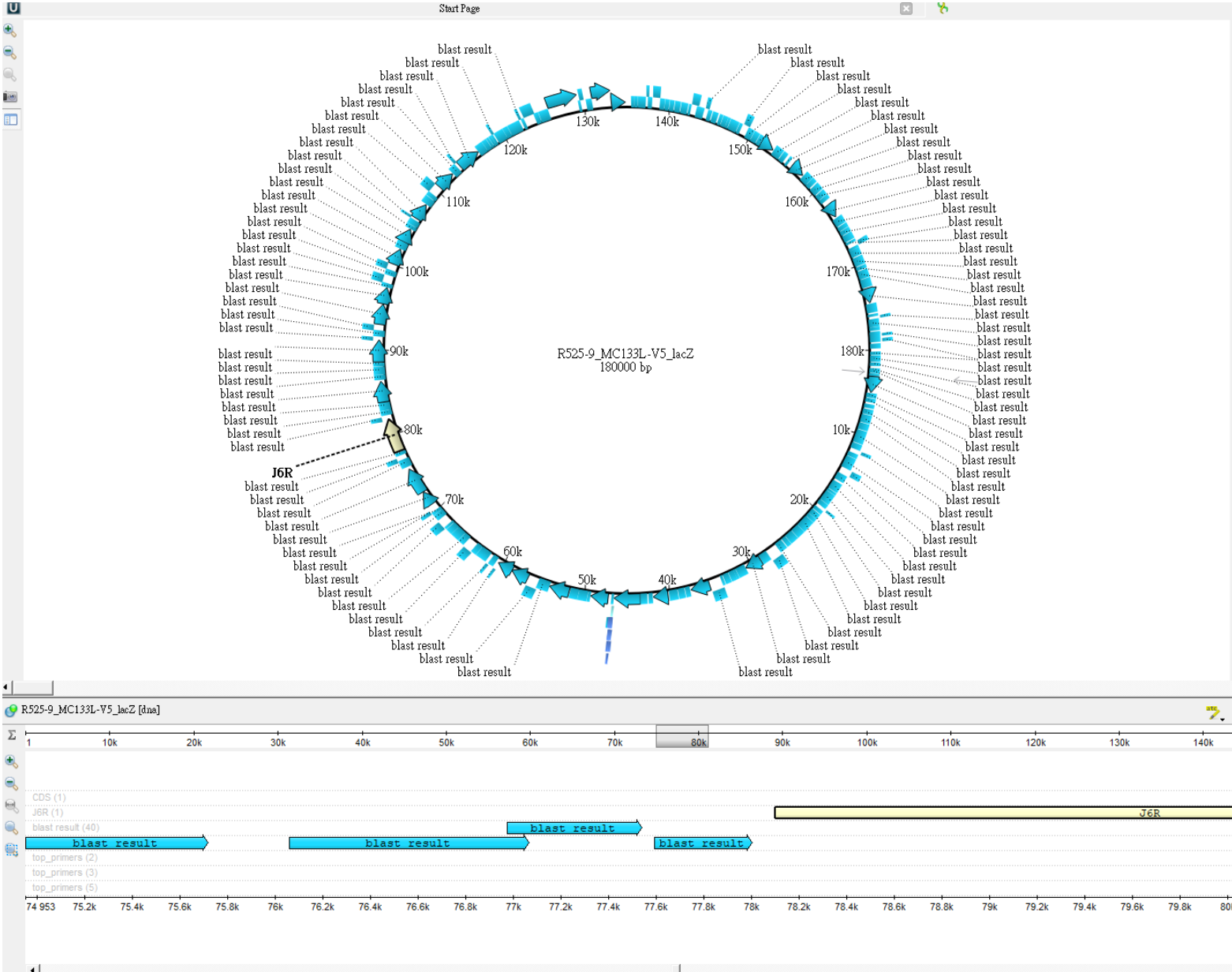
- using *BLAST+* open *Tools* ▶ *BLAST* ▶ *BLAST+ make DB*.
- You can choose either *protein* or *nucleotide* type of the files.
- Then you must select the path to save the database file and specify a *Base name for BLAST files* and a *Title for database file*.



BLAST search



BLAST search result



Making Request to Public Databases Database

The screenshot illustrates the steps to connect to a public database in UGENE. The main window shows the 'Welcome to UGENE' screen with a menu bar (File, Actions, Settings, Tools, Window, Help) and several icons: 'Open File(s)', 'Create Sequence', and 'Create Workflow'. A 'Shared Databases Connections' dialog box is open, displaying a list with 'UGENE public database' selected. A 'File' menu is open, with 'Connect to shared database' (Ctrl+L) highlighted. Green arrows and numbered circles (1, 2, 3) indicate the sequence of steps: 1. Clicking 'Connect to shared database' in the File menu; 2. Selecting 'UGENE public database' in the dialog; 3. Clicking the 'Connect' button.

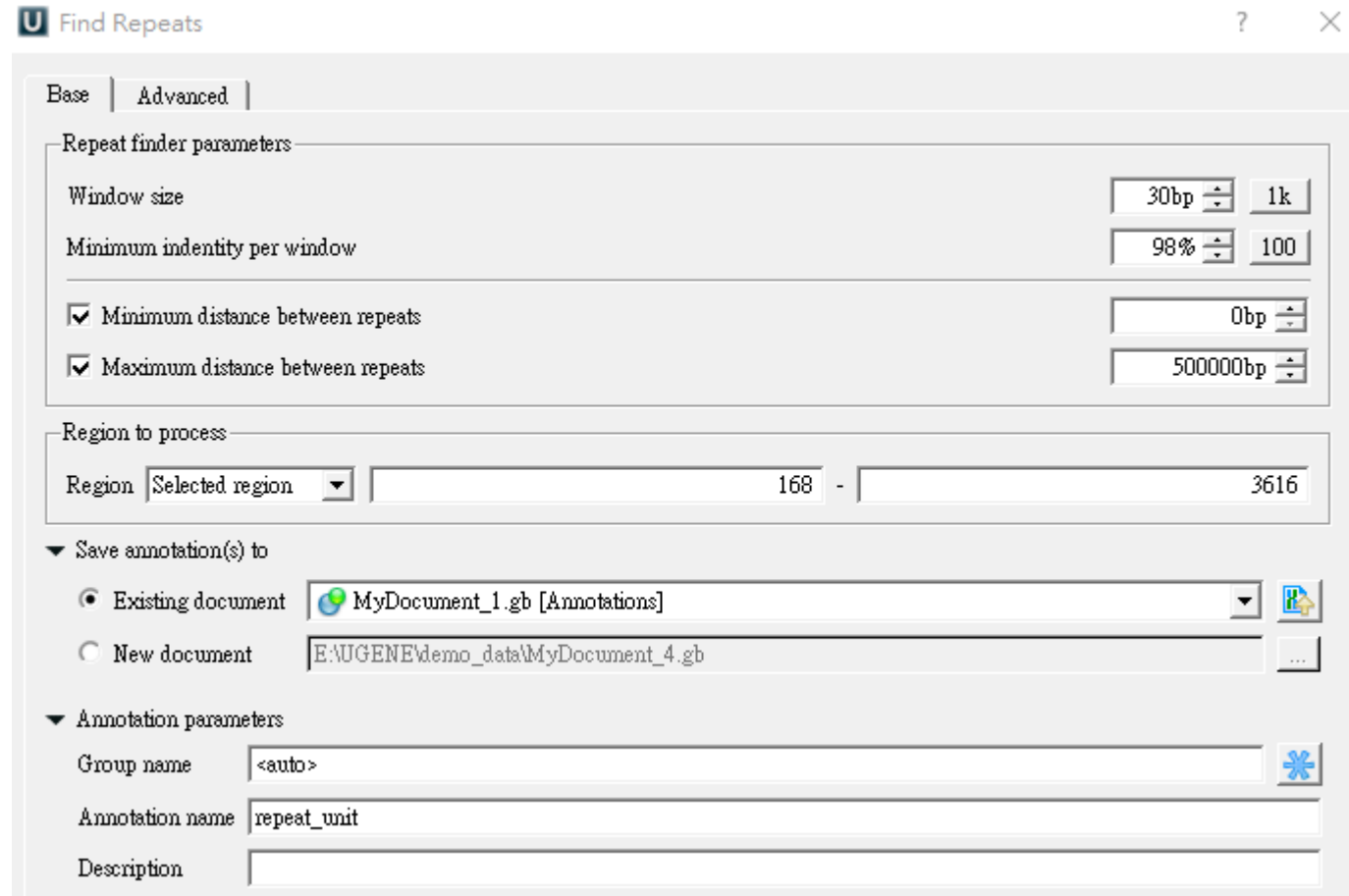
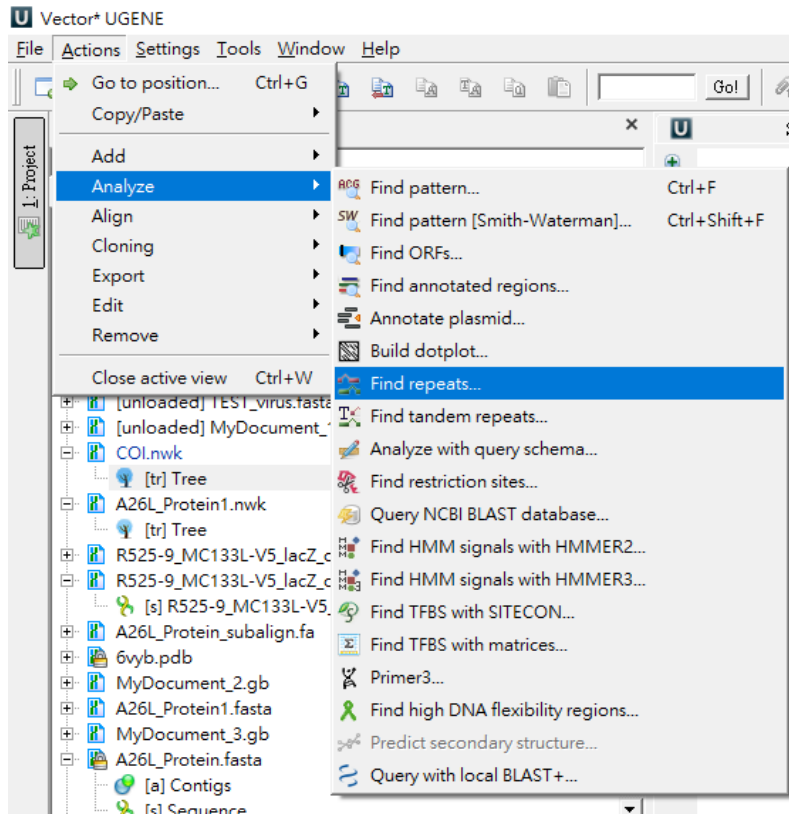
Cite UGENE:
"Unipro UGENE: a unified bioinformatics toolkit"
Okonechnikov, Golosova, Fursov, the UGENE team
Bioinformatics 2012 28: 1166-1167

Follow UGENE:
[Social media icons: Facebook, Twitter, LinkedIn, Google+, YouTube, Blogger, RSS]

Making Request to Public Databases Database

The screenshot displays the UGENE software interface. On the left, the 'Project' pane shows a tree view of the 'UGENE public database' with folders for 'genomes' and 'plasmids'. A callout box points to the 'plasmids' folder, stating: "This folder contains plasmids from the NCBI UniVec database". Another callout box points to the search bar in the 'Project' pane, stating: "Type an organism name of a plasmid property to find corresponding sequences in the database". A third callout box points to the 'UGENE public database' folder, stating: "The database's shown up in the project as a regular document". Below the 'Project' pane, there are buttons for "Open File(s)" and "Create Workflow". At the bottom, a "Cite UGENE:" section provides citation information: "Unipro UGENE: a unified bioinformatics tool Okonechnikov, Golosova, Fursov; the UGENE Bioinformatics 2012 28: 1166-1167". On the right, the 'Objects' pane shows a list of retrieved sequences, including: "AY597270_Gentamycin resistance FRT vector pFGM1, complete sequence.", "AY597271_Chloramphenicol resistance FRT vector pFCM1, complete sequence.", "AY597272_Kanamycin resistance FRT vector pFKM1, complete sequence.", "AY597273_Flp expression vector pFLP3, complete sequence.", "AY599226_Cloning vector pUC18-mini-Tn7, complete sequence.", "AY599227_Cloning vector pUC18-mini-Tn7T, complete sequence.", "AY599228_Cloning vector pUC18-R6K-mini-Tn7T, complete sequence.", "AY599229_Cloning vector pUC18Pv-mini-Tn7T, complete sequence.", "AY599230_Cloning vector pUC18T-mini-Tn7T, complete sequence.", "AY599231_Cloning vector pUC18-mini-Tn7T-Gm, complete sequence.", "AY599232_Cloning vector pUC18T-mini-Tn7T-Gm, complete sequence.", "AY599233_Cloning vector pUC18-mini-Tn7T-Gm-lacZ, complete sequence.", "AY619004_Cloning vector pUC18-mini-Tn7-Gm, complete sequence.", "AY619005_Cloning vector pUC18-mini-Tn7-LACM15, complete sequence.", "[s] AY619005_Cloning vector pUC18-mini-Tn7-LACM15, complete sequence", "[a] AY619005_Cloning vector pUC18-mini-Tn7-LACM15, complete sequence features", "AY619006_Cloning vector pUC18R6K-mini-Tn7T-Km, complete sequence.", "AY643800_Large-insert cloning vector pSMART VC, complete sequence.", "AY712950_Tetracycline resistance FRT vector pFTC1, complete sequence.", "AY712951_Trimethoprim resistance FRT vector pFTP1, complete sequence.", "AY712952_Cloning vector pUC18T-mini-Tn7T-Gm-REP, complete sequence.", "AY712953_Mini-Tn7 delivery vector pUC18R6KT-mini-Tn7T, complete sequence.", "AY737004_Cloning vector pUC18-mini-Tn7T-Gm-Gateway, complete sequence.", "AY737005_Cloning vector pUC18-mini-Tn7T-Tp, complete sequence."

Search for Sequence Repeats



Search for Sequence Repeats

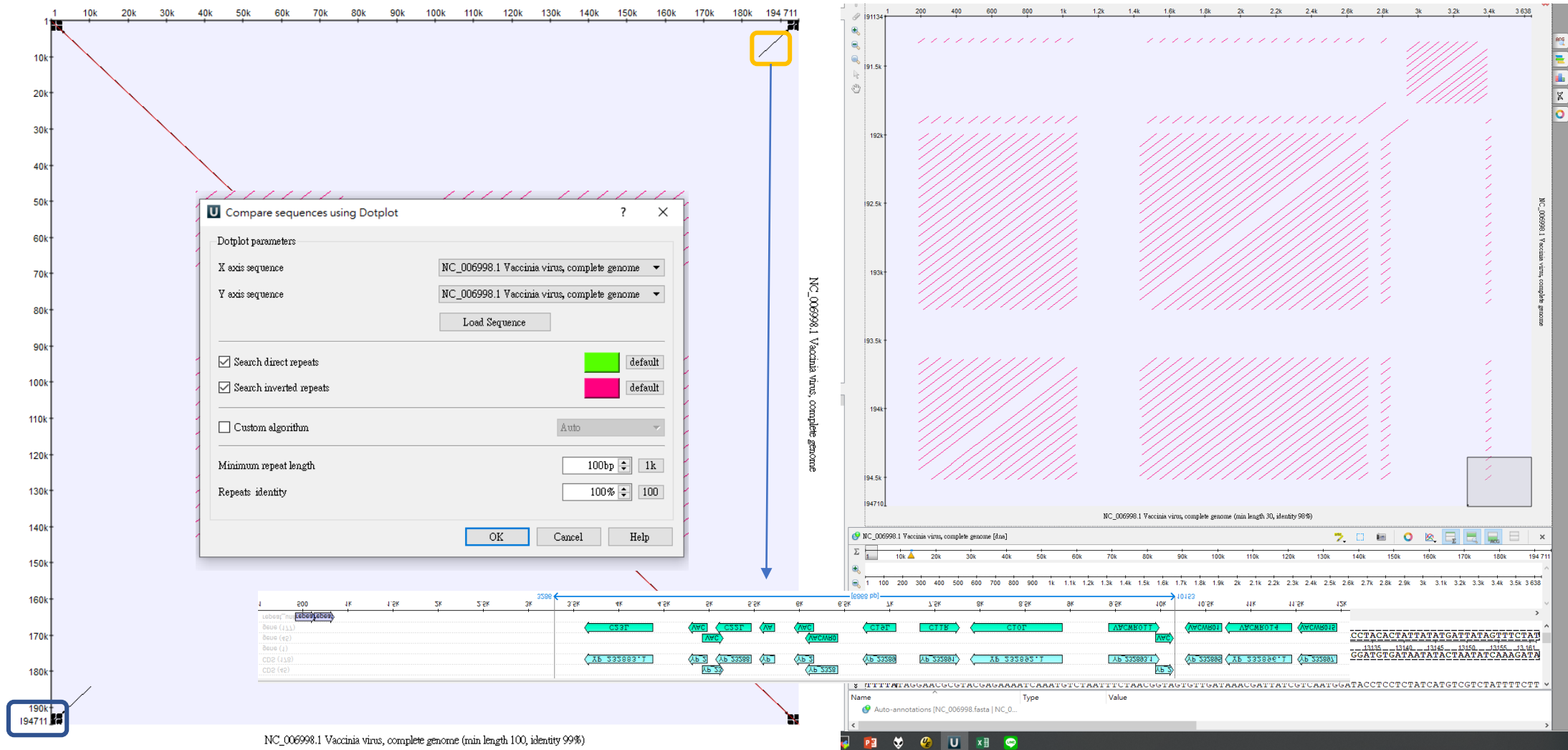
The screenshot displays the UGENE software interface for sequence analysis. The main window shows a genome map of the Vaccinia virus complete genome (NC_006998) with purple bars indicating identified repeats. The bottom panel provides a detailed view of a repeat unit, showing sequence alignment and a table of repeat regions.

Name	Type	Value
repeat_unit	Repeat Region	join(423..631,632..840)
repeat_unit	Repeat Region	join(734..833,833..841)

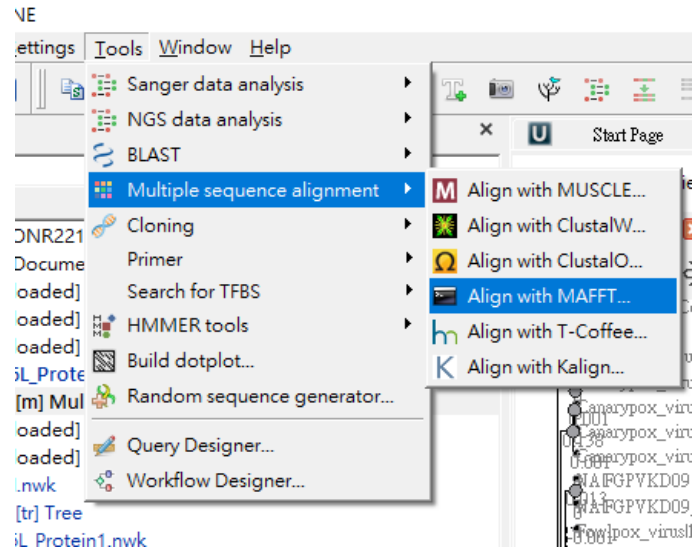
Tasks:

Task name	Task state description	Task progress	Actions

Search for Sequence Repeats using Dotplot



Multiple sequence alignment

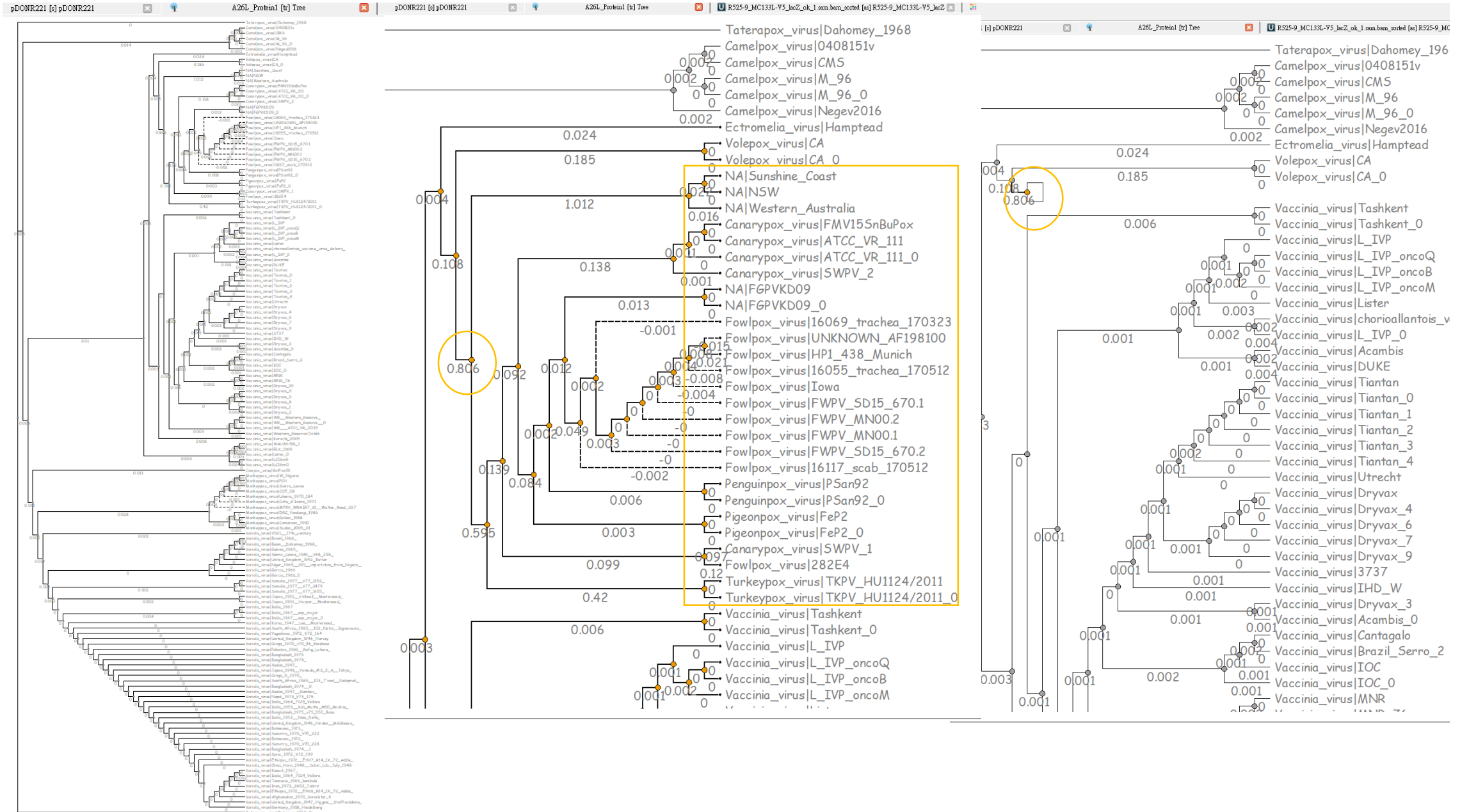


- Clustal is a widely used multiple sequence alignment program. It is used for both nucleotide and protein sequences.
- Kalign is a fast and accurate multiple sequence package designed to align large numbers of protein sequences.
- MAFFT is used for both nucleotide and protein sequences.

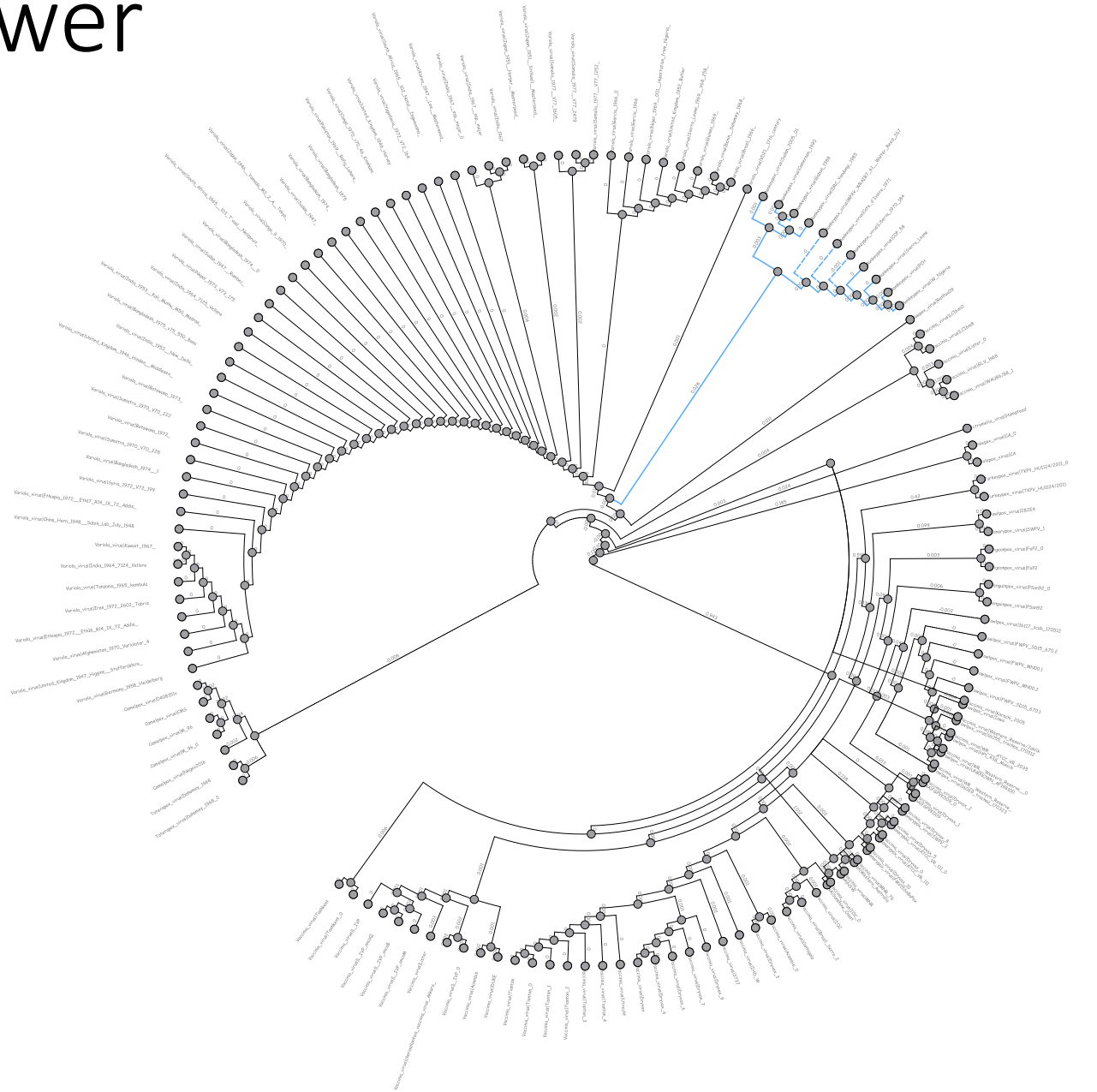
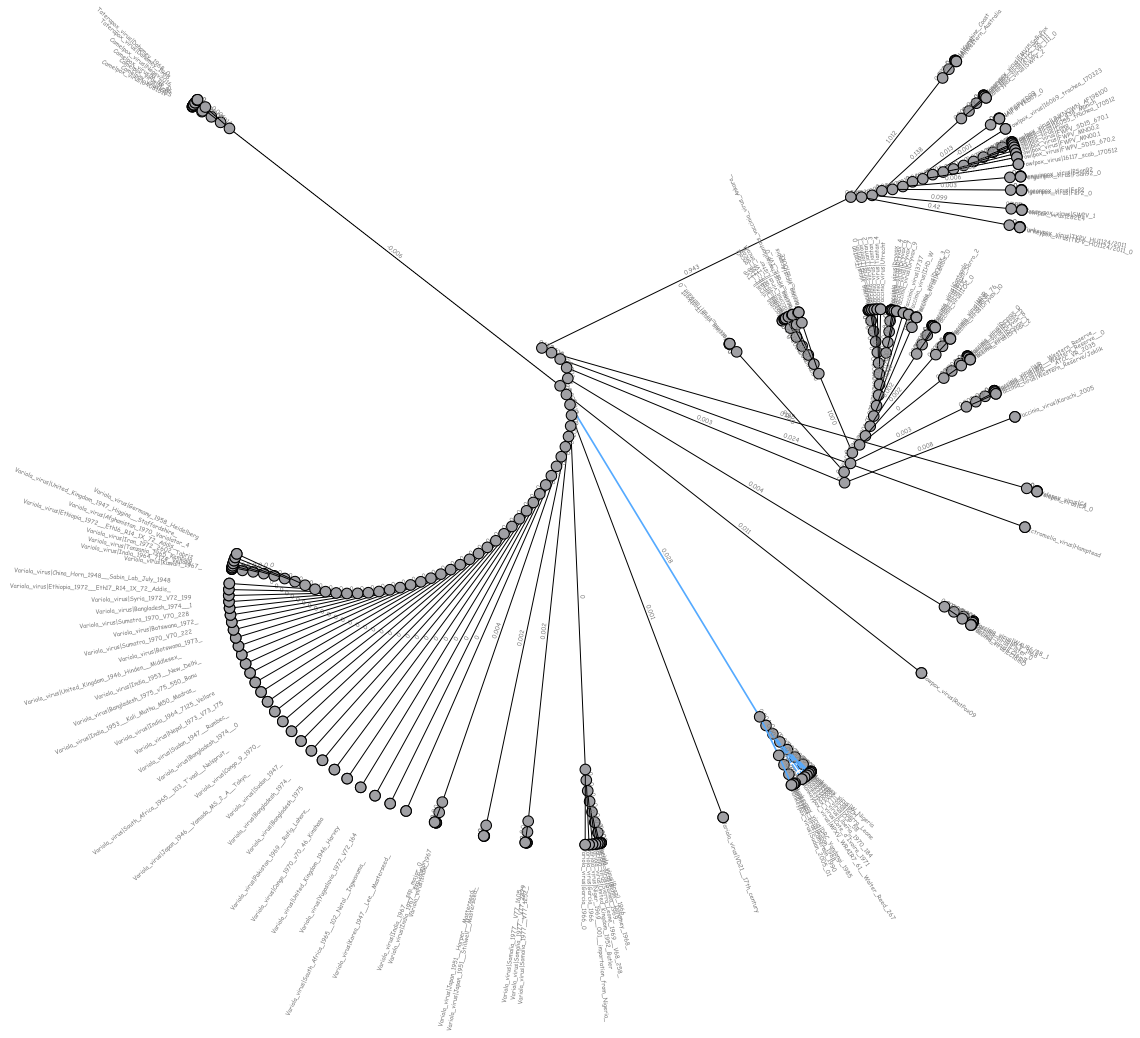
Alignment Editor

The screenshot displays the Vector UGENE software interface for editing a multiple sequence alignment. The main window is titled "A26L_Protein [m] Multiple alignment" and shows a consensus sequence at the top: **MANIINLWNGageeeddvIVPTVQDVNVASITARdFKSMIDETWDKKIEANTCISRKHR**. Below this, a grid of amino acid sequences is shown, with each row representing a different virus strain. The sequences are aligned, with gaps indicated by dashes. A vertical bar chart at the top of the alignment grid shows the frequency of each amino acid across the sequences. The left sidebar contains a "Tree view" of the alignment, listing various poxvirus strains such as *Taterapox virus*, *Camelox virus*, *Canarypox virus*, *Fowlpox virus*, *Penquinpox virus*, and *Vaccinia virus*. The right sidebar contains a "General" panel with "Reference sequence" (selected as "Select and add"), "Alignment info" (Alphabet: Standard amino acid, Length: 613, Sequences: 151), "Consensus mode" (Consensus type: Default, Threshold: 62%), "Copy to clipboard" (Format: CLUSTALW, Copy button), and "Sort sequences" (Sort by: Name, Sort order: Ascending, Sort button). The bottom status bar shows "Seq 1 / 151", "Col 35 / 613", "Pos gap / 510", and "Sel 1x151".

Phylogenetic Tree Viewer



Phylogenetic Tree Viewer



3D Structure Viewer

Vector* UGENE

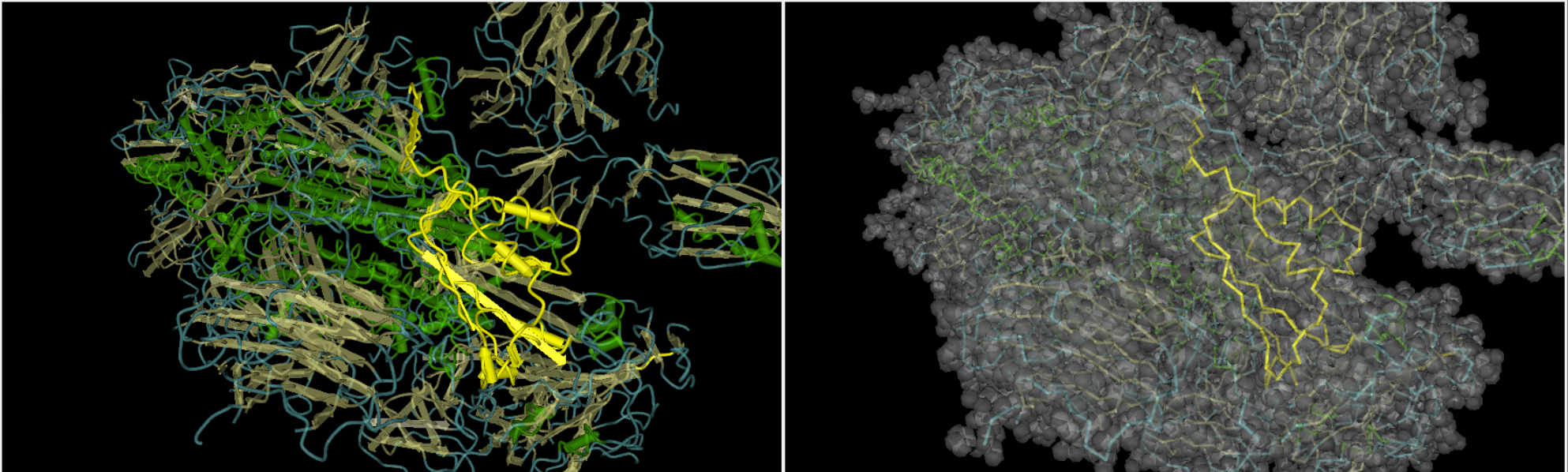
File Actions Settings Tools Window Help

Project: U Start Page COI [tr] Tree pDONR221 [s] pDONR221 A26L_Protein1 [tr] Tree A26L_Protein [m] Multiple alignment A26L_Protein_subalign [m] Multiple alignment Workflow Designer - Search for inverted repeats 6vyb.pdb

3D Structure Viewer

Objects

Active view: 1: XXXX Display Links



Bookmarks

- CC
- pC
- A2
- A2
- A2
- A2
- 6V

XXXX chain C sequence [amino]

beta_str alpha_helix beta_strand beta_str alpha_helix alpha_helix alpha_helix alpha_helix beta_strand alpha_helix alpha_helix beta_str alpha_helix beta_strand

259 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 498

NRKRISNCVADYSVLYNSASFSTFKCYGVSPSTKLNLDLCEINVVYADSFVIRGDEVROIA PQQTGKIA DYNKLPDDFTGCVIAWNSNLDKSNVNYLYRKPFFERDITSTPLQSYCFQPTNGVGYQPYRVVVLSELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPQQQFGRDIADTTDAVRDPQTLLEILDITPCSFGGVSVITPGINTSNEVAVLYQDVNCTEVNVF

XXXX chain A sequence [amino]

sec_struct (8)

sec_struct beta_str

422 424 426 428 430 432 434 436 438 440 442 444 446 448 450 452 454 456 458 460 462 464 466 468 470 472 474 476 478 480 482 484 486 488 490 492 494 496 498 500 502 504 506 508 510 512 514 516 518 520 522 524 526 528 530 532 534 536 538 540 542

NFNGLTGTGVLTESNKKFLPQQQFGRDIADTTDAVRDPQTLLEILDITPCSFGGVSVITPGINTSNEVAVLYQDVNCTEVNVFQTRAGCLIGAHEVNNSYECDIPIGAGICASYQTQSI IAY

GT+ CAT

A YTN SFTRGVYYPDKVFRSS VLIHSTODLFLPFSS NVTWTE HAIHPVLPFNDGVYFAS DEK SNII RGMIFGT TLDSKSL LLI VNNATNVVIVKVEFQFCND EFLGVCTFEYVSFKNLRFEVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIG

Name	Type	Value
sec_struct	Misc. Feature	368..377

2 Tasks Log

No active tasks

Tutorials

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

The screenshot shows the UniproUGENE website's navigation menu. The 'Learn' tab is highlighted with a green box. A dropdown menu is open under 'Documentation', with 'YouTube Podcast' highlighted by a yellow box. Other items in the dropdown include 'UGENE Key Features', 'Quick Start Guide', 'UGENE Manual', 'Version History', and 'FAQ'. The 'Home' tab is also visible.

The screenshot shows the YouTube channel page for UniproUGENE, which has 1210 subscribers. The page features a grid of video thumbnails. The first row includes a video titled 'Unipro UGENE podcast #56: What's new in UGENE 33.0 ...' (4:33) and several other podcasts. The second row shows 'Unipro UGENE podcast #50: What's new in UGENE 1.20 ...' (2:13) and 'Unipro UGENE podcast #49: What's new in UGENE 1.18 ...' (3:54). The third row contains 'Unipro UGENE podcast #48: What's new in UGENE 1.17' (3:46) and 'Unipro UGENE podcast #47: What's new in UGENE 1.16' (4:06). The page also includes a search bar and navigation tabs for 'Home', 'Videos', 'Playlists', 'Community', 'Channels', and 'About'.