

# QC on Consensus Sequences

Clyde Dapat



WHO Collaborating Centre  
for Reference and  
Research on Influenza  
**VIDRL**



A joint venture between The University of Melbourne and The Royal Melbourne Hospital

# **Objective**

- **To learn how to assess the quality of consensus sequence**

## Consensus assembly

The output of the process of generating **a single representative sequence** from a mapped set of reads to a reference sequence. In other cases, from a set of similar sequences that are also multiple sequence aligned.

## Challenges in consensus assembly

- **Genetic variability:** Viruses often exhibit high mutation rates and genetic diversity.
- **Sequence errors:** Errors introduced during sequencing, such as base calling errors, insertions, deletions, sequence repeats, and chimeric sequences.
- **Mixed infections:** Presence of multiple viral variants within the same sample.
- **Laboratory contamination:** Unintended contamination due to human error in the laboratory.
- **Computational artifacts:** Bioinformatic tools used to analyze data and create a final genome can sometimes introduce errors.

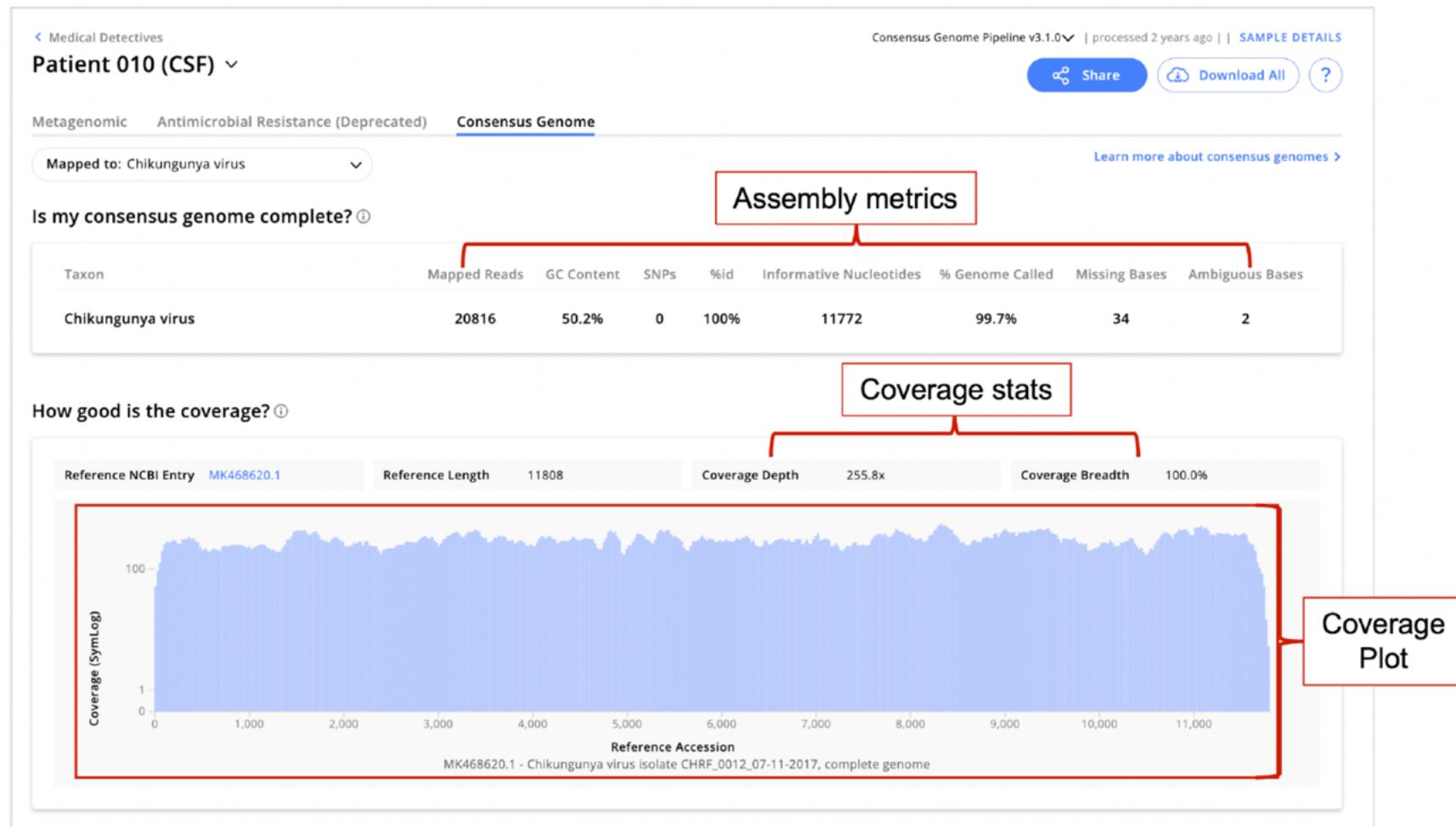
# Methods for quality control

- **Pre-assembly**
  - **Read Filtering:** Exclude low-quality reads and adapter sequences to improve assembly accuracy.
  - **Error Correction:** Utilize tools for error correction, such as k-mer-based correction or consensus-based correction algorithms, e.g. minimum frequency threshold, minimum depth to call consensus
- **Alignment or read mapping**
  - **Alignment-based Filtering:** Align reads and/or consensus sequences to a reference genome to identify and remove erroneous reads or artifacts or gaps
  - **Variant Calling:** Identify and filter out variants that are likely sequencing errors rather than true genetic variation.
  - **Depth of Coverage Analysis:** Assess the depth of coverage across the genome to identify regions with low coverage or potential errors.
- **Cross-Validation:** Validate the consensus sequence with independent sequencing data or experimental methods like PCR and Sanger sequencing.
- **Recombination Detection:** Detect potential product of recombination.

## Assembly metrics

- Coverage plot
- % Genome Called
- SNPs
- Informative bases
- Mapped reads
- GC content

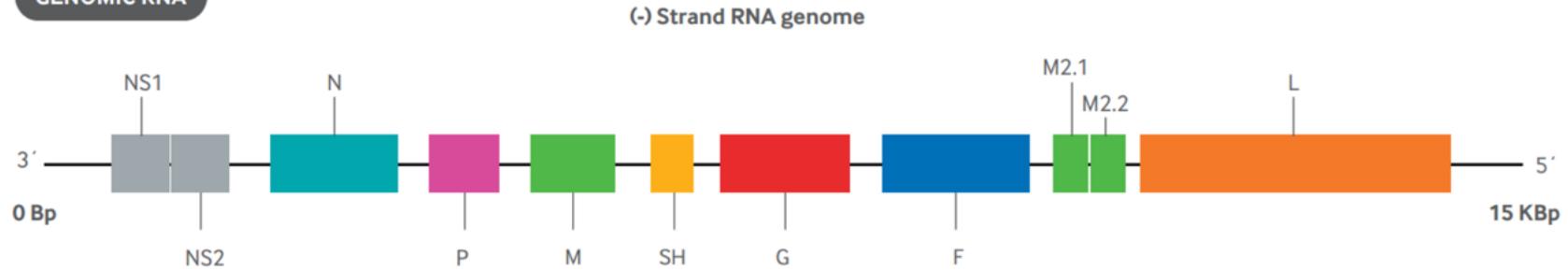
# Assembly metrics



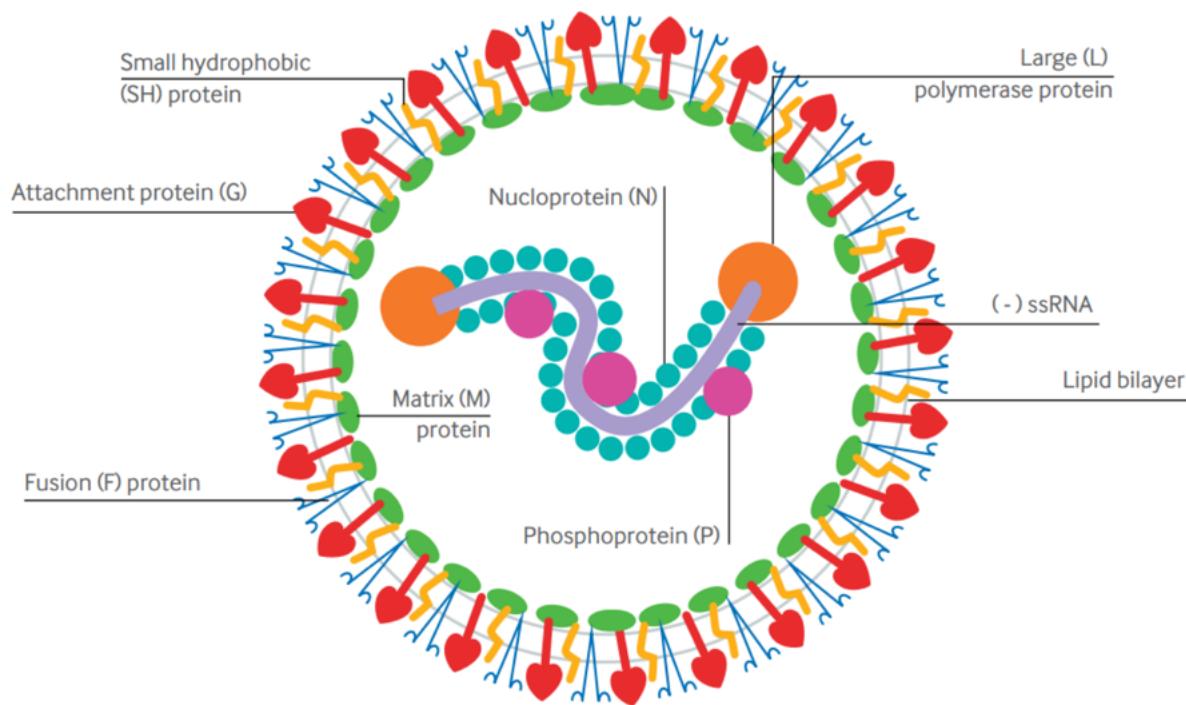
## Organism-specific considerations

- **Genome size:** How many base pairs is it?
- **Genome organization:** How many open reading frames (ORFs)? What is their orientation?
- **Repeat regions:** Are there known repeat regions? What are their positions? If reads don't span this region (i.e., region covered by single reads) the assembly or consensus sequence over these regions should not be trusted.
- **Low or high GC content areas:** Are there genomic regions with low or high GC content? Is important to inspect the assembly over these regions because they are prone to have sequencing bias or errors.

### GENOMIC RNA



### RSV VIRION STRUCTURE



**Non-structural proteins:** NS1, NS2

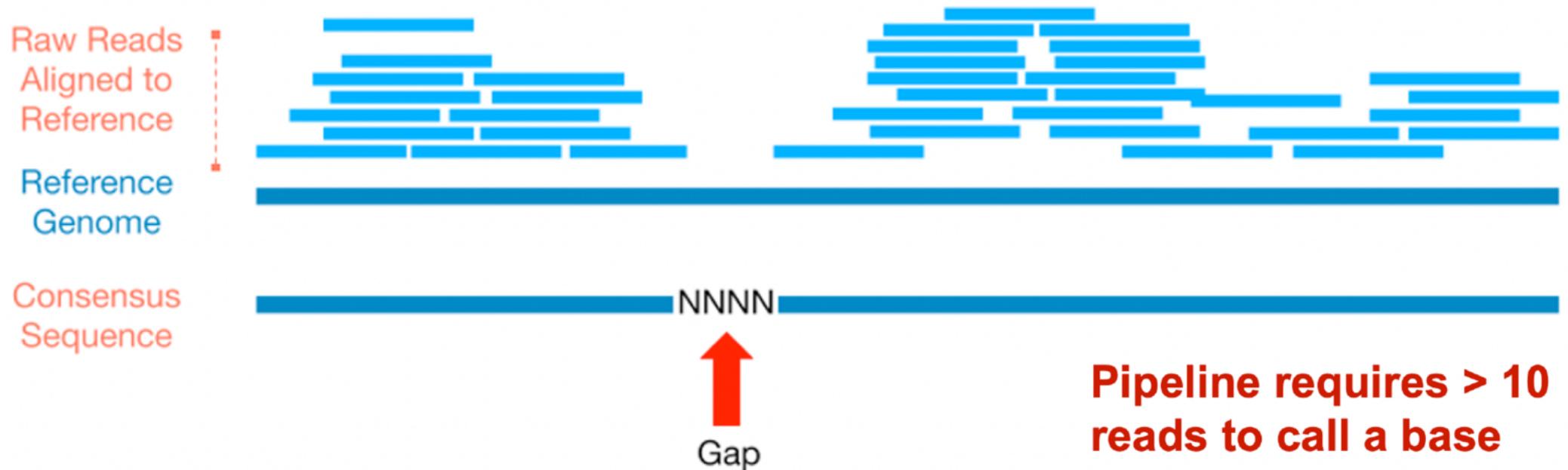
**Nucleocapsid and regulatory proteins:** N (nucleoprotein), P (phosphoprotein), M2.1, M2.2, L (large polymerase)

**Envelope proteins:** SH (small hydrophobic), G (attachment), F (fusion)

**Inner envelope protein:** M (matrix)

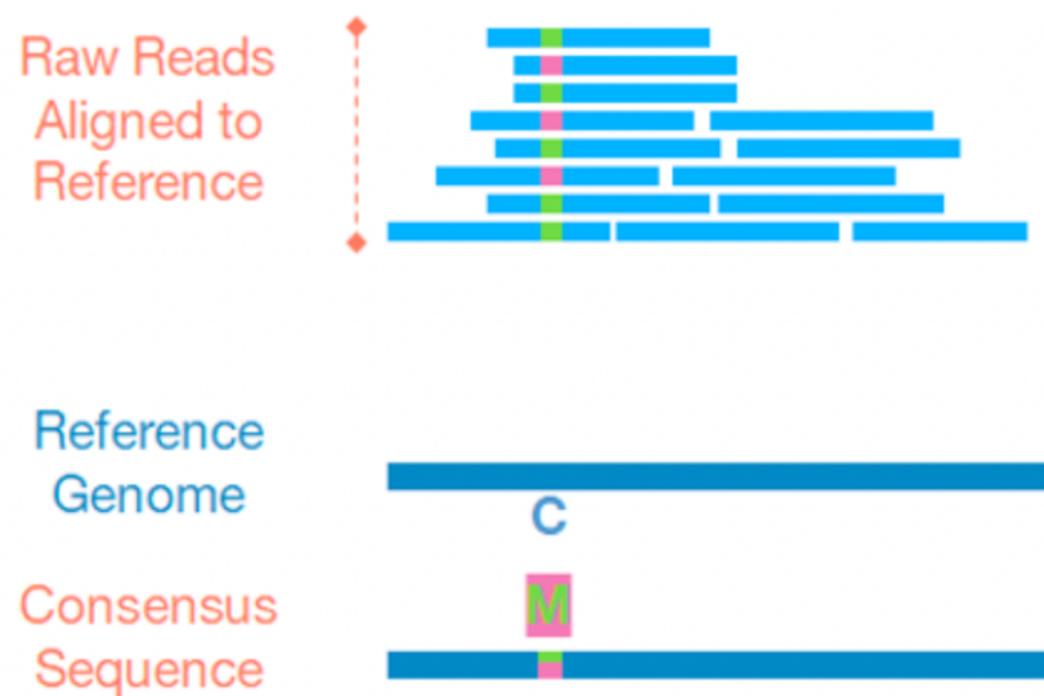
## Checking read alignments

- Gaps or low coverage of consensus genome



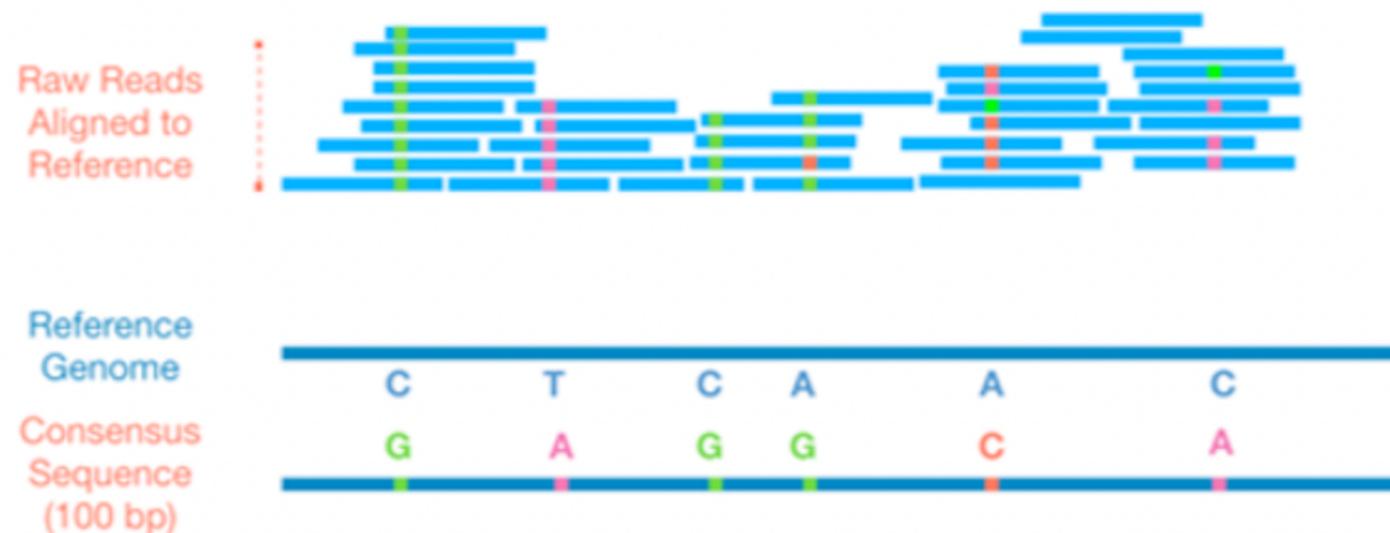
# Checking read alignments

- Ambiguous bases



# Checking read alignments

- Single nucleotide polymorphisms (SNPs)



# GISAID quality control

	Virus name	Passage de	Accession ID	Collection da	Submission d		Length	Host	Location	Originating	
<input type="checkbox"/>	hCoV-19/Philippines/PH-PGC-132210/2022	Original	EPI_ISL_17389138	2022-09-26	2023-04-04		29,867	Human	Asia / Philippines	DAVAO INT	
<input type="checkbox"/>	hCoV-19/Philippines/PH-PGC-132213/2022	Original	EPI_ISL_17389137	2022-09-26	2023-04-04		This submission requires further investigation! It appears to contain markers multiple lineages from both Delta and Omicron variants				
<input type="checkbox"/>	hCoV-19/Philippines/PH-PGC-131761/2022	Original	EPI_ISL_17389136	2022-09-26	2023-04-04		29,838	Human	Asia / Philippines	KIDAPAWA	
<input type="checkbox"/>	hCoV-19/Philippines/PH-PGC-132196/2022	Original	EPI_ISL_17389135	2022-09-26	2023-04-04		29,847	Human	Asia / Philippines	OTHERS - I	
<input type="checkbox"/>	hCoV-19/Philippines/PH-PGC-131762/2022	Original	EPI_ISL_17389134	2022-09-26	2023-04-04		29,838	Human	Asia / Philippines	KIDAPAWA	
<input type="checkbox"/>	hCoV-19/Philippines/PH-PGC-132215/2022	Original	EPI_ISL_17389133	2022-09-26	2023-04-04		29,878	Human	Asia / Philippines	DAVAO INT	
<input type="checkbox"/>	hCoV-19/Philippines/PH-PGC-131727/2022	Original	EPI_ISL_17389132	2022-09-25	2023-04-04		29,880	Human	Asia / Philippines	ST. ELIZAB	
<input type="checkbox"/>	hCoV-19/Philippines/PH-PGC-131723/2022	Original	EPI_ISL_17389131	2022-09-25	2023-04-04		29,882	Human	Asia / Philippines	ST. ELIZAB	
<input type="checkbox"/>	hCoV-19/Philippines/PH-PGC-131728/2022	Original	EPI_ISL_17389130	2022-09-25	2023-04-04		29,854	Human	Asia / Philippines	ST. ELIZAB	
<input type="checkbox"/>	hCoV-19/Philippines/PH-PGC-131726/2022	Original	EPI_ISL_17389129	2022-09-25	2023-04-04		29,877	Human	Asia / Philippines	ST. ELIZAB	
<input type="checkbox"/>	hCoV-19/Philippines/PH-PGC-131725/2022	Original	EPI_ISL_17389128	2022-09-25	2023-04-04		29,846	Human	Asia / Philippines	ST. ELIZAB	
<input type="checkbox"/>	hCoV-19/Philippines/PH-PGC-131680/2022	Original	EPI_ISL_17389127	2022-09-24	2023-04-04		29,866	Human	Asia / Philippines	DR. ARTUF	
<input type="checkbox"/>	hCoV-19/Philippines/PH-PGC-131682/2022	Original	EPI_ISL_17389126	2022-09-24	2023-04-04		29,249	Human	Asia / Philippines	DR. ARTUF	

# Nextclade quality control

The screenshot shows the Nextclade software interface. At the top, there are navigation buttons: Start, Dataset, Results (which is highlighted in blue), Tree, Export, and a status message "Done. Total sequences: 5. Succeeded: 5". To the right of the status are links for Settings, About, Citation, Docs, CLI, and language selection (EN). Below the header is a table with the following columns:

i	Sequence name	QC	Clade	Pango lineage (Nextclade)	WHO name	Mut.	non-ACGTN	Ns	Cov.	Gaps	Ins.	FS	SC	Nucleotide sequence
2	✓ hCoV-19/Philippines/PH-PGC-38292/2021	N M P C F S	20B	B.1.1.28		25	0	13	99.8%	0	0	0	0	[Barcode]
3	✓ hCoV-19/Philippines/PH-PGC-38288/2021	N M P C F S	20B	B.1.1.28		25	0	11	99.8%	0	0	0	0	[Barcode]
1	✓ hCoV-19/Philippines/PH-PGC-14224/2021	N M P C F S	20B	B.1.1.263		26	0	13	99.8%	0	0	0	0	[Barcode]
4	✓ hCoV-19/Philippines/PH-PGC-113124/2021	N M P C F S	21I	B.1.617.2	Delta	70	0	184	99.1%	22	0	0	0	[Barcode]
0	✓ hCoV-19/Philippines/PH-PGC-113286/2021	N M P C F S	21I	B.1.617.2	Delta	52	0	0	99.7%	13	0	0	0	[Barcode]

# Nextclade quality scores

- **Individual Scores**

- **Missing data (N)**: If your sequence misses more than 3000 sites (N characters), it will be flagged as bad
- **Mixed sites (M)**: Ambiguous nucleotides are often indicative of contamination
- **Private mutations (P)**: As a by-product of the phylogenetic placement method of assigning lineages, Nextclade identifies the mutations, called “private mutations”, that differ between the query sequence and the nearest neighbor sequence.
- **Mutation clusters (C)**: To be more sensitive for quality problems in a narrow area of a genome, the mutation cluster rule counts the number of private within all possible 100-nucleotide windows
- **Stop codons (S)**: Premature stops
- **Frame shifts (F)**: Wrong grouping of codons

- **Overall QC score**

- multiple mildly concerning scores don't result in a bad overall score, but a single bad score guarantees a bad overall score.
- lower value means better quality, higher value means worse quality

# Practical: Quality assessment on consensus

Open file: msa\_consensus\_rsv.fasta

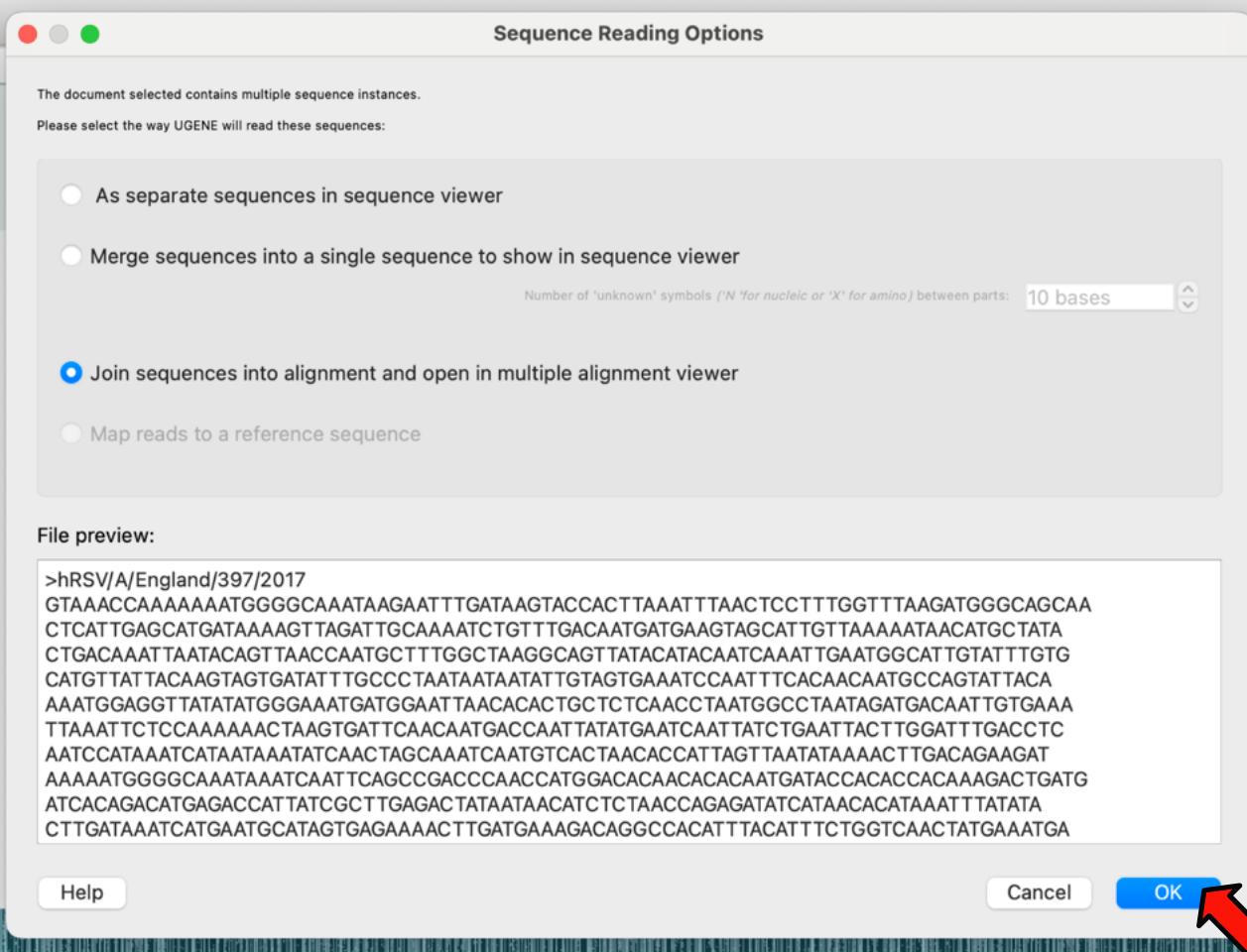
-\* UGENE

Welcome to UGENE

1: Project

Open File(s)

Run or Create Workflow



Cite UGENE:

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

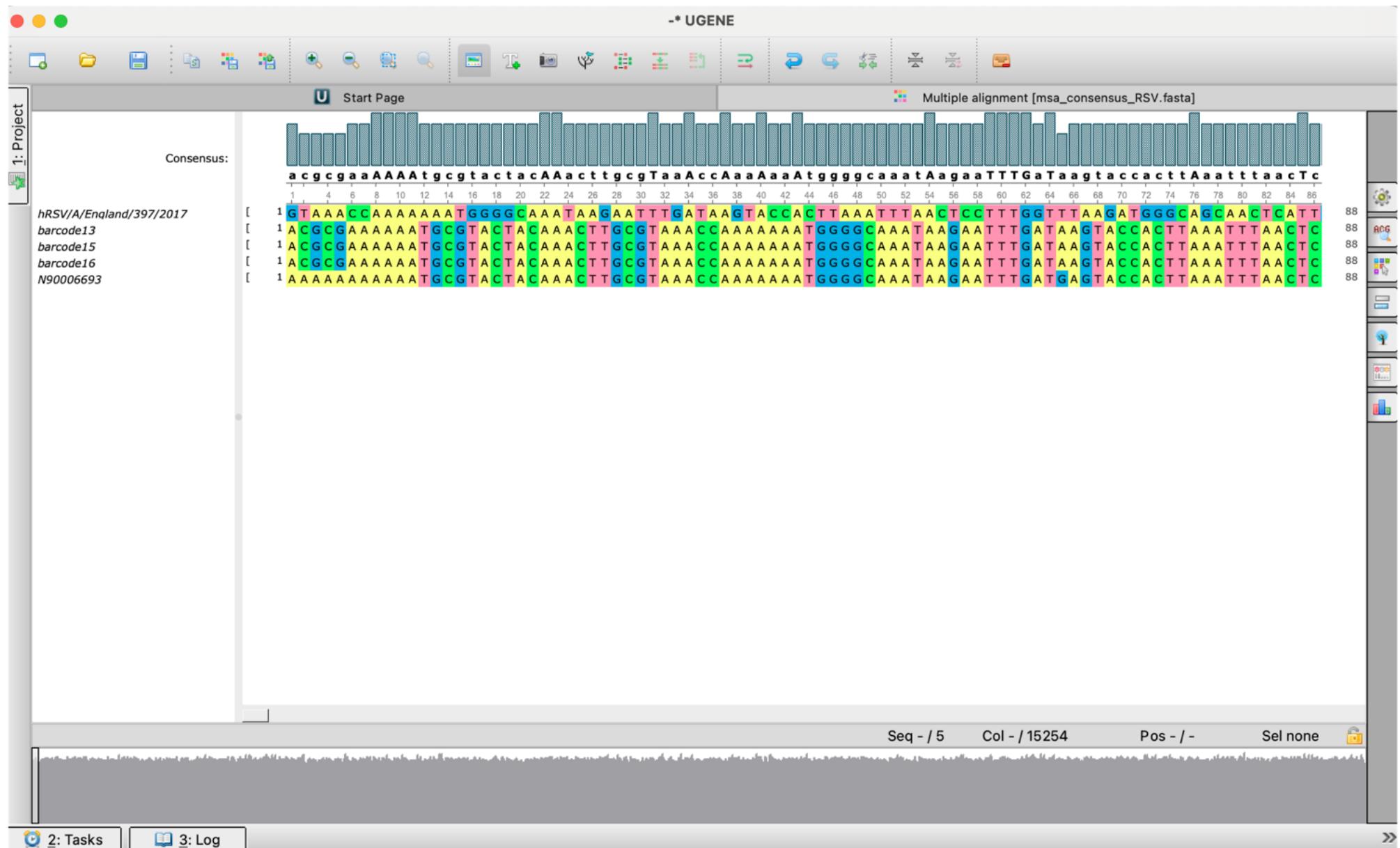
Follow UGENE:



2: Tasks

3: Log

>>



Unipro UGENE

File Actions Settings Tools Window Help

Project 1: Consensus mode... hRSV/A/England/397/2017 barcode13 barcode15 barcode16 N90006693

Multiple alignment [msa\_consensus\_RSV.fasta]

Align with MAFFT...

Align sequences to alignment with MAFFT...

Align selected sequences to alignment with MAFFT...

Align sequences to alignment with MUSCLE...

Align selected sequences to alignment with MUSCLE...

Align with ClustalO...

Align with ClustalW...

Align with Kalign...

Align with T-Coffee...

Align with MAFFT...

Align with MUSCLE...

Align alignment to alignment with MUSCLE...

Align alignment to alignment with ClustalO...

Align selected sequences to alignment with MAFFT...

Align selected sequences to alignment with MUSCLE...

Seq - / 5 Col - / 15254 Pos - / - Sel none

2: Tasks 3: Log

Align sequences using MAFFT

Unipro UGENE

File Actions Settings Tools Window Help

1: Project

Consensus:

hRSV/A/England/397/2017  
barcode13  
barcode15  
barcode16  
N90006693

Start Page

Align with MAFFT

Advanced options

- Gap opening penalty: 1.53
- Offset (works like gap extension penalty): 0.00
- Maximum number of iterative refinement: 0

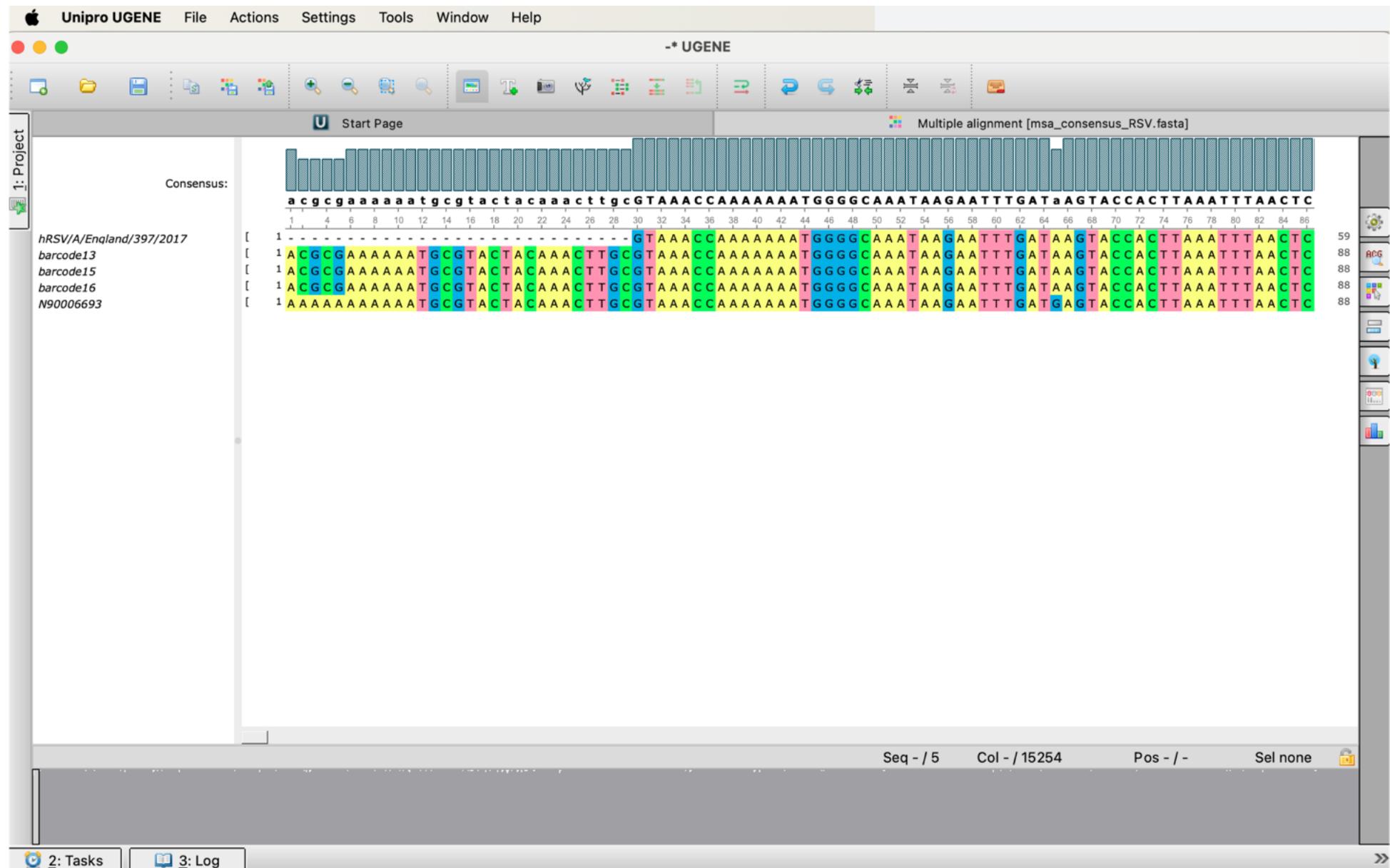
Help Cancel Align

Multiple alignment [msa\_consensus\_RSV.fasta]

Seq - / 5 Col - / 15254 Pos - / - Sel none

2: Tasks 3: Log

The screenshot shows the Unipro UGENE interface with a project titled '1: Project'. On the left, there's a tree view with items like 'hRSV/A/England/397/2017', 'barcode13', 'barcode15', 'barcode16', and 'N90006693'. The main workspace displays a sequence alignment. A dialog box titled 'Align with MAFFT' is open, containing 'Advanced options' with three checkboxes: 'Gap opening penalty' (set to 1.53), 'Offset (works like gap extension penalty)' (set to 0.00), and 'Maximum number of iterative refinement' (set to 0). Below these are 'Help' and 'Cancel' buttons, followed by a large blue 'Align' button which has a red arrow pointing to it. To the right of the dialog is a 'Multiple alignment [msa\_consensus\_RSV.fasta]' viewer showing sequence data from positions 56 to 86. The bottom status bar indicates 'Seq - / 5', 'Col - / 15254', 'Pos - / -', and 'Sel none'. At the bottom, there are tabs for '2: Tasks' and '3: Log'.



Unipro UGENE File Actions Settings Tools Window Help

-\* UGENE

1: Project

Consensus:

hRSV/A/England/397/2017  
barcode13  
barcode15  
barcode16  
N90006693

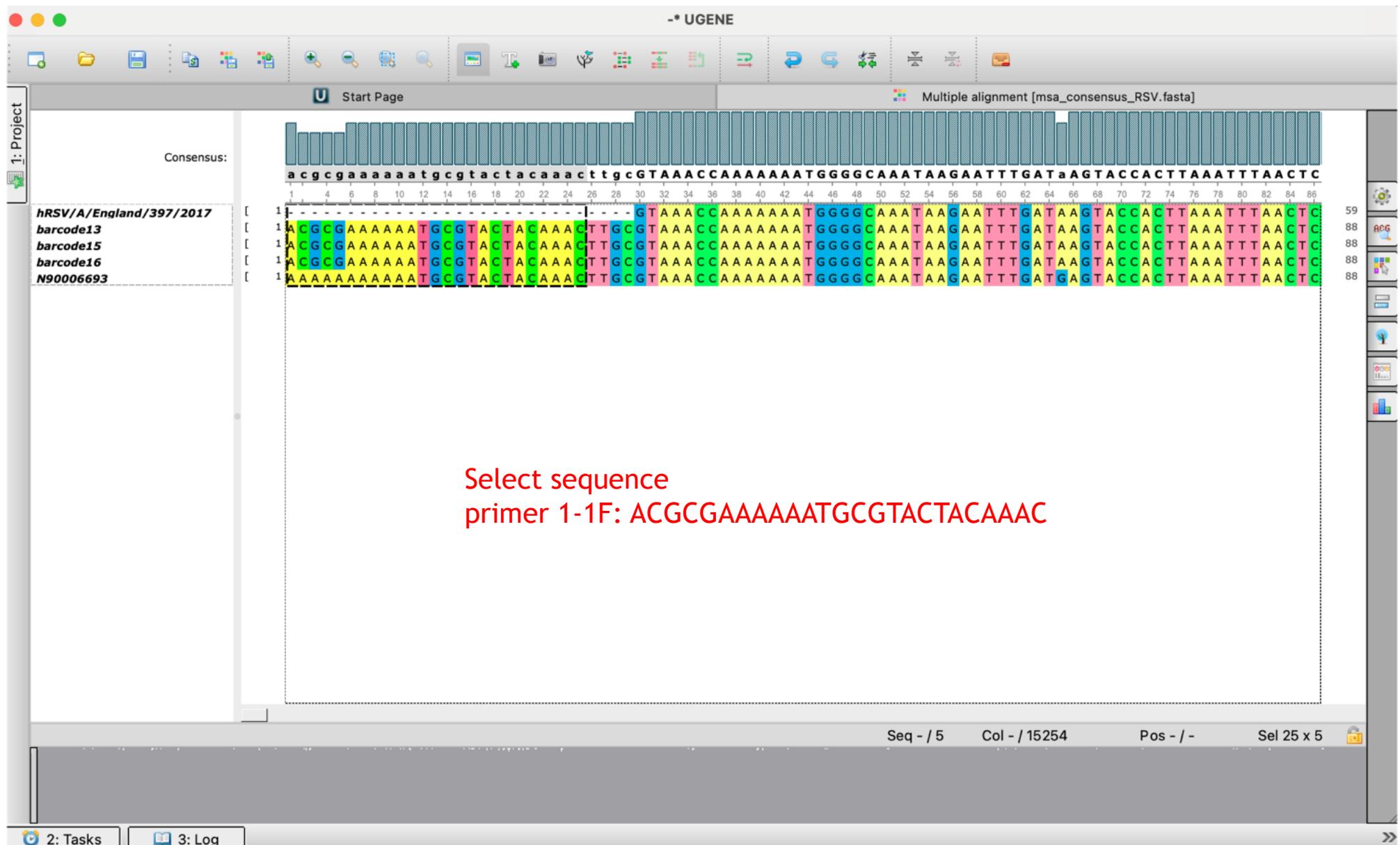
Multiple alignment [msa\_consensus\_RSV.fasta]

Look for and remove 1-1F:  
ACGCGAAAAAATGCGTACTACAAAC

Seq - / 5 Col - / 15254 Pos - / - Sel none

2: Tasks 3: Log >>

The screenshot displays the Unipro UGENE bioinformatics platform. The main window shows a multiple sequence alignment titled "Multiple alignment [msa\_consensus\_RSV.fasta]". The alignment consists of a consensus sequence at the top and several individual sequences below. A specific sequence, "1 ACGCGAAAAAATGCGTACTACAAAC", is highlighted with a red box and labeled "1-1F". A red annotation text "Look for and remove 1-1F:" is overlaid on the screen, followed by the sequence "ACGCGAAAAAATGCGTACTACAAAC". The UGENE interface includes a toolbar, a project tree on the left, and various status indicators at the bottom.



Unipro UGENE

File Actions Settings Tools Window Help

1: Project

Consensus mode... N90006693

hRSV/A/England/397/2017 barcode13 barcode15 barcode16

Multiple alignment [msa\_consensus\_RSV.fasta]

-\* UGENE

Edit sequence name F2

Replace with gaps ⌘ Space

Replace selected character ⌘ R

Replace selected rows with reverse-complement

Replace selected rows with reverse

Replace selected rows with complement

Remove columns of gaps...

Remove all gaps

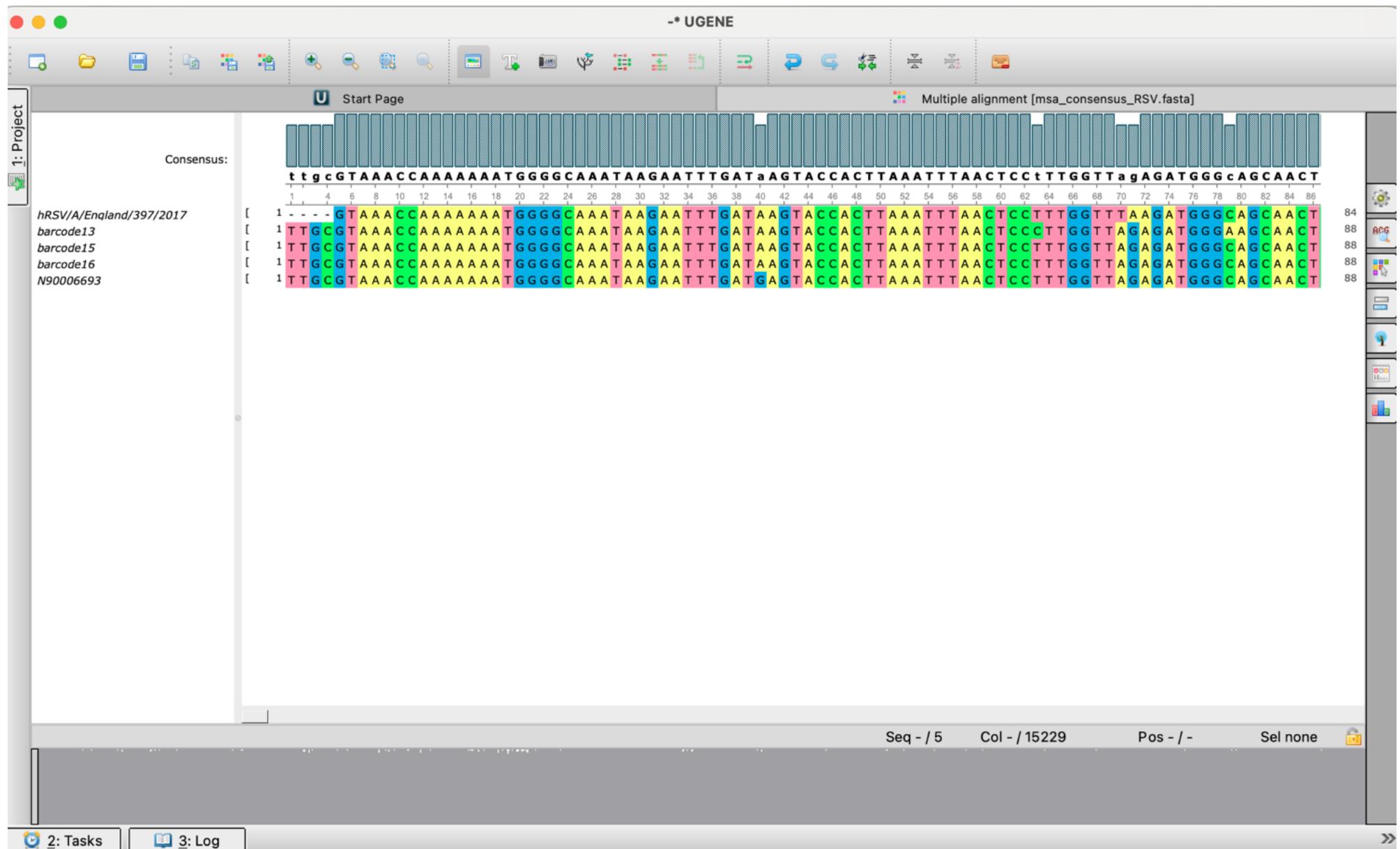
Remove sequence(s)

Remove selection

Seq - / 5 Col - / 15254 Pos - / - Sel 25 x 5

2: Tasks 3: Log

The screenshot shows the Unipro UGENE interface with a multiple sequence alignment window titled "Multiple alignment [msa\_consensus\_RSV.fasta]". The alignment displays several DNA sequences with color-coded bases (A, T, C, G) across 15254 positions. A context menu is open, listing various edit options like "Edit sequence name", "Replace with gaps", and "Remove selection". A red arrow points to the "Remove selection" option at the bottom of the menu.



\* UGENE

Start Page

Multiple alignment [msa\_consensus\_RSV.fasta]

Consensus:

hRSV/A/England/397/2017  
barcode13  
barcode15  
barcode16  
N90006693

15140 AAAATTATCATTGATCTAGGAAGAATAAGTTAAATCCAAATCTAATTGGTTATATGTATATTaACTAAATTACGAGAcATTA 15225  
15144 AAAATTATCATTGATCTAGGAAGAATAAGTTAAATCCAAATCTAATTGGTTATATGTATATTaACTAAATTACGAGAcATTA 15229  
15144 AAAATTATCATTGATCTAGGAAGAATAAGTTAAATCCAAATCTAATTGGTTATATGTATATTaACTAAATTACGAGAcATTA 15229  
15144 AAAATTATCATTGATCTAGGAAGAATAAGTTAAATCCAAATCTAATTGGTTATATGTATATTaACTAAATTACGAGAcATTA 15229  
15144 AAAATTATCATTGATCTAGGAAGAATAAGTTAAATCCAAATCTAATTGGTTATATGTATATTaACTAAATTACGAGAcATTA 15229

Look for and remove 2-6R:AGTGTCAAAAACTAATRTCTCGT

Seq - / 5 Col - / 15229 Pos - / - Sel none

2: Tasks 3: Log

\* UGENE

Start Page

Multiple alignment [msa\_consensus\_RSV.fasta]

Consensus:

hRSV/A/England/397/2017  
barcode13  
barcode15  
barcode16  
N90006693

15140 AAAATTATCATTGATCTAGGAAGAATAAGTTAAATCCAAATCTAATTGGTTATATGTATATTaACTAAATTACGAGAcATTA 15225

15144 AAAATTATCATTGATCTAGGAAGAATAAGTTAAATCCAAATCTAATTGGTTATATGTATATTaACTAAATTACGAGAcATTA 15229

15144 AAAATTATCATTGATCTAGGAAGAATAAGTTAAATCCAAATCTAATTGGTTATATGTATATTaACTAAATTACGAGAcATTA 15229

15144 AAAATTATCATTGATCTAGGAAGAATAAGTTAAATCCAAATCTAATTGGTTATATGTATATTaACTAAATTACGAGAcATTA 15229

15144 AAAATTATCATTGATCTAGGAAGAATAAGTTAAATCCAAATCTAATTGGTTATATGTATATTaACTAAATTACGAGAcATTA 15229

Select primer sequence  
2-6R:AGTGTCAAAAACTAATRTCTCGT

Seq - / 5 Col - / 15229 Pos - / - Sel 11 x 5

2: Tasks 3: Log

Unipro UGENE

File Actions Settings Tools Window Help

1: Project

Consensus mode... 15195 15.2k 15205 15210 15215 15220 15225 15229

hRSV/A/England/397/2017 barcode13 barcode15 barcode16 N90006693

Multiple alignment [msa\_consensus\_RSV.fasta]

-\* UGENE

Edit sequence name F2

Replace with gaps ⌘ Space

Replace selected character ⌘ R

Replace selected rows with reverse-complement

Replace selected rows with reverse

Replace selected rows with complement

Remove columns of gaps...

Remove all gaps

Remove sequence(s)

Remove selection

Seq - / 5 Col - / 15229 Pos - / - Sel 11 x 5

2: Tasks 3: Log >>

The screenshot shows the Unipro UGENE interface. A context menu is open under the 'Edit' option in the main menu. The 'Remove selection' option at the bottom of the submenu is highlighted with a red arrow. The main workspace displays a multiple sequence alignment titled 'Multiple alignment [msa\_consensus\_RSV.fasta]'. The alignment consists of several DNA sequences with various mutations and colors. The bottom status bar shows sequence information: Seq - / 5, Col - / 15229, Pos - / -, and Sel 11 x 5.

-\* UGENE

Start Page

Multiple alignment [msa\_consensus\_RSV.fasta]

1: Project

Consensus:

hRSV/A/England/397/2017  
barcode13  
barcode15  
barcode16  
N90006693

15133 15140 15145 15150 15155 15160 15165 15170 15175 15180 15185 15190 15195 15.k 15205 15210 15 218 15214] 15218] ACG 15218] 15218] 15218] 15218]

Check ORF  
Select one sequence, Ctrl+C

Seq - / 5 Col - / 15218 Pos - / - Sel none

2: Tasks 3: Log >>

Unipro UGENE

File Actions Settings Tools Window Help

New project... **New document from text...** ↑

New workflow...

Open... Open as... Open from clipboard...

Access remote database... Search NCBI GenBank...

Recent files Recent projects

Save all Save project as... Export project... Close project

-\* UGENE

Multiple alignment [msa\_consensus\_RSV.fasta]

1: Project

hRSV/A/England/397/2015  
barcode13  
barcode15  
barcode16  
N90006693

CCTAAAAATTATCATTGGATCTAGGAAGAAATAAGTTAAATCCAAATCTAATTGGTTTATATGTATATTAA

15145 15150 15155 15160 15165 15170 15175 15180 15185 15190 15.2k 15205 15.211 15209 15213 15213 15213 15213 15213

Seq 2 / 5 Col - / 15218 Pos - / - Sel 15218 x 1

2: Tasks 3: Log

Unipro UGENE

File Actions Settings Tools Window Help

Start Page

Consensus:

hRSV/A/England/397/2017

barcode13

barcode15

barcode16

N90006693

Paste data here

```
CTGCATTGTCTAAATTAAAGAGTGTGTTAGTGGAGATACTATCATATTCTA  
TAGCTGGCGTAATGAAGTTCAGCAATAAAACTTATAATCATAAGCATATG  
AACATCTTAAAGTGGTCAATCATGTTAAATTTCAGATCAACAGAAATAAA  
CTATAATCTTATATGGTAGAATCTACTTATCCATCATTAAGTGAATTGTTA  
AACAGCTTGACAACCAATGAACTAAAAAACTGATTAACAGGTAGTTT  
GTATACAACTTTATAATGAATAATGAGCAAAATCTTATAACAGAAATAGCT  
ACACACTAACATGTATTCAATTATAGTTTTAAATTAAATAATTATATAATT  
TTAATAACTCTAGTGAACCAATCTAAATTATCATTTGATCTAGGAAGAAT  
AAGTTAAATCCAATCTAATTGGTTATATGTATATTAACTAAATT
```

Custom settings

Alphabet: Standard DNA

Skip unknown symbols

Replace unknown symbols with

Save sequence to file  ...

File format FASTA

Sequence name barcode13\_qc

Save file immediately

Help Cancel Create

then, Ctrl+V

Seq 2 / 5 Col - / 15218 Pos - / - Sel 15218 x 1

2: Tasks 3: Log

Project

msa\_consensus\_RSV.fasta

CTAATTGGTTTATATGTATATTAA 15209

CTAATTGGTTTATATGTATATTAA 15213

CTAATTGGTTTATATGTATATTAA 15213

CTAATTGGTTTATATGTATATTAA 15213

Unipro UGENE File Actions Settings Tools Window Help

-\* UGENE

Project

barcode13\_qc [dna]

Multiple alignment [msa\_consensus\_RSV.fasta]

barcode13\_qc [barcode13\_qc.fasta]

Start Page

Go

1:1 ACG SW

15.218

ACG

GT CAT C AAC T T T G G T T T T T A C C C G T T T A T T C T A A C T A T T C A T G G T G A A T T A A T G G A A C C A A T C T C A C C C T C G T T G A G T A A C T C A T A C T A T T T C A A T C T A A C G T T T A G A C A A A C T G T T A

T G A A G T A G C A T T G T T A A A A T A A C A T G C T A T A C T G A C A A A T T A A T A C A G T T A A C T A A T G C T T A G C T A A G G C A G T T A T A C A T A C A A T C A A A T T G A A T G G C A T T G T A T T T G T G C A T G T T A T T A C A A G T A

CT A C T T C A T C G T A A C A A T T T T A T T G T A C G A T A T G A C T G T T A A T T A T G T C A A T T G A T T C G A A A T C G A T T C C G T C A A T A T G T A T G T T A G T T A A C T T A C C G T A A C A A C A C G T A C A A T A A T G T T C A T

Name Type Value

Auto-annotations [barcode13\_qc.fasta ...]

2: Tasks 3: Log

Open GenBank file: RSVA\_annotations.gb  
Drag annotations to target sequence: barcode13\_qc.fasta

The screenshot shows the UGENE software interface. On the left, the Project panel displays several files: msa\_consensus\_RSV.fasta, barcode13\_qc.fasta, and RSVA\_annotations.gb. A blue arrow points from the RSVA\_annotations.gb entry towards the sequence viewer. The main workspace shows a sequence viewer for barcode13\_qc.fasta [dna]. The sequence is displayed with a ruler at the top ranging from 1k to 15k. Below the sequence, two lines of nucleotide sequence are shown with their corresponding amino acid translations (GT, CAT, C, T, T). At the bottom of the sequence viewer, there is a table for "Auto-annotations [barcode13\_qc.fasta ...]". The bottom navigation bar includes tabs for Tasks (2) and Log (3).

Unipro UGENE File Actions Settings Tools Window Help

-\* UGENE

Project

Search... Objects

- msa\_consensus\_RSV.fasta [m] Multiple alignment
- barcode13\_qc.fasta [s] barcode13\_qc
- RSVA\_annotations.gb [a] Annotations

Bookmarks

- Multiple alignment [msa\_consensus\_RSV.fasta]
- barcode13\_qc [barcode13\_qc.fasta]

Start Page Multiple alignmer

barcode13\_qc [dna]

Select sequence to associate annotations with:

barcode13\_qc

Cancel OK

GT CAT TTGCGTAAACCAAAAAAATGGGGCAAATAAGA C 1 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 AACGCATTGGTTTTTACCCGTTATTCT T CAACTCATTGAGTATGATAAAAGTTAGATTGCAAATCTGTTGACAATGATGAAGTAGCATTGTTAAAATACATGCTA GTTGAGTAACTCATACTATTTCAATCTAACGTTTAGACAAACTGTTACTTCATCGTAACAATTTTATTGTACGAT

Name Type Value

Auto-annotations [barcode13\_qc.fasta ...]

2: Tasks 3: Log

37% 10:51PM

Wednesday, June 5, 2019

A red arrow points to the "OK" button in the "Edit Object Relations" dialog box.

Unipro UGENE File Actions Settings Tools Window Help

-\* UGENE

Multiple alignment [msa\_consensus\_RSV.fasta] barcode13\_qc [barcode13\_qc.fasta]

Project 1: barcode13\_qc [dna]

Click on each gene

Name	Type	Value
> F	gene	5701..7422
> G (0,1)		
> L (0,1)		

GT  
CAT ATCTCTATCTCCTCCAACACAACAAAA TGATAGTCATTAAAAAGCGTATTGGCAAAAAACATGACCAAATAAAACAGAATCAAATTAACTCTGGGGCAAATAACA ATGGAGTTGCCAATCCTCAA  
C 5591 5595 5.6k 5605 5610 5615 5620 5625 5630 5635 5640 5645 5650 5655 5660 5665 5670 5675 5680 5685 5690 5695 5.7k 5705 5710 5715 5720  
TAGAGATAGAAGGAGGTTGTGTTACTATCAGTAATTTCGCTAACACGTTGGTACTGGTTATTGTCTTAGTTAATTGAGACCCGTTATTGTTACCTAACCGTTAGGAGTT

2: Tasks 3: Log

Unipro UGENE File Actions Settings Tools Window Help

-\* UGENE

Multiple alignment [msa\_consensus\_RSV.fasta] barcode13\_qc [barcode13\_qc.fasta]

Project 1: barcode13\_qc [dna]

Genomic track visualization showing genes F, G, L, M, N, P, NS1, NS2, M2-1, M2-2, SH, and K across a genomic region from 1 to 15,218 bp. Gene F is green, gene G is yellow, gene L is pink, gene M is purple, gene N is green, gene P is pink, gene NS1 is yellow, gene NS2 is purple, gene M2-1 is pink, gene M2-2 is purple, gene SH is blue, and gene K is orange.

Check for complete ORF  
Correct reading frame, no premature stop codon

Sequence alignment details:

- Gene F: 4551-4656 bp
- Gene G: 4565-4605 bp
- Gene L: 4620-4680 bp
- Gene M: 4635-4675 bp
- Gene N: 4645-4665 bp
- Gene P: 4655-4675 bp
- Gene NS1: 4665-4675 bp
- Gene NS2: 4675-4680 bp
- Gene M2-1: 4680-4690 bp
- Gene M2-2: 4690-4700 bp
- Gene SH: 4700-4710 bp
- Gene K: 4710-4720 bp

Sequence details:

```

GT
CAT
CACCATGCAAGCCATCATCCATACCAAAAGTAGTTAATTAAAAAAATAGTCATAACAATGAACCTAGGATATTAAGACCAAAAACACGCTGGGGCAATGCAAACATGTCCAAAACCAAGGACCAACGCA
T
4551 4555 4560 4565 4570 4575 4580 4585 4590 4595 46k 4605 4610 4615 4620 4625 4630 4635 4640 4645 4655 4660 4665 4670 4675 4680
GTGGTACGTTCGTAGAGGTATGGTATTTCATCAATTAAATTTTTATCAGTATTGTTACTTGATCCTATAATTCTGGTTTTGTTGCACCCGTTACGTTTGACAGGTTGGTTCTGGTTGC
T
T
CCGCCAAGACTCTAGAAAGGACCTGGGACACTCTCAATCATCTGTTATTCAATCATCGTGCTTATACAAGTTAAATCTTAAATCTATAGCACAAATCACATTATCTATTGGCAATGATAATCTCAAC
4681 4685 4690 4695 4.7k 4705 4710 4715 4720 4725 4730 4735 4740 4745 4750 4755 4760 4765 4770 4775 4780 4785 4790 4795 4.8k 4805 4810
GGCGGTTCTGAGATCTTCCTGGACCCCTGTGAGAGTTAGTAGACAATAAGTATAGTAGCAGAACATATGTTCAATTAGAATTAGATATCGTGTAGTGTAAATAGATAACCGTTACTATTAGAGTTG
CTCACTTATAATTGCAGCCATCATATTCAAGCCTCGGAAACCACAAAGTCACACTAACACTGCAATCATACAAGATGCAACGAACCAAGATCAAGAACACAAACCCCACATACCTCACCCAGAACCCC

```

Name	Type	Value
> F	gene	5701.7422
> G (0,1)	gene	4656..5618
> G	gene	

Tasks Log

# Thank you !

[clyde.dapat@influenzacentre.org](mailto:clyde.dapat@influenzacentre.org)



WHO Collaborating Centre  
for Reference and  
Research on Influenza  
**VIDRL**



A joint venture between The University of Melbourne and The Royal Melbourne Hospital

