Annex 17 Generating Consensus Sequences with Sequencher

1. Purpose

The purpose of the procedure is to derive the sequence of the 450 nucleotides encoding the carboxyl terminus of the N gene of measles virus, or the 739 nucleotides encoding the sequencing window of the E1 gene of rubella virus, or the approximately 1018 nucleotides from the non-coding region of the measles virus between the M and F coding regions (MF-NCR), from an existing PCR product.

2. Related Documents

Title
Measles Genotyping RT-PCR protocol
Rubella Genotyping RT-PCR protocol
Measles MF-NCR RT-PCR protocol
Agarose Gel Electrophoresis Protocol for Analysis of PCR Products
PCR Products Clean-Up Protocol
Sequencing Reaction Set-up Protocol
Sequencing Reactions Clean-Up Protocol

3. Definitions

Term	Definition
WHO	World Health Organization
MeV	Measles Virus
RuV	Rubella Virus
MF-NCR	Non-Coding Region between M and F coding regions of measles genome
MeaNS	Measles Nucleotide Sequence Database
RubeNS	Rubella Nucleotide Sequence Database
ID	Identification
Indel	Insertion or deletion
Guide sequence	N-450 or MF-NCR or E1-739 sequence FASTA file cropped to the correct window

4. Equipment/Software

- Computer for using Sequencher and Mega software
- USB/flash drive to transfer chromatogram (.abi) files
- Microsoft Office Power Point software (for editing phylogenetic trees)
- Sequencher sequence analysis software USB toggle

5. Supplies for Using Sequencher

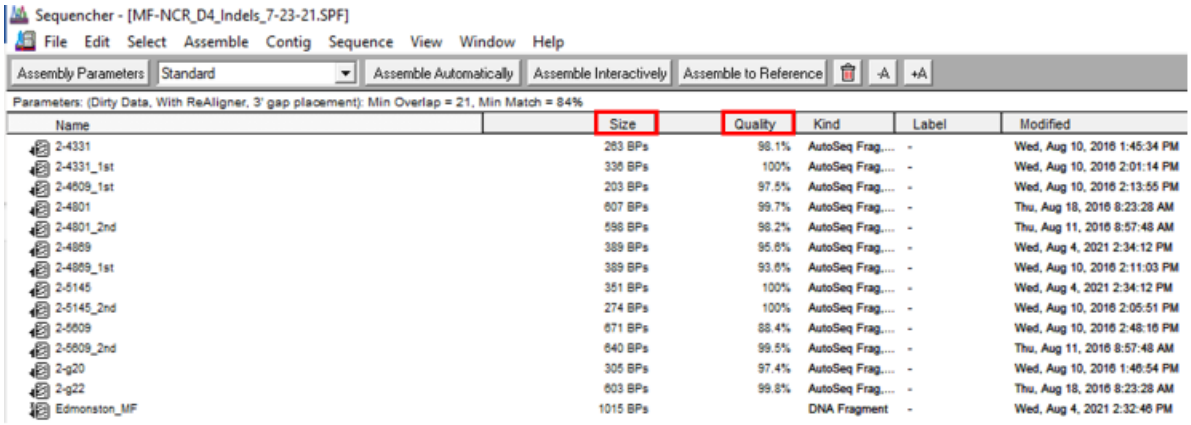
One of the following, depending on the analysis being performed:

- Measles N450 guide sequence
- Rubella E1-739 guide sequence
- Measles MF-NCR guide sequence

6. **Safety Precautions**
   To protect the computers from viruses, make sure to update anti-virus software. It is recommended to use a secure flash drive to ensure the computers remain virus-free.

7. **Sample Transfer**
   A. Transfer the chromatogram files to the flash drive once the sequencing run has completed.
   B. Copy the chromatogram files (.abi) from the flash drive into a separate folder on a computer before beginning the sequence analysis. Add FASTA files of the guide sequences to this folder. Create a subfolder for each sample.
   C. Save the output files from Sequencher into this folder so that they are easily identified.
   D. Multiple samples can be imported into Sequencher software at once but only edit one sample at a time.

8. **Importation of data**
   A. Insert the Sequencher USB into a port on your computer.
   B. Open the Sequencher program. A window labeled **Untitled project** opens.
   C. Under **File>Import>sequences** find the folder that contains your chromatograms. Highlight the chromatograms you want to import, plus the guide sequence, then click **open**. For the rubella genotyping window and the measles MF-NCR, import all chromatograms for the specimen, even if it was amplified in two PCR fragments. The chromatograms and guide sequence are imported into the main parameters window.



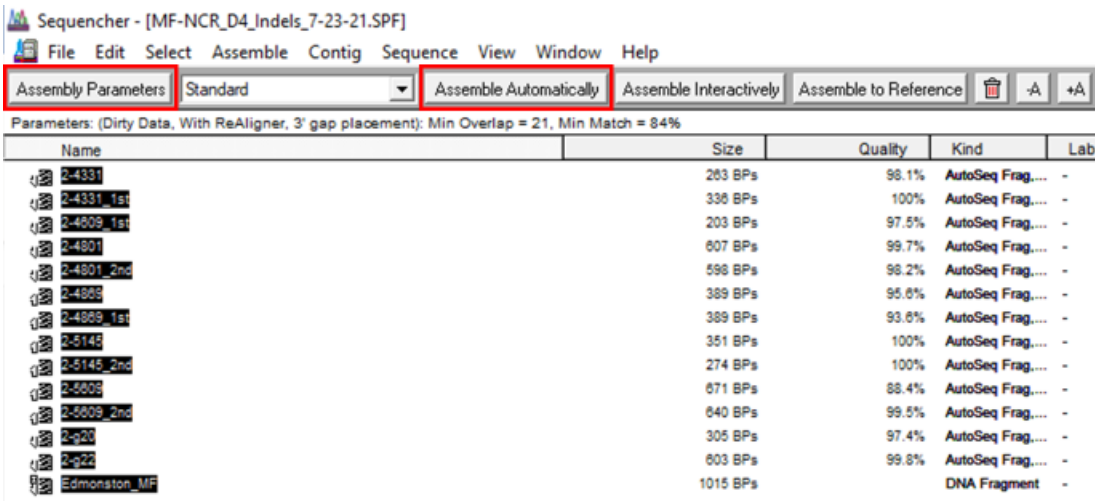
Name	Size	Quality	Kind	Label	Modified
2-4331	283 BPs	98.1%	AutoSeq Frag,...	-	Wed, Aug 10, 2016 1:45:34 PM
2-4331_1st	336 BPs	100%	AutoSeq Frag,...	-	Wed, Aug 10, 2016 2:01:14 PM
2-4609_1st	203 BPs	97.5%	AutoSeq Frag,...	-	Wed, Aug 10, 2016 2:13:55 PM
2-4801	607 BPs	99.7%	AutoSeq Frag,...	-	Thu, Aug 18, 2016 8:23:28 AM
2-4801_2nd	598 BPs	98.2%	AutoSeq Frag,...	-	Thu, Aug 11, 2016 8:57:48 AM
2-4869	389 BPs	95.6%	AutoSeq Frag,...	-	Wed, Aug 4, 2021 2:34:12 PM
2-4869_1st	389 BPs	93.6%	AutoSeq Frag,...	-	Wed, Aug 10, 2016 2:11:03 PM
2-5145	351 BPs	100%	AutoSeq Frag,...	-	Wed, Aug 4, 2021 2:34:12 PM
2-5145_2nd	274 BPs	100%	AutoSeq Frag,...	-	Wed, Aug 10, 2016 2:05:51 PM
2-5609	671 BPs	88.4%	AutoSeq Frag,...	-	Wed, Aug 10, 2016 2:48:16 PM
2-5609_2nd	640 BPs	99.5%	AutoSeq Frag,...	-	Thu, Aug 11, 2016 8:57:48 AM
2-g20	305 BPs	97.4%	AutoSeq Frag,...	-	Wed, Aug 10, 2016 1:46:54 PM
2-g22	603 BPs	99.8%	AutoSeq Frag,...	-	Thu, Aug 18, 2016 8:23:28 AM
Edmonston_MF	1015 BPs		DNA Fragment	-	Wed, Aug 4, 2021 2:32:46 PM

   D. Important: The program has a 'Trim' function. If your sequences are imported into an 'Ends trimming' window first, close that window. Trim parameters can be changed: Click **window>user preference**. Under **add imported sequences to trim window uncheck all imported sequences** box.
   E. Once the sequences are imported, the **Quality** column is a good indication of the quality of each chromatogram. The **Size** column indicates how many nucleotides are in each sequence. Both of these column values will change when the sequences have been edited or trimmed. See picture in section 9.C above. Low quality sequences may not assemble into a contig. It may be possible to crop chromatograms to remove the low-quality regions (see section 11), then assemble again.

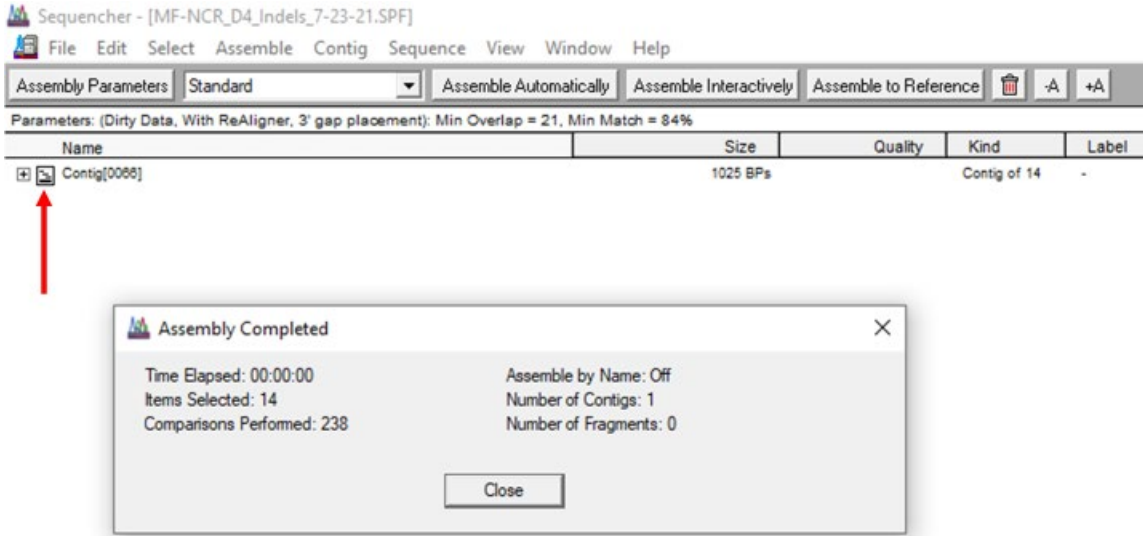
9. Generating a Contig

A. Once the sequences are displayed in the main window, click **assembly parameters**. Verify that the **minimum match percentage** is set to 85%. Minimum overlap should be set to20%. Click **OK** to return to the main window.

B. Highlight the sequences to align. Choose all sequences derived from your sample plus the guide sequence. Click on **Assemble automatically**.

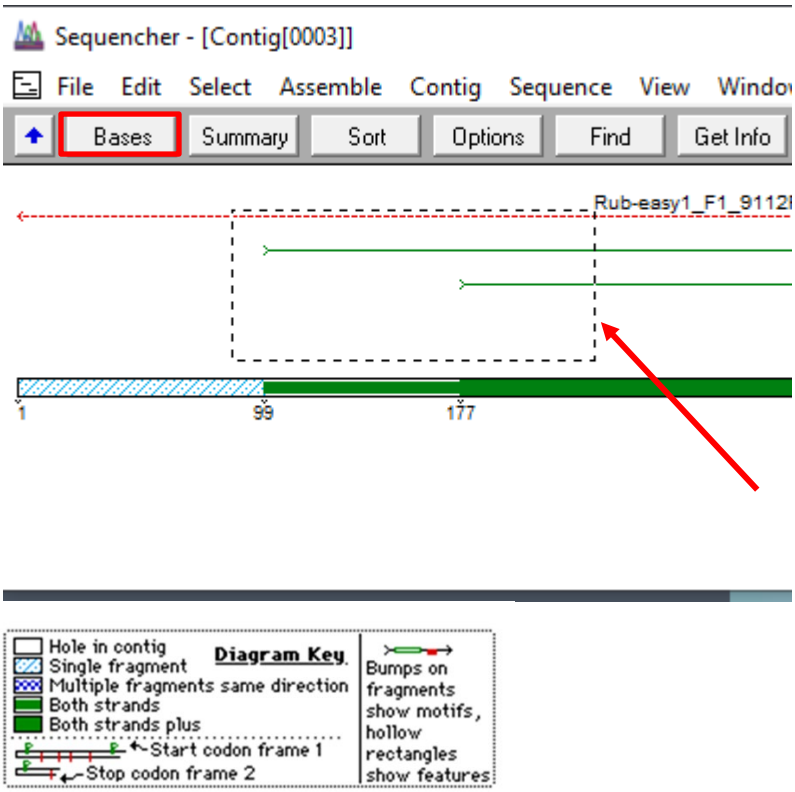


C. A new file will be created, labeled 'contig'.

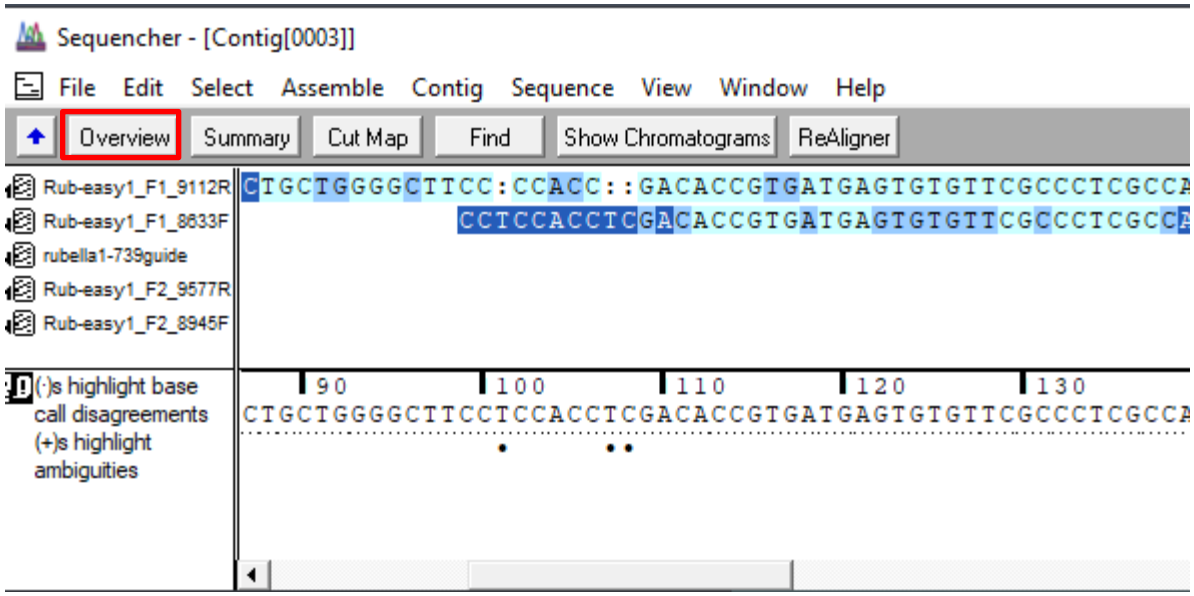


D. If the chosen files do not create a single contig, or if individual sequences will not assemble, first check the quality percent listed. If it is very low, the software may not be able to find sufficient similarity with the other sequences. It may be possible to trim low quality ends (see section 11) and then try to re-assemble.

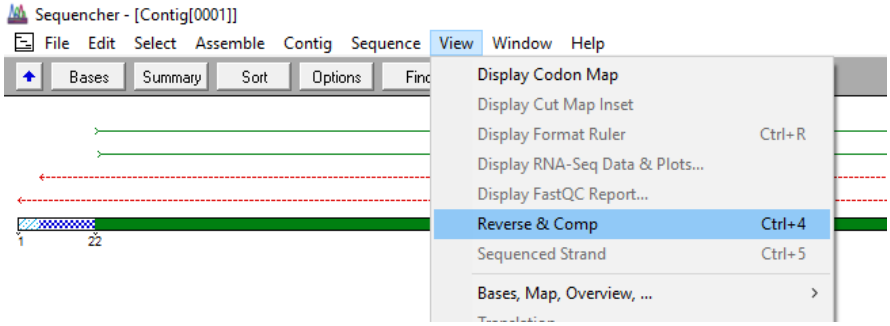
E. Open the contig by double clicking on the square next to its name. A new overview window opens to show a line diagram representing the sequences aligned to each other. Each horizontal line represents a sequence in the contig. The line will be green if the fragment is in the forward orientation and red if it is in the reverse orientation. Below the sequences you can see the coverage map and below the coverage map you will see the start and stop map. The diagram key is at the base of the window.



F. In an open contig window, it is possible to switch from the Overview (with the lines representing sequences) to the Bases view (with the nucleotide sequences) by clicking on 'Bases' (see above). To return to the Overview, click on 'Overview' (see below) at the top left of the contig window. In the Overview, a frame (dashed line, see above) indicates the part of the contig that will be visible when the Bases view is chosen. Click and hold the frame to move it. In the Bases view, the sequence names are listed on the left side and the consensus sequence is located at the bottom. Dots below the consensus sequence indicate mismatches in the alignment. As long as the guide sequence is part of the alignment, there will be many dots because the guide sequence may be a different genotype.



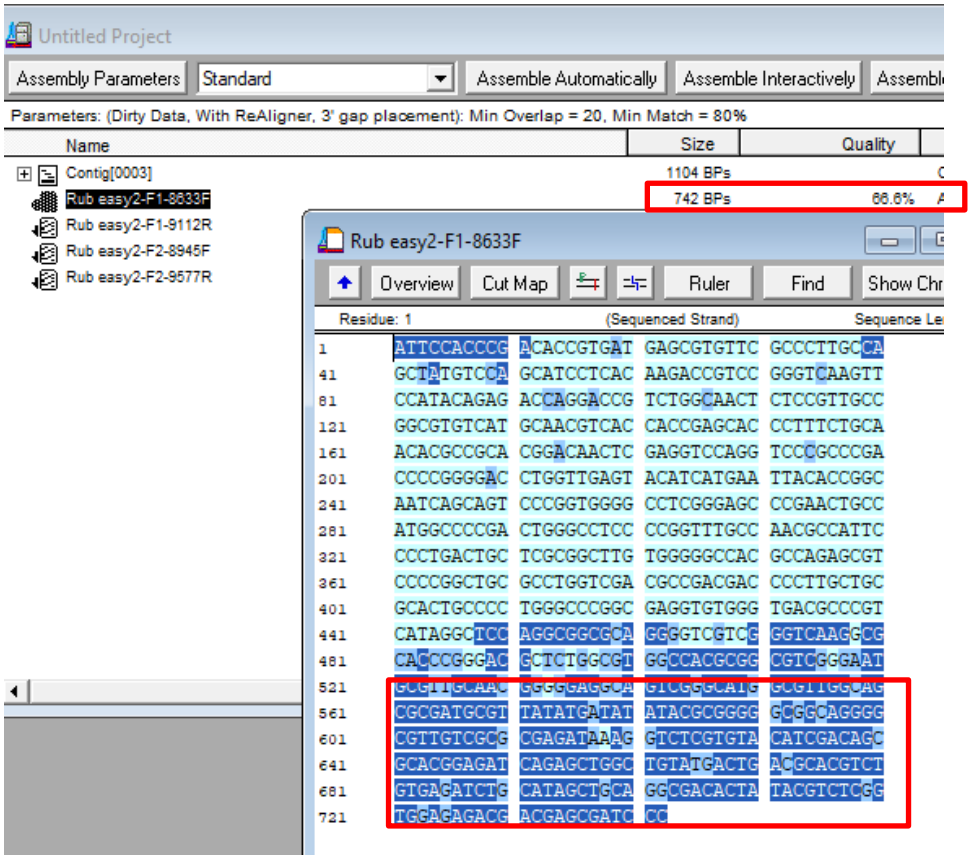
G. The guide sequence should always be in the forward orientation. The orientation of the abi files depends on the primer used, e.g. if MV214 was used, the orientation should be reverse, with MV216 it should be forward. If the orientation of the sequences is not correct, on the top toolbar, click **View** and then select **Reverse & Comp**, to change the orientation. Verify the orientation after each reassembly.



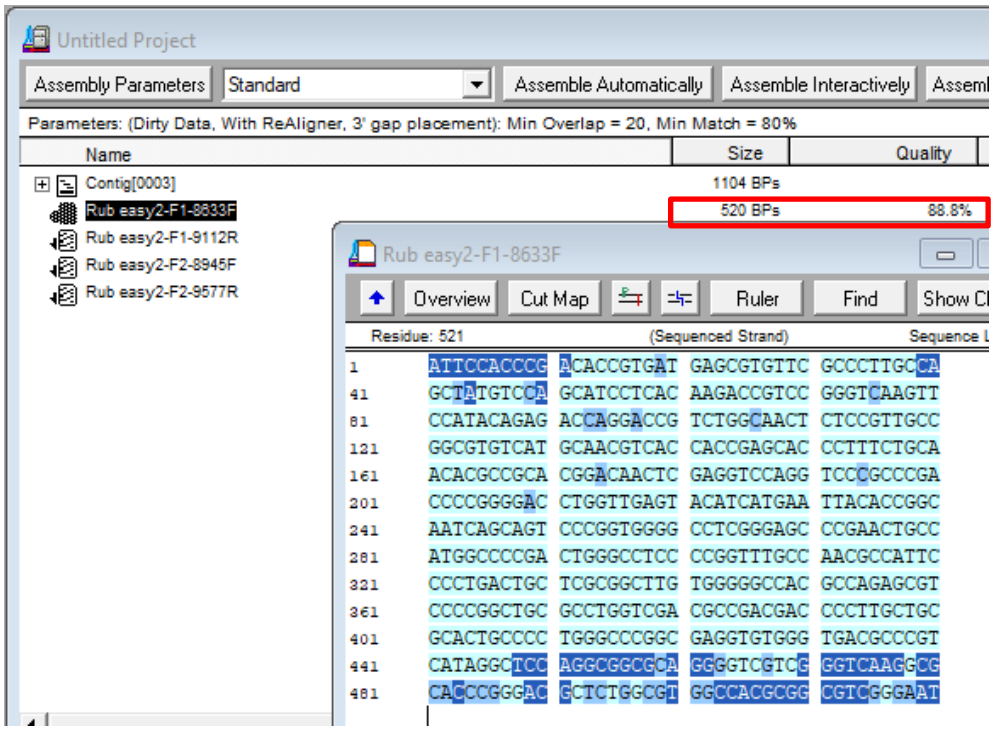
10. Cropping individual sequences

A. If a sequence is not assembling with the other sequences from a sample, double click on the box next to the sequence name. A new window containing the sequence bases will open.

B. Highlight poor-quality bases at the beginning and end of the sequence and click delete on keyboard.
Notice that after deletion of low-quality bases the overall quality of the chromatogram increases but the size of the sequence decreases.



Notice that after deletion of low-quality bases the overall quality of the chromatogram increases and the size decreases.

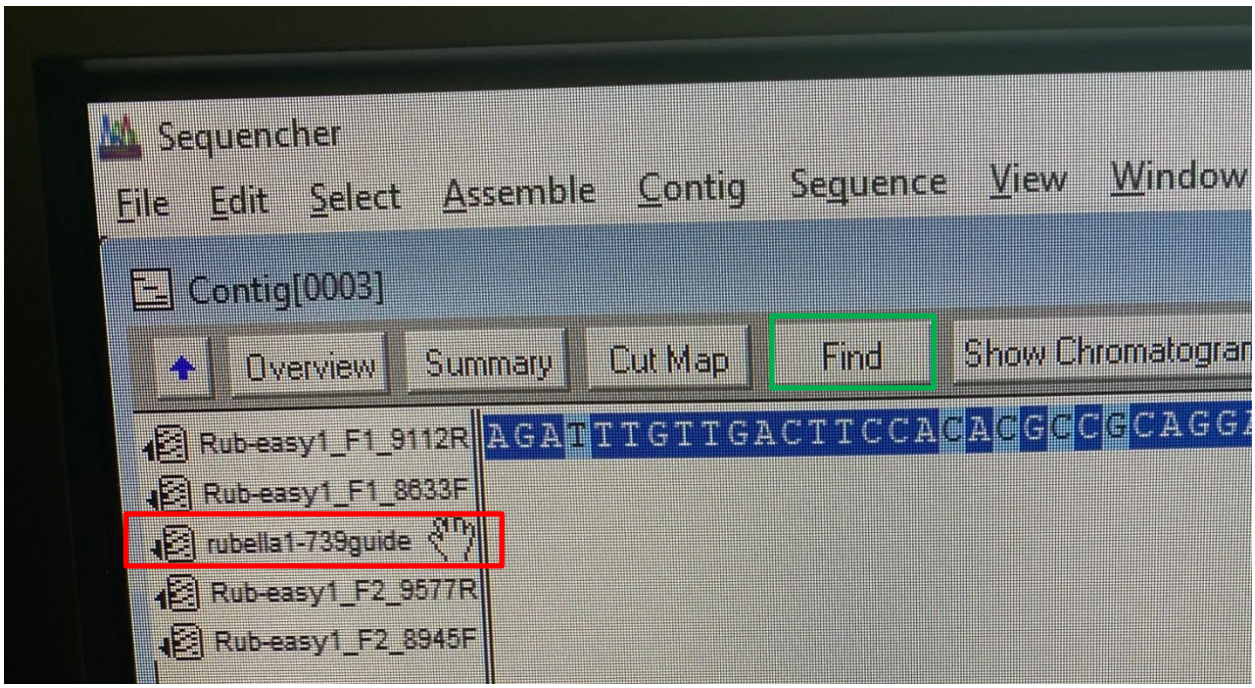


C. Close the window and try to assemble with contig by highlighting both the individual sequence and the assembled contig.

D. Chromatograms that are very short or of overall low quality may not assemble and should be deleted from the Sequencher file.

Cropping sequences in a contig

A. The order of sequences can be changed. Drag the guide sequence to the top of the list using the mouse (a little hand image will appear).

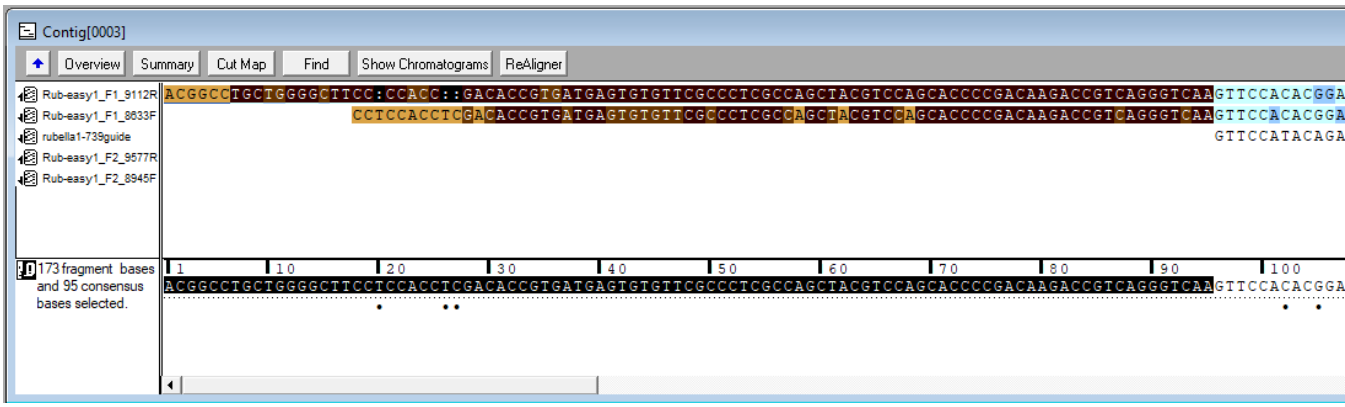


B. Find the beginning of the sequencing window using the guide sequence. Alternately, click **Find** and type the start sequence from the table below into the search window and click **Find**.

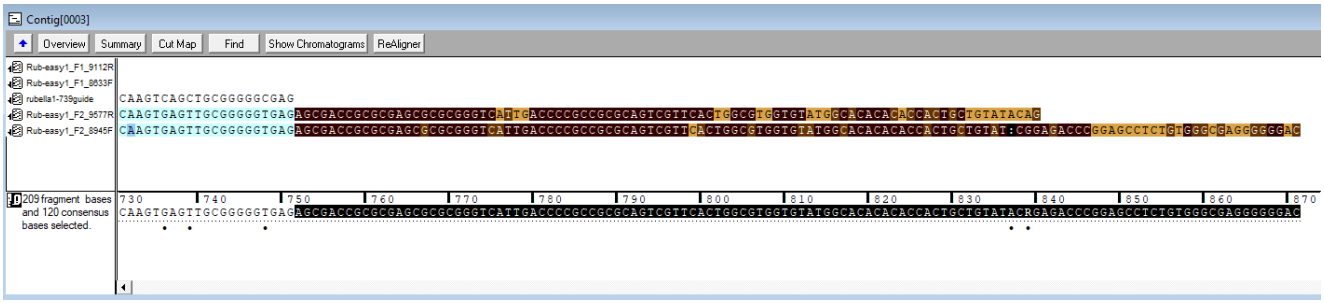
Virus	Region	Start	End	Length
Measles	N450	AAG**G**TCAGT	CTAGA**C**TAG	450
Rubella	E1	CAA**G**TTYC	GGGYGA**G**AGCG	739
Measles	MF-NCR	TTCTG**T**AG	TCAT**G**GG	1018*

MF-NCR length can vary because of insertions and deletions.

C. Once the beginning of the window is identified, delete all bases in all sequences that precede the beginning of the guide sequence or are before the **Bold** nucleotide in the search sequence. Highlight the nucleotides to be deleted in the consensus sequence and pressing the **delete** button on the keyboard.



D. Find the end of the sequencing window, using the guide sequence or sequence search information provided above, and delete all bases in all sequences that continue past the guide sequence or are after the **Bold** nucleotide in the search sequence. To do this, highlight the nucleotides to be deleted in the consensus sequence and press the **backspace** key on the keyboard. **Do not press the delete key, this will shift the sequence alignment.** If the delete key is used instead of backspace, go to **Edit** and click **Undo.**

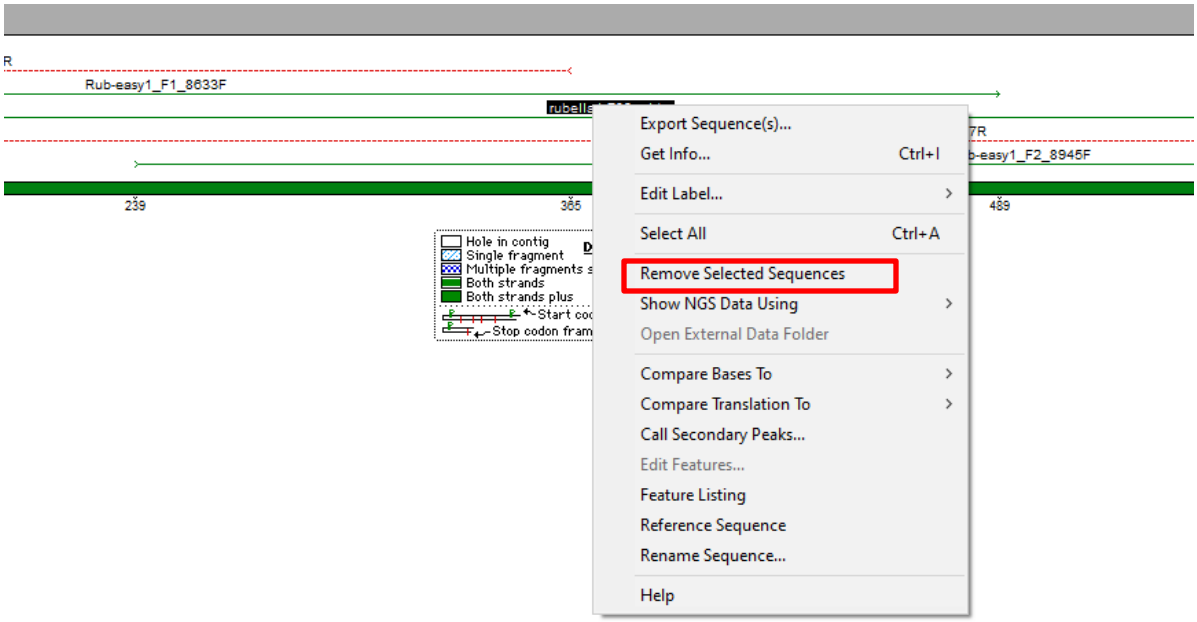


11. Removing the guide sequence and cropping overlapping regions

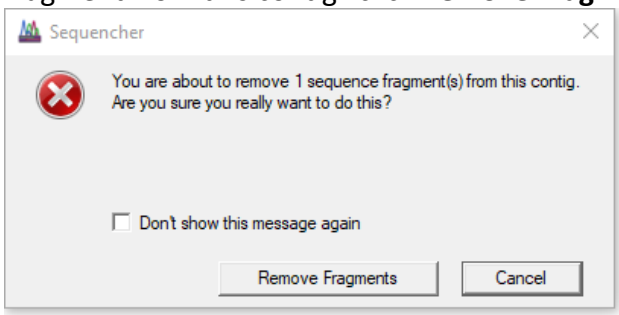
A. Once the sequences are cropped, remove the guide sequence. This can be done in Overview or Bases view.

B. Highlight the name of the guide sequence

C. Right click on the name and select **Remove Selected Sequences**

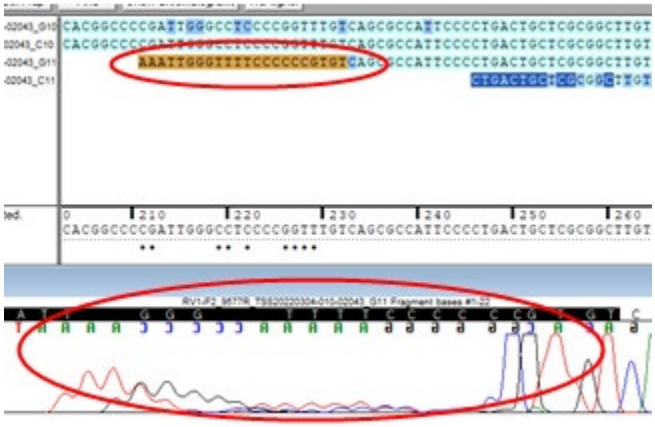


D. A warning window will open stating "You are about to remove 1 sequence fragment from this contig" click **Remove Fragments**.

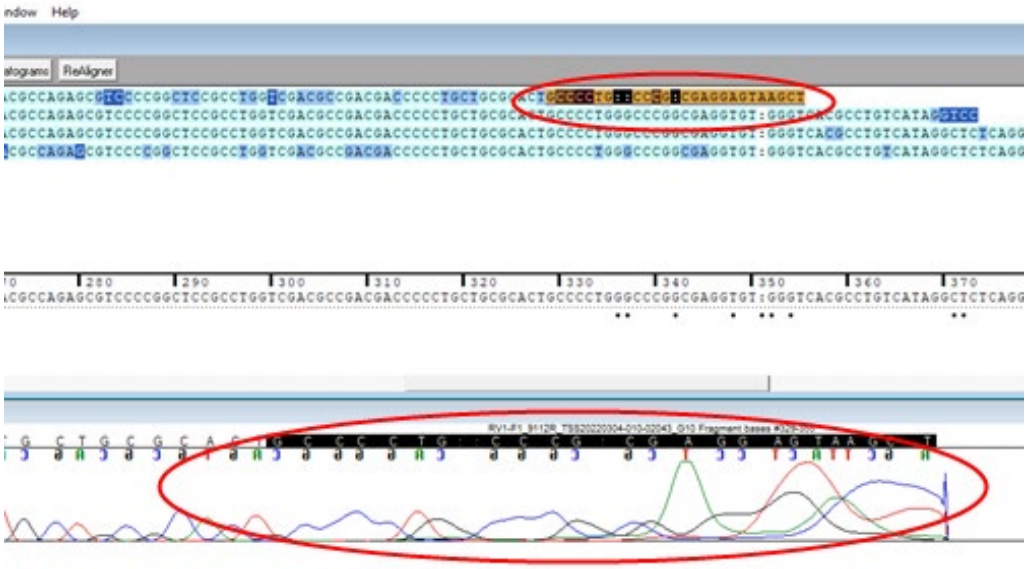


E. Removing low quality sequences within the sequencing window
Crop the beginnings and ends of poor-quality sequences for each sequence individually, not on the consensus sequence. This should not reduce the overall length of the alignment. Do this by first clicking on the **Show Chromatogram** button. Identify low-quality nucleotides, highlight the individual nucleotides in each sequence and use the **delete** button at the beginning and the **backspace** button at the end to remove them (see below).

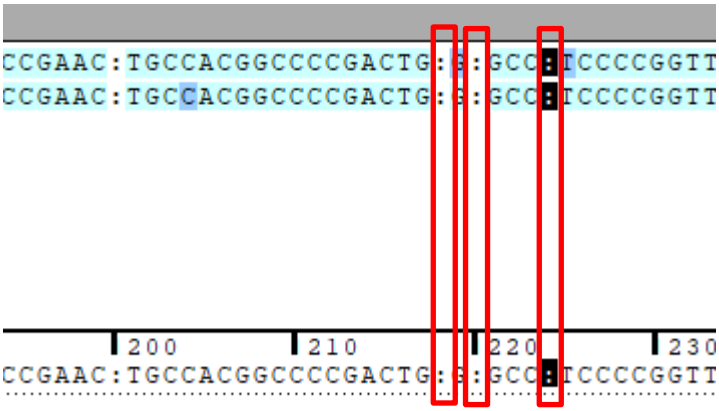
1. Original view with poor quality sequences in the overlapping region.



2. Highlighting and removing the ends of low-quality sequences.



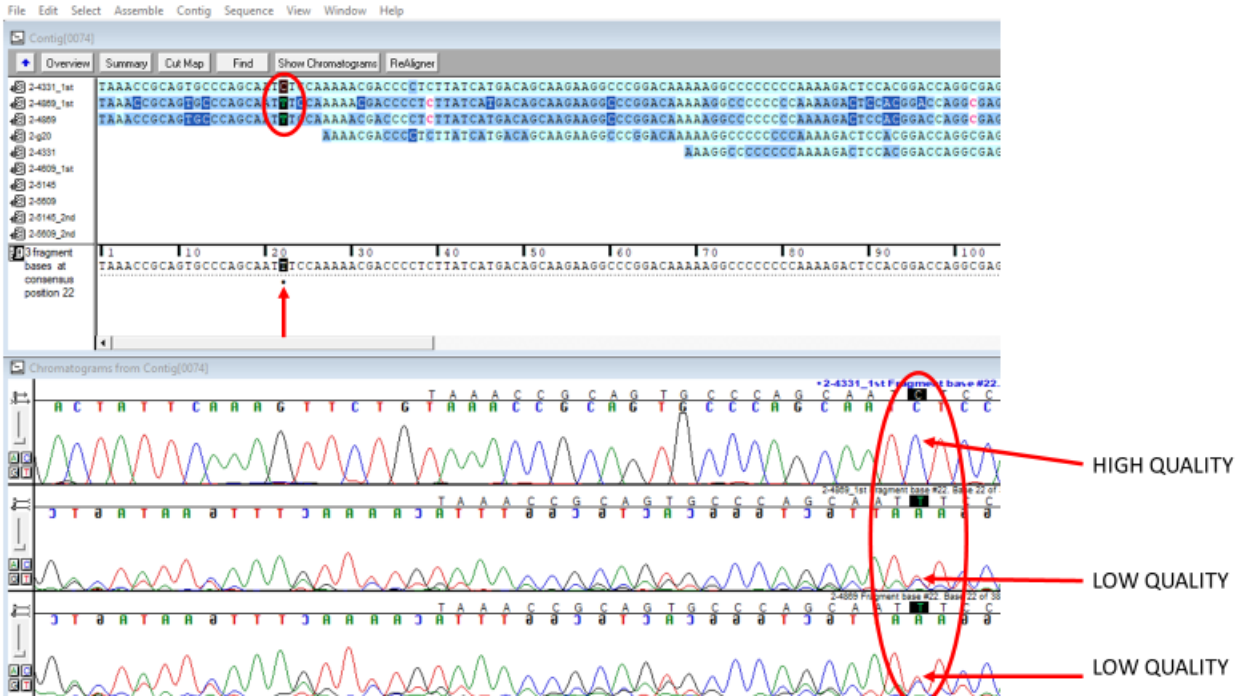
F. Remove any gaps the poor-quality sequences may have caused in these overlapping areas by highlighting the gaps in the <u>consensus sequence</u> and click **delete** on the keyboard. It is necessary to look at the chromatograms while doing this to avoid accidentally deleting a nucleotide. See below for how to open chromatograms. If there are many gaps, it may be better to re-assemble after removing the low-quality sequence as described above. Note: Verify the orientation of the contig after each reassembly.



12. Quality control of chromatograms

Choose the Bases view to examine confidence of base calls: Not all of the base calls will be of equivalent quality. The color shading of the nucleotides can be used as a general guide, but all chromatograms should be inspected, regardless of the color (see below). Sequencher confidence ranges: The lighter the color, the better the quality of the peak as determined by the software. Preset confidence ranges as described as percent quality are dark blue <20%, medium blue is between 20-40%, and light blue is >40% confident. These can be changed in the User Preference section of the software; however, it is strongly discouraged to make them too lenient.

A. Guidelines for handling discrepancies:
   a. If at least one chromatogram has a high-quality peak, the consensus can be edited to match the high-quality chromatogram (see below).
   b. Sequencing of samples may have to be repeated, depending on the overall quality of the sequencing results.
B. Examine chromatograms: Check the entire length of the cropped sequence for quality and disparities by clicking **Show Chromatograms**. The peak height can be adjusted with the slider bars on the left of the chromatogram. This does not change the quality of the chromatogram.
C. Check the consensus sequence for discrepancies: If the sequences vary from each other or if the program could not determine a base, a mark (●) will be displayed under the consensus sequence at this position. Ambiguous bases are labeled with N, R, K, Y, M, or S, depending on the mixed bases.
D. Click on the nucleotide in the consensus sequence to highlight the position in all aligned sequences. Compare the chromatograms.
E. Compare the mismatch position for all aligned sequences. If at least one sequence has a high-quality peak, the other sequences can be edited to match.



F. Click on the letter of the base you wish to change in the consensus sequence and type the correct letter. Close the chromatograms. The altered base will be highlighted in red.

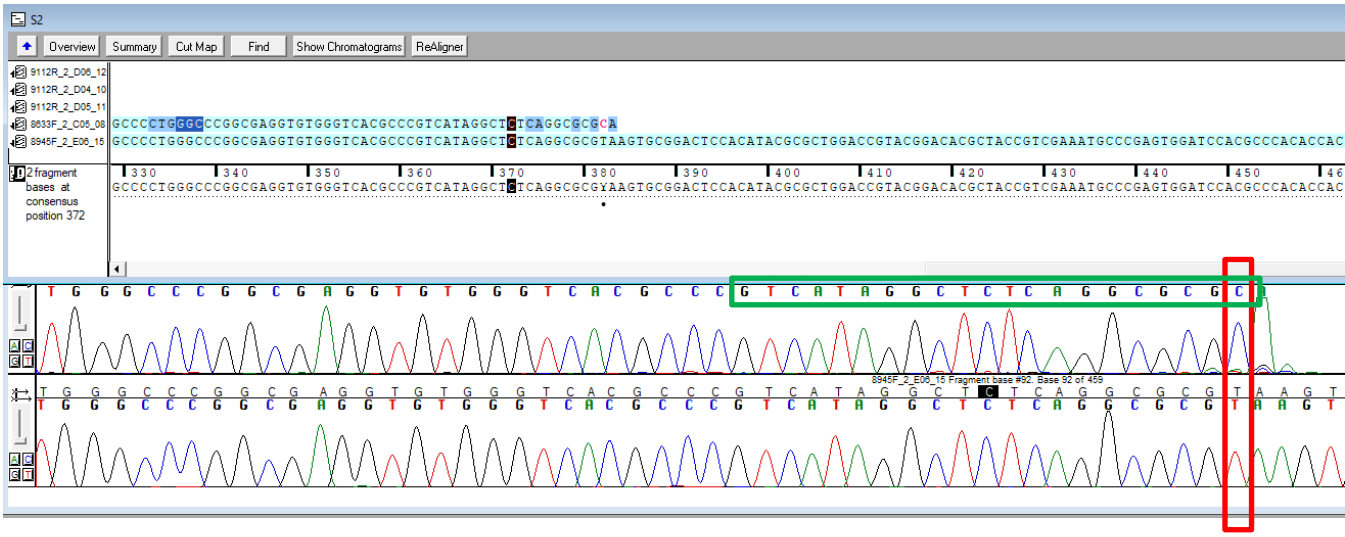
<u>Important</u>: Only alter a base if you can clearly determine it, based on the chromatograms. Do not alter a nucleotide based on the base call of the analysis program alone, the quality of the chromatogram must support it.

Note: Measles N-450 and rubella E1-739 sequences are always exactly 450 and 739 nucleotides long. Very rarely MF-NCR sequences may have indels. If high-quality chromatograms show evidence of indels, please discuss with GSL.

13. Special considerations for contigs composed of chromatograms from multiple PCR products

For the rubella genotyping window and the MF-NCR, the contig may consist of chromatograms from two PCR fragments. In the overlap between the two fragments, it is possible to find discrepancies that are caused by the PCR primer at the end of a fragment. Amplification can still occur if there are a few mismatches between the primer and the template, this results in PCR amplicons with the primer sequence at the end and not the 'original' sequence in the template. In this case, it is best to trust the chromatogram that is not close to the end of a fragment (provided the sequence quality is good). It is also useful to verify the presence of the primer sequence, and then delete it from the end of the chromatogram. To locate the primer sequence, enter the sequence into the **Find** window.

The example below shows the end of rubella fragment 1 (top sequence, 8633F primer), overlapping with fragment 2 (bottom sequence, 8945F primer). The highlighted nucleotide in position 382 is different in both chromatograms. However, the top nucleotide is part of primer 9112R. Therefore, the bottom sequence is the correct one.



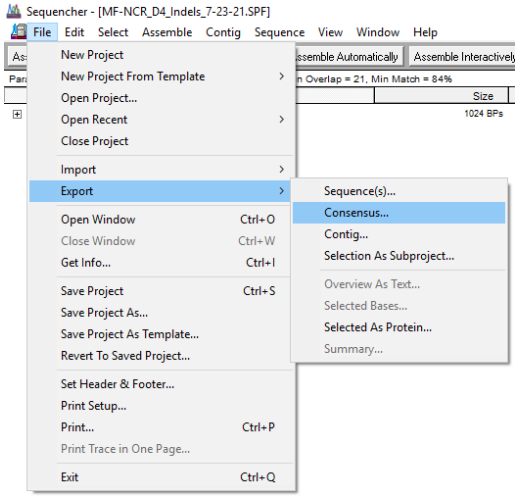
List of primers at the internal end of fragments

Amplicon	Primer name	Primer sequence	Reverse complemented sequence
RuVF1	9112R	GCG CGC CTG AGA GCC TAT GAC	GTC ATA GGC TCT CAG GCG CGC
RuVF2	8945F	TGG GCC TCC CCG GTT TG	
MF-NCR	G22F	CAC AAG CGA CCG AGG TGA CC	
MF-NCR	4801F	CAC AAG CGA CCG AGG TGA C	
MF-NCR	4869R	CTT GGC CCT TAG TTT TGT TTA G	CTA AAC AAA ACT AAG GGC CAA G
MF-NCR	4869R-H1	CTT GGC CCT CAG TTT TGT TTA G	CTA AAC AAA ACT GAG GGC CAA G
MF-NCR	5145R	GGT TGC CGT GGT CGT GTG TG	CAC ACA CGA CCA CGG CAA CC
MF-NCR	5145R-B3	GGT TGC CGT GGT GGT GTG TG	CAC ACA CCA CCA CGG CAA CC

14. SEQUENCHER: Exporting the consensus sequence

A. Highlight the edited contig, click on **contig>rename contig** or simply double click on the word contig to rename. Give the contig the name of the sample.

B. Click on **File>export>Consensus**. Under **format>FASTA**.



C. Click **export** to export, chose file location and save.

D. The edited Sequencher file can be saved for easy reference.

15. Sample Retention and Storage

All sequencing chromatogram files (.abi) and Sequencher files should be backed up onto the server or external drive. Do not rely on the sequencer computer or personal computer for storage of these files.

16. Reporting Results

Results should be reviewed by supervisor/SME prior to uploading to MeaNS/RubeNS databases. For the measles N450 and rubella E1-739, this should be done within 2 months of specimen receipt.

17. References

A. Bankamp, B, Byrd-Leotis, LA, Lopareva, EN, et al. *Improving molecular tools for global surveillance of measles virus*. J Clin Virol. 2013 Sep; 58(1): 176-82.
B. Penedos AR, Myers R, Hadef B, Aladin F, Brown KE (2015) *Assessment of the Utility of Whole Genome Sequencing of Measles Virus in the Characterisation of Outbreaks.* PLoS ONE 10(11): e0143081.
C. Thomas, Shari; Hiebert, Joanne, et al. *Measles Outbreak with Unique Virus Genotyping, Ontario, Canada, 2015*. EID. 2017 July; 23(7): 1063-1069.