# Suggestions for SARS-CoV-2 variant analysis of Sanger sequencing traces using SeqScape Software

Detecting and confirming mutations in newly arising SARS-CoV-2 strains is an important strategy for controlling the spread of new viral lineages. Recently we published four protocols that use Sanger sequencing to analyze SARS-CoV-2 isolates. Although the data contained in the sequence files can be compared against known sequences using a BLAST<sup>™</sup> search, in some cases a simpler and more definitive workflow is needed. In this protocol, we present options for SARS-CoV-2 variant analysis using Sanger sequencing files and Applied Biosystems<sup>™</sup> SeqScape<sup>™</sup> Software for secondary analysis. Although this workflow is not comprehensive, it is presented to help users begin to better understand their results. To take advantage of all the complete features of the software, please see the user guide.

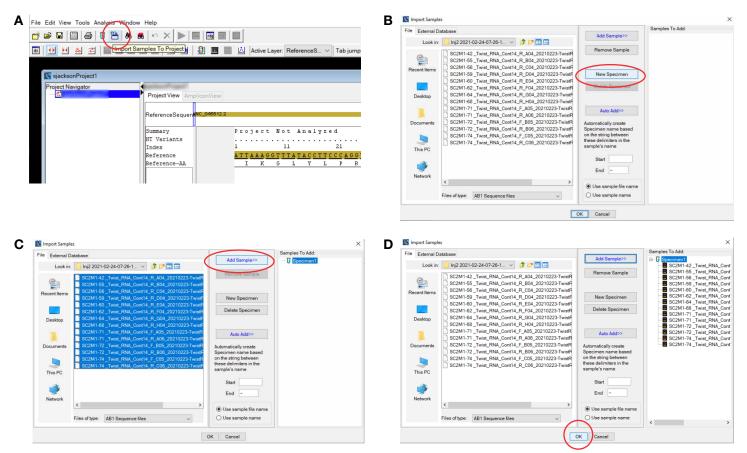
# 1. Import files and templates

File Edit View Tools					
📸 🖻 🗏 🎒					
New Project			Active Layer:	2	Tab jumps to nex
New Project Wizard					×
					×
New Project Wizard Select Reference Data Grou	p				×
Select Reference Data Grou	p Group you want to use for the new proj	ject.			×
Select Reference Data Grou Select the Reference Data	Group you want to use for the new proj	ject.			×
Select Reference Data Grou Select the Reference Data	Group you want to use for the new proj	ject.			×
Select Reference Data Grou Select the Reference Data	Group you want to use for the new proj	ject.		Brows	
Select Reference Data Grou Select the Reference Data	Group you want to use for the new proj ng Reference Data Group data:	ject.		Brows	
Select Reference Data Grou Select the Reference Data Use a GenBank file containin File Name:	Group you want to use for the new proj ng Reference Data Group data:	ject. Created By	Modified	Brows Modified By	
Select Reference Data Grou Select the Reference Data O Use a GenBank file containir File Name: Use an existing Reference D	Group you want to use for the new proj og Reference Data Group data: ata Group:	•	Modified 01 Feb 2021 at 13:32:30 PST		
Select Reference Data Grou Select the Reference Data () Use a GenBank file containin File Name: (@) Use an existing Reference D RDO Name 🖄	Group you want to use for the new proj ig Reference Data Group data: ata Group: Created	•			

A project template, DemoProject, can be downloaded from thermofisher.com/us/en/home/lifescience/sequencing/sanger-sequencing/applications/sars-cov-2-research.html. This template contains the SARS-CoV-2 reference sequence and known mutations in variants of concern. (A) Once the template is downloaded and SeqScape Software has been started, go to the Tools menu and select the SeqScape Software 4 Manager. Go to the Project Templates tab and click Import at the bottom of the window. Select the DemoProject.pt.ctf file and click Import. To load a project template in the SeqScape Software, select File -> New Project. (B) When the New Project window opens, select DemoProject from the list and enter a name for the project, and click New. The named project using the selected template will open.

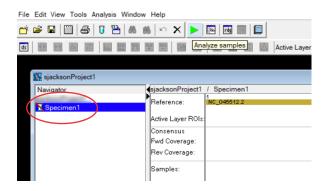


# 2. Import specimens, samples, and reference sequences to a project



In SeqScape Software, a specimen is a biological sample that could have many sequencing traces for different genomic regions assigned to it. A sample is a trace (.ab1 file) that was derived from a specimen. (A) To add specimens and samples, click the Import Samples to Project (red oval) button. (B) In the Import Samples window, click the New Specimen button (red oval). (C) A specimen (test tube) should appear in the Samples To Add window. The specimen can be renamed by double-clicking on the name. Once the specimen appears in the window, navigate to the location containing the appropriate set of .ab1 files. Highlight all the files for that specimen and choose Add Sample (red oval). Note that the software can create a specimen based on file names; for more information see the SeqScape Software 4.0 User Guide. (D) Once all files have been added, they should appear under the specimen name. Repeat for other specimens and samples as needed. Once all specimens and samples have been imported, click OK (red oval).

# 3. Analyze samples



The specimen is ready to be analyzed if the green arrow icon for Analyze samples is active. To analyze your specimen, click the green arrow. SeqScape Software will align all of the imported traces to the reference genome. The specimen icon in the Navigator window will have a red line across it (red oval) until the analysis is complete.

# 4. View results

File Edit View Tools Analysis Window H	lelp	
📸 🛋 🗐 😂 🗊 🔭 🕷 🕷	∽ X ▶	
	· · · ·	🚼 🔳 🔛 Active Layer: ReferenceS 🗸 Tab jumps to next Multiple 🗸
		The second secon
0		
Project Navigator	Project View Amp	
🖻 🚺 Specimen1	ridject fiew party	
Unassembled	ReferenceSequen	NC_0485122
SC2M1-64Twist_		
	Summary IT Variants	TGAAGGETTTTAATTGTTACTTTCCTTTACAATCATATGGTTTCCAACCCACTTATGGTGTTACCAACCA
- SC2M1-74Twist]	Index	1 23021 23031 23041 23051 23061 23071 23081 23091 23101 23111 2
	Reference	TG&AGGTITTAATTGTTACTTTCCTTTACAATCATATGGTTTCCAACCAA
- SC2M1-71Twist_	<ul> <li>Specimen1</li> </ul>	T GAAGGTTTTAATTGTTACTTTCCTTTACAATCATATGGTTTCCAACCCACTTATGGTGTTGGTTACGAACCATACGAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGT
SC2M1-62Twist_		
SC2M1-59Twist_		
SC2M1-42Twist_		
	< >	
	AT.	' son contain a library
	me active layer do	n nu cunan a nuary
<	.egend:	Known Variants All Variants
< >>	-	

When the analysis is complete, SeqScape Software will display all the aligned and unaligned (unassembled) traces in the leftmost Project Navigator panel. In the Project View panel, the reference sequence is shown (top), with red and blue vertical bars above it. The red bars highlight the positions of variants defined in the Reference Data Group (RDG) template; the blue bars highlight variants all variants, including those not defined in the RDG template.

# 5. View assembled sequences

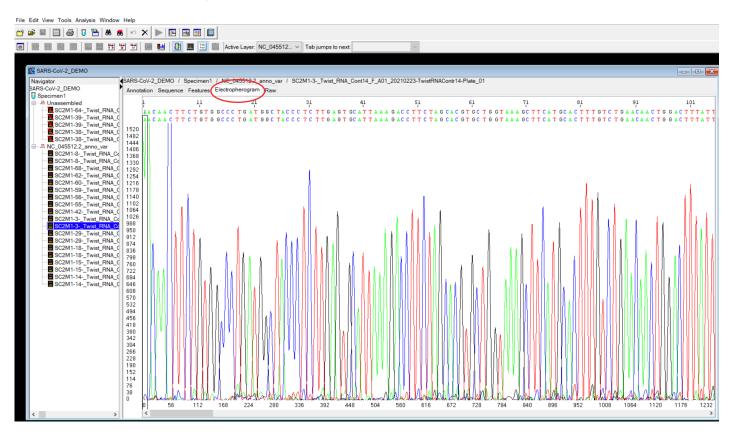
### Α File Edit View Tools Analysis Window Help NC 045512.2 Layout J CR-End 2345 22986 23299 24046 26462 27650 27710 23519 23874 24655 27004 28203 28203 BC2M1-60-578 614 543 554 62 C2M1-62-RN revers 610 543 554 494 SC2M1-68-\_Twist\_RN reverse SC2M1-71-\_Twist\_RN reverse SC2M1-71-\_Twist\_RN forward 0.0 -56-\_ -55-\_ -42-16065 16865 17665 18465 19265 20065 20865 21665 22474 23274 24074 24874 25674 26474 27274 28074 28875 sc... s,... s<u>c...</u> s<u>c...</u> sc.. SC2.... sc...

### File Edit View Tools Analysis Window Help

t Navigator	Specimen1 / NC_	045512.2								C
U Specimen1 ⊡- # Unassembled	Layout Assembly Sample Name Orientation	Ref.Start	Ref.End	CR-Length	CR-Start	CR-End	%N's	Sample Score	%Mixed Bases	
SC2M1-72Twist_RN		16065	16647	583	584	2	0.0	58	0.0	c
SC2M1-64Twist_RM		21421	21914	488	490	3	0.0	58	0.0	1
NC_045512.2	SC2M1-56- Twist BN reverse	21775	22345	568	568	1	0.0	58	0.0	-
SC2M1-74Twist_RM	SC2M1-59, Turist DN revenue	22986	23519	534	534	1	0.0	58	0.0	-
SC2M1-74Twist_RM SC2M1-72Twist_RM	0.001 // 0.0 7 1 0.01	23299	23874	576	578	3	0.0	43	0.0	-
SC2M1-72Twist_RM		24046	24655	610	614	5	0.0	44	0.0	-
SC2M1-71Twist_RM	SC2M1-68Twist_RN reverse	26462	27004	543	543	1	0.0	41	0.0	-
SC2M1-68Twist_RM	SC2M1-71Twist_RN reverse	27650	28203	554	554	1	0.0	59	0.0	-
SC2M1-62Twist_RM SC2M1-60Twist_RM		27710	28203	494	62	555	0.0	52	0.0	
									-	
										1
SC2M1-59Twist_RM SC2M1-56Twist_RM										1
SC2M1-59Twist_RM		23394 : t g t t c t t t a t	23404 C A G G G T G T T A A	23414 C T G C A C A G A A G	23424	23434	23444 ; a t c a a c t t a c	23454 234 TCCTACTTGGCG	464 2347	
SC2M1-59Twist_RM SC2M1-56Twist_RM SC2M1-55Twist_RM	23974 23984 TACTTCTAACCAGOTTGC TACTTCTAACCAGOTTGC 2444 2626 26 4 A CAGOTTGC	T G TT C TT T A T 124 T G TT C TT T A T	CAGGGTGTTAA	<u>стесасавале</u> стесасавале стесасавале	23424 <u>TCCCTGTTGC</u> TCCCTGTTGC	23434 <u>TATTCATGCAG</u> FATTCATGCAG	<u>атскасттас</u> атскасттас 74		64 234 TGTTTATTCTA TGTTTATTCTA TGTTTATTCTA 54	<u>a (</u>

(A) Highlighting the area with the aligned sequences gives details about how the traces align to the reference. The top of the window shows sequencing quality metrics for each of the traces. In the Layout tab (red oval), the position of each of the amplicons relative to the reference can be displayed. (B) In the Assembly tab (red oval), the sequences are shown. The highlighted position can be changed by sliding the box or moving the scroll bar at the bottom of the window.

# 6. View individual electropherograms



The quality of a sequencing run can be visualized by viewing the electropherograms. To see an analyzed electropherogram, highlight the sample listed on the left, under the specimen name. Choose the Electropherogram tab (red oval) at the top of the Results window.

# 7. Generate report

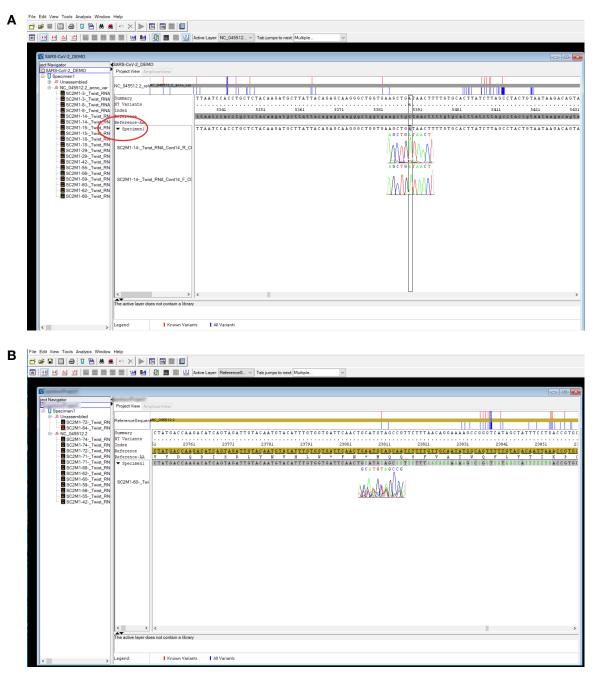
SARS-CoV-2_DEMO									
	SARS-CoV-2_DEMO / Specimen1 /	NC_045512.2_anno_	var						
II Specimen1	Layout Assembly	1	1	1	1		1	1.0	1
□ - <sup>24</sup> Unassembled	SC2M1-3Twist_RNA reverse SC2M1-3- Twist_RNA forward	826	1392	567	569	3	0.0	43	0.0
SC2M1-64Twist_RNA_	SC2M1-3Twist_RNA_torward SC2M1-8Twist_RNA_reverse	2931	3482	552	552	571	0.0	50	0.0
	SC2M1-8Twist_RNA forward	2932	3483	552	1	552	0.0	50	0.0
SC2M1-38Twist_RNA_	SC2M1-14- Twist BN reverse	5258	5810	553	561	9	0.0	39	0.0
SC2M1-38Twist_RNA_									
SC2M1-8Twist_RNA_C	826 836	846	856	866	876	886	896 90	5 916	926
SC2M1-68Twist_RNA_	TAACAACTTCTGTGGC	CCTGATGGCT	ACCCTCTTGAG	TOCATTAAAG	ACCTTCTAGCAC	GTGCTGGTAA	ACCTTCATCCAC	TTTGTCTGAAC	AACTGGACTTT
SC2M1-62Twist_RNA_ SC2M1-60Twist_RNA	TAACAACTTCTGTGGC	CCTGATGGCT	ACCCTCTTGAG	TOCATTAAAG	ACCTTCTAGCAC	GTGCTGGTAA	AGCTTCATGCAC	TTTGTCTGAAC	AACTGGACTTT
SC2M1-59Twist_RNA	569 559	549	539	529	519		499 48		469
SC2M1-56Twist_RNA_	AACAACTTCTGTGGC	CCTGATGGCT	ACCCTCTTGAG	TGCATTAAAG	ACCTTCTAGCAC	GTGCTGGTAA	AGCTTCATGCAC	TTTGTCTGAAC	<b>A A C T G G A C T T T</b>
	1 11	21	31	41	51	61		81 91	101
SC2M1-42TWISt_RNA_									
SC2M1-3- Twist RNA C									
SC2M1-29Twist_RNA									
SC2M1-29- Twist RNA									
SC2M1-18Twist_RNA									
SC2M1-18Twist_RNA									
SC2M1-15-Twist_RNA									
SC2M1-15Twist_RNA_									
SC2M1-14Twist_RNA_									
SC2M1-14Twist_RNA_									
	<								
	_egend: Clear Range	Known Variants		Algorithm Edited	Unknown Variant				

다 같은 문 [] 중 전 원 성 성 약 X 》 면 면 문 [] 전 문 문 전 법 것 것 것 면 번 번 및 문 문 실 Active Layer, NC,045512. ~ Tab jumps to next Mul

Reports										
Analysis QC Report					nmary					
Mutations Report	Active Layer	NC_045512.2_reference			Project			CoV-2_DEMO		
AA Variants Report	Project Creation Date	04 Mar 2021 at 12:59:33 PST			Project Modifio			2021 at 12:59:56 PST		
- D Specimen Statistics Report	Project Template (PT)	SARS-CoV-2_DEMO_Template			PT Creation Da			2021 at 12:59:33 PST		
Sequence Confirmation Report	PT Modification Date	04 Mar 2021 at 12:59:56 PST			Reference Data			0V2_NC_045512.2_0210		
Base Frequency Report	RDG Creation Date	10 Feb 2021 at 12:44:58 PST			RDG Modificati			2021 at 12:59:50 PST		
RDG Report	Display Settings (DS)	newProjectWizardSettings			DS Creation Da			2021 at 08:43:25 PST		
Audit Trail Report	DS Modification Date	04 Mar 2021 at 12:59:56 PST			Analysis Default			CoV-2_DEMO		
- Belectronic Signature History Repo	rt AD Creation Date	04 Mar 2021 at 12-50-22 PCT			<b>AD Modification</b>	Date	04 M 9	2021 # 12-50-58 PCT		
Genotyping Report	Specimen1									
				Muta	ations					
	- Specimen	Base Change	ROI	Muta		QV	Known	Effect As Change	Description	
Report Settings	Specimen Specimen1	Base Change 913crt	ROI NC_045512.2_anno_var			QV 43	Known yes	Effect As Change non-cod - ing	Description B.1.1.7 2_C913T:syn	
The content in this report:				Position 913	Length Type		yes	non-cod -		
The content in this report: is dependent on the Active Layer Project/Specimen level:	Specimen1	913o>t 3287o>t	NC_045512.2_anno_var	Position 913	Length Type 1 Sub	43	yes	non-cod - ing non-cod -	B.1.1.7 2_C913T:syn	
The content in this report: is dependent on the Active Layer Project/Specimen level: is independent of defined Layers	Specimen1 Specimen1	913o>t 3287o>t 913c>t	NC_045512.2_anno_var NC_045512.2_anno_var	Position 913 3287 5388	Length Type 1 Sub 1 Sub	43 48	yes yes yes	non-cod - ing non-cod - ing non-cod -	8.1.1.7 2_C013Tisyn 8.1.1.7 2_C3267TiT10011	
The content in this report. Is dependent on the Active Layer Project/Specimen level: is independent of defined Layers This report reflects information for the selected Consensus Segments:	Speciment Speciment Speciment Speciment Speciment	913eH 3207eH 933eH 6388eH 6666H	NC_046512.2_anno_var NC_046512.2_anno_var NC_046512.2_anno_var	Position 913 3287 5388 5986 6954	Length Type 1 Sub 1 Sub 1 Sub	43 48 43	yes yes yes	non-cod - ing non-cod - ing non-cod - ing non-cod -	B.1.17 2_C013T.syn B.1.17 2_C3207T.:10011 B.1.17 2_C5308A:A17081 B.1.17 2_C5308A:A17081 B.1.17 2_C5308A:A17081 B.1.17 2_T5954C:12230T	>
The content in this report. Is dependent on the Active Layer Project/Specimen level: is independent of defined Layers This report reflects information for the	Specimen1 Specimen1 Specimen1 Specimen1	913eH 3207eH 5308eH 5006eH	NC_045512.2_anno_var NC_045512.2_anno_var NC_045512.2_anno_var NC_045512.2_anno_var	Position 913 3267 5388 5996	Length Type 1 Sub 1 Sub 1 Sub 1 Sub 1 Sub 1 Sub	43 48 43 47	yes yes yes yes	nen-cod - ing - ing - nen-cod - ing - nen-cod - ing - nen-cod -	B.1.1.7 2_C913T:syn B.1.1.7 2_C3267T:T10011 B.1.1.7 2_C5386A:A17081 B.1.1.7 2_C5386A:A17081 B.1.1.7 2_C5996T:syn	
The content in this report. Is dependent on the Active Layer Project/Specimen level: is independent of defined Layers This report reflects information for the selected Consensus Segments:	Speciment Speciment Speciment Speciment Speciment	913eH 3207eH 933eH 6388eH 6666H	NC_048512.2_anno_var NC_049512.2_anno_var NC_049512.2_anno_var NC_049512.2_anno_var NC_049512.2_anno_var	Position 913 3287 5388 5986 6954	Length Type 1 Sub 1 Sub 1 Sub 1 Sub 1 Sub 1 Sub	43 48 43 47 40	yes yes yes yes	nen-cod - ing non-cod - ing non-cod - ing non-cod - ing nen-cod - ing nen-cod -	0.1.17 2_0013Tayn 0.1.17 2_02307T:11001 0.1.17 2_05380A:A17081 0.1.17 2_05680Tayn 0.1.17 2_056961:22307 0.1.17 2_4411238-112301:44150	

A report listing the nature of variants can be generated. (A) From any window, click on the Report Manager button (red oval). (B) Various reports, including quality control metrics and variants identified, can be seen in the resulting screen. To view variants, click on Mutations Report. A list of all variants identified in each specimen is generated. Note that variants that are defined in the RDG file are shown as Yes in the Known column; all other variants are shown as No in that column. The blue text in the Base Change column indicates that it contains a link to the variant's electropherogram and alignment to the reference sequence. Click on the base change to view the sequence trace. To print the Mutations Report, choose File -> Print. Alternatively, choose File -> Export -> Report and export the Mutations Report as a PDF file.

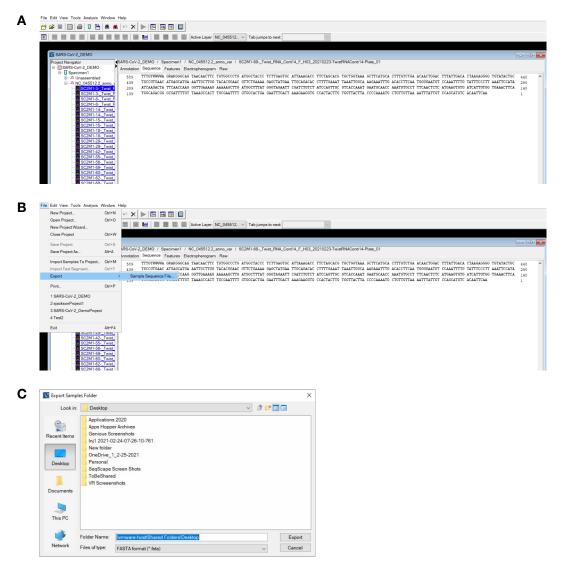
# 8. Examine variants



Clicking on blue text in the Base Change column of the report will display the relevant portion of the electropherogram aligned to the reference sequence. (A) Note that the Specimens triangle needs to be expanded (red oval) to view the electropherograms. This view can be used to quickly examine the quality of the variant definition. An example of an acceptable variant (5388C>A) is shown. (B) A variant that is inconclusive (and should be rejected) is displayed.



# 9. Mapping sequences to known strain lineages



Results can be exported into FASTA files for querying public databases. (A) To generate a FASTA file, select all the files to export in the Project Navigator window. (B) From the main menu, select File -> Export -> Sample Sequence File. (C) Specify the location for the files. A separate FASTA text file will be generated for each of the files selected. Sequences in these files can then be used to query strain lineage databases. One of these, the GISAID database (epicov.org/epi3/frontend#637476), provides BLAST search results against all known SARS-CoV-2 strains. Up to 10 FASTA sequences can be queried at once. Registration is required, but it is free to use. The Coronavirus Typing Tool (genomedetective.com/app/typingtool/cov) also accepts FASTA files. Single FASTA files can be queried at no charge, but batch uploading requires premium access.

# Find out more at thermofisher.com/sangercoronavirus

For Research Use Only. Not for use in diagnostic procedures. © 2021 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. BLAST is a trademark of the National Library of Medicine. COL34127 0421

