Sequencing Standards for SARS-CoV-2

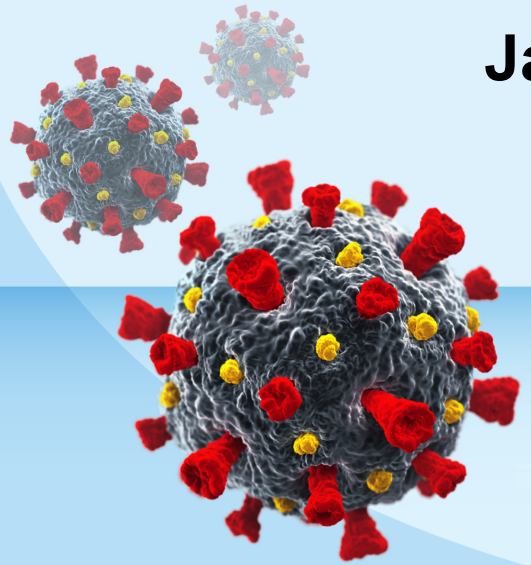
and

(the next viral pathogen?)

Discussion for Standards

Jack Collins, Jim Cherry, Ligia Pinto, Jayne Christen

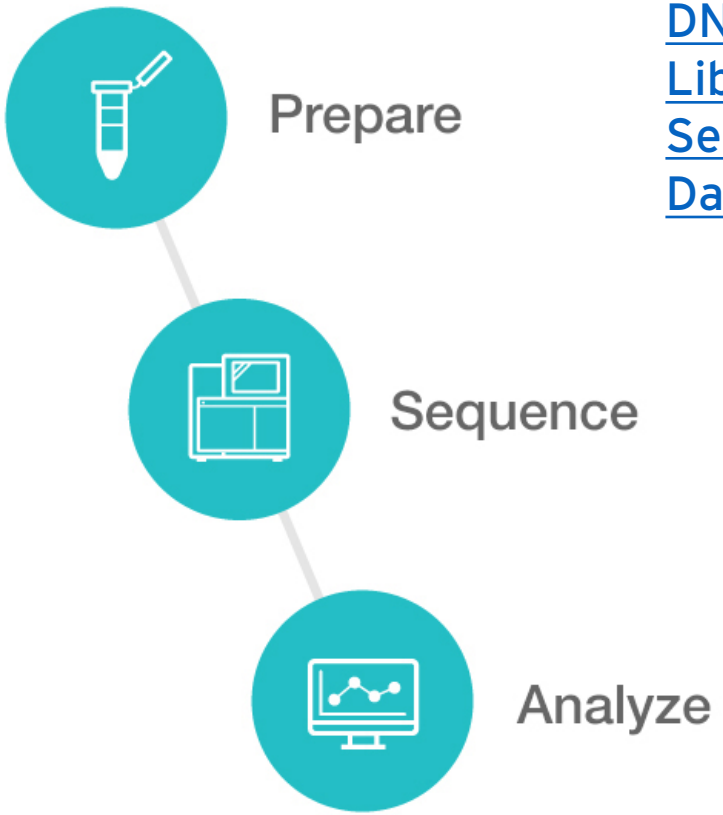
Frederick National Lab for Cancer Research



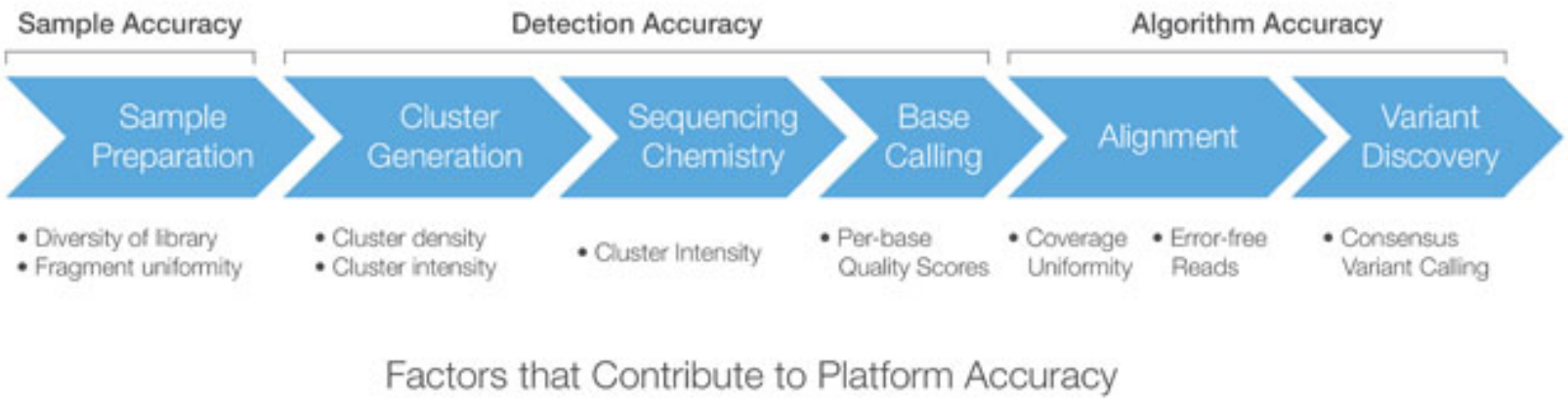
Clinical and Translational
Serology Task Force

From Bench to Bedside to Improve Public Health

Steps in the workflow that affect ‘accuracy’



[DNA Isolation]
[Library Prep]
[Sequencing]
[Data Analysis]



Factors that Contribute to Platform Accuracy

https://www.illumina.com/science/education/sequencing-workflow-accuracy.html

https://www.illumina.com/science/technology/next-generation-sequencing/beginners/ngs-workflow.html

Sequencing SARS-CoV-2

ASSOCIATION OF PUBLIC HEALTH LABORATORIES Recommendations

Recommendations for SARS-CoV-2 Sequence Data Quality & Reporting

Version 1 • March 1, 2021

APHL™

www.aphl.org

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of coronavirus disease 2019 (COVID-19), emerged in late 2019 in Wuhan, China.[1] Since being identified, SARS-CoV-2 has spread across the globe resulting in over 113 million cases of COVID-19 and 2.5 million deaths.[2] In the US, public health laboratories have played a critical role in the response to the COVID-19 pandemic. As the availability of molecular and serologic/antibody based testing diagnostics has expanded, so too has the work performed by public health laboratories.

Currently, efforts are underway to improve the coordination of genomic sequencing of SARS-CoV-2 in the US. Next generation sequencing (NGS) is a tool commonly used in public health surveillance and outbreak response. There are currently multiple efforts in public and private sectors to sequence SARS-CoV-2 genomes, but they are not yet well coordinated. This includes increasing the amount of sequencing performed in state and local public health laboratories. Currently, laboratories are using different sequencing strategies and protocols, including but not limited to differences in sample selection, library preparation and sequencing platforms, bioinformatics workflows, and data interpretation, which are leading to inconsistent data quality standards and biases among SARS-CoV-2 sequences generated by public health laboratories in public databases. While these issues are not unique to public health laboratories, providing basic guidance on SARS-CoV-2 sequencing and data sharing practices will improve coordination of laboratories that have been conducting sequencing and provide expectations for public health laboratories that are currently building capacity.

Recommendations

In order to provide clear expectations and a basic standardization across public health laboratories, APHL puts forward the following recommendations.

Methods

There are a number of methods and platforms currently available for SARS-CoV-2 sequencing. APHL has a compilation[3] of publicly available protocols, training resources and other information. The US Food and Drug Administration (FDA), in collaboration with other federal partners, has developed quality-controlled reference sequence data for the SARS-CoV-2 reference strain for the US.

SARS-CoV-2 Reference Sequence Data:

- **Reference sequence** from shotgun data under GenBank accession MT233526.1* or the original sequence from Wuhan MN908947.3**
- **Reference sequence** from target-capture data under GenBank accession MT246667.1

Data Quality and Sharing Parameters

Currently, efforts are underway to improve the coordination of genomic sequencing of SARS-CoV-2 in the US. Next generation sequencing (NGS) is a tool commonly used in public health surveillance and outbreak response. There are currently multiple efforts in public and private sectors to sequence SARS-CoV-2 genomes, but they are not yet well coordinated. This includes increasing the amount of sequencing performed in state and local public health laboratories. Currently, laboratories are using different sequencing strategies and protocols, including but not limited to differences in sample selection, library preparation and sequencing platforms, bioinformatics workflows, and data interpretation, which are leading to inconsistent data quality standards and biases among SARS-CoV-2 sequences generated by public health laboratories in public databases

https://www.aphl.org/programs/preparedness/Crisis-Management/Documents/APHL-SARS-CoV-2-Sequencing.pdf

Available SARS-CoV-2 Reference Strains

b|e|i RESOURCES

SUPPORTING INFECTIOUS DISEASE RESEARCH

Advanced Search Login

Home | Catalog | Deposits | Register | MR4 | About

Catalog | Search | Choose a Resource | Choose an Organism

Currently Available SARS-CoV-2 Materials

*Please see the Appendix section in the COA for each virus lot on variant analysis on the specific virus stock

BEI Number	Description	Lineage	GISAID Clade	GISAID ID	Clinical Information Available	Registration	AA Substitutions per GISAID*
Virus							
NR-54982	SARS-CoV-2 Isolate hCoV-19/Japan/TY7-503/2021	P.1. (or 20J/501Y.V3)	GR	EPI_ISL_792683 (P0) and EPI_ISL_877769 (passage 1 in Vero E6 /TMPSS2*) *contains a F184V mutation compared with sequence from original sample.	Isolated in airport quarantine station in Japan from a COVID-19 positive passenger from Brazil in January 2021.	BEI Level 3	Link to Mutations
NR-54011	SARS-CoV-2 Isolate hCoV-19/USA/CA_CDC_5574/2020	B.1.1.7	GR	EPI_ISL_751801	Isolated from a nasopharyngeal swab collected on December 29, 2020 in San Diego County, California, USA	BEI Level 3	Link to Mutations
NR-54009	SARS-CoV-2 Isolate hCoV-19/South Africa/KRISP-K005325/2020	B.1.351	GH	EPI_ISL_678615	Isolated from an oropharyngeal swab from a 40-year-old human male in Ugu district, KwaZulu-Natal, South Africa on November 16, 2020	BEI Level 3	Link to Mutations
NR-54008	SARS-CoV-2 hCoV-19/South Africa/KRISP-EC-K005321/2020	B.1.351	GH	EPI_ISL_678570	Isolated from an oropharyngeal swab from a 57-year-old human male in Harry Gwala district, KwaZulu-Natal, South Africa on November 15, 2020	BEI Level 3	Link to Mutations
NR-54000	SARS-CoV-2 hCoV-19/England/204820464/2020	B.1.1.7	GR	EPI_ISL_683466	Isolated from a 58-year-old human male on November 24, 2020 in England, United Kingdom.	BEI Level 3	Link to Mutations
NR-53953	SARS-CoV-2, Isolate hCoV-19/Denmark/DCGC-3024/2020 (also referred to as SARS-CoV-2/hu/DK/CL-5/1)	B.1.1.298	GR	EPI_ISL_616802	Isolated from a human who was exposed to a COVID-19 infected European mink (*Mustela lutreola*) in Northern Jutland, Denmark on October 5, 2020.	BEI Level 3	Link to Mutations
NR-53945	SARS-CoV-2, Isolate hCoV-19/Scotland/CVR2224/2020	B.1.222	G	EPI_ISL_448167	Isolated from a throat swab from a human patient diagnosed with COVID-19, on July 17, 2020 in Scotland, United Kingdom	BEI Level 3	Link to Mutations
NR-53944	SARS-CoV-2, Isolate hCoV-19/Scotland/CVR837/2020	B.1.5	G	EPI_ISL_461705	Isolated from a throat swab from a human patient diagnosed with COVID-19, on July 17, 2020 in Scotland, United Kingdom	BEI Level 3	Link to Mutations
NR-52281	SARS-CoV-2, Isolate USA-WA1/2020	A	S	EPI_ISL_404895	Male in 30s, returning traveler from Wuhan. Mild disease; recovered.	BEI Level 3	
	SARS-CoV-2, Isolate Hong				Isolated from a nasopharyngeal aspirate and		

Open Questions
(Not in a particular order)

- Optimal conditions/protocols to balance “speed, correct, cost”
- Protocol and workflow to fit purpose / scientific question
- Published standards so that data can be compared and harmonized
- Obtain high-quality data for SARS-CoV-2
- Be prepared for future viral (or bacterial?) pathogens