



NOV 22, 2022

WORKS FOR ME

1

## Introduction and Lineage Assignment of Assembled Sequences

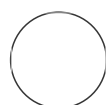
COMMENTS 0

**This protocol is published without a DOI.**

Paul Lorenzo A Gaite<sup>1</sup>,  
Dr Ritchie Mae T Gamot<sup>1</sup>

<sup>1</sup>PGC Mindanao

PHA4GE Subgrant - Philippines



phagesubgrantph

### ABSTRACT

As the SARS-CoV-2 pandemic emerged, Philippine Genome Center Mindanao (PGC Mindanao), in collaboration with Project Accessible Genomics and Genomic Epidemiology of COVID in the Philippines (GECO), has acquired an Oxford Nanopore MinION sequencer to be able to sequence SARS-CoV-2 whole genomes from different samples in Mindanao in a previous project. Specifically, the samples originated and were collected from actual patients in various Sub-National Laboratories (SNLs). A workflow was developed by PGC Mindanao to generate and identify SARS-CoV-2 whole genome sequences from these samples, up to submission of the sequences, along with the associated metadata, to the public database GISAID. The previous project was successful, generating and publicly-releasing about 100 sequences from these samples.

However, the previously-developed workflow, though functional, still requires a considerable amount of time to run since each step of the workflow would require much human intervention (e.g. generation of scripts and issuing of terminal commands) to start the next part of the workflow. In line with this, PGC Mindanao and Project Accessible Genomics has sought a collaborator that offers an automated workflow, which covers the input of raw sequencing data up to assembled sequences and automated report generation. This was the basis for applying under the Public Health Alliance for Genomics Epidemiology (PHA4GE) subgrant.

For this grant, PGC Mindanao collaborated with BugSeq due to the latter's capacity to automate the workflow and previous experience with submission to public databases. The collaboration with BugSeq, with their automated workflow, will definitely decrease runtime and improve bioinformatics capacity to generate immediate actionable insights from the data and ease submission to databases. In addition, the PHA4GE grant funds went to the improvement of the bioinformatics infrastructure within PGC Mindanao such as fiber optic structured cabling for the sequencers and internet connectivity and upgrade of bioinformatics workstations.

### PROTOCOL CITATION

Paul Lorenzo A Gaite, Dr Ritchie Mae T Gamot 2022. Introduction and Lineage Assignment of Assembled Sequences . **protocols.io**  
<https://protocols.io/view/introduction-and-lineage-assignment-of-assembled-s-cgfttttn>



————— This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Sep 12, 2022

Nov 22, 2022

69843

## 1 Abstract/Introduction

As the SARS-CoV-2 pandemic emerged, Philippine Genome Center Mindanao (PGC Mindanao), in collaboration with Project Accessible Genomics and Genomic Epidemiology of COVID in the Philippines (GECO), has acquired an Oxford Nanopore MinION sequencer to be able to sequence SARS-CoV-2 whole genomes from different samples in Mindanao in a previous project. Specifically, the samples originated and were collected from actual patients in various Sub-National Laboratories (SNLs). A workflow was developed by PGC Mindanao to generate and identify SARS-CoV-2 whole genome sequences from these samples, up to submission of the sequences, along with the associated metadata, to the public database GISAID. The previous project was successful, generating and publicly-releasing about 100 sequences from these samples.

However, the previously-developed workflow, though functional, still requires a considerable amount of time to run since each step of the workflow would require much human intervention (e.g. generation of scripts and issuing of terminal commands) to start the next part of the workflow. In line with this, PGC Mindanao and Project Accessible Genomics has sought a collaborator that offers an automated workflow, which covers the input of raw sequencing data up to assembled sequences and automated report generation. This was the basis for applying under the Public Health Alliance for Genomics Epidemiology (PHA4GE) subgrant.

For this grant, PGC Mindanao collaborated with BugSeq due to the latter's capacity to automate the workflow and previous experience with submission to public databases. The collaboration with BugSeq, with their automated workflow, will definitely decrease runtime and improve bioinformatics capacity to generate immediate actionable insights from the data and ease submission to databases. In addition, the PHA4GE grant funds went to the improvement of the bioinformatics infrastructure within PGC Mindanao such as fiber optic structured cabling for the sequencers and internet connectivity and upgrade of bioinformatics workstations.

The protocol below outlines the PGC Mindanao workflow (Section 2) for the lineage assignment

of assembled sequences, starting from patient sample collection (Section 2.1), to whole-genome sequencing of samples (Section 2.2), generation, quality control, and processing of raw sequence data (2.3), quality assessment and control of assembled sequence data (Section 2.4), then to the different approaches to annotation of the assembled sequence data such as by PANGOLIN (Section 2.5), Nextclade (Section 2.6), and GISAID (Section 2.7). The BugSeq workflow (Section 3) was also then outlined below, starting from FASTQ sequence file upload through the BugSeq website interface (Section 3.1), to quality control and processing of FASTQ data (Section 3.2), generation of results, output data (especially the assembled sequence FASTA files), and reports which also included PANGO lineage assignment (Section 3.3) for annotation of the assembled sequence data. Comparison of results from PGC Mindanao and BugSeq workflows (Section 4) was also done.

## 2 PGC Mindanao Workflow

Figure 1 illustrates the overview of the PGC Mindanao workflow.

The process starts with sample collection from the patient by swabbing, then subsequent processing and transport of the sample to PGC Mindanao. This was followed by pre-processing and whole-genome sequencing of the transported samples by the MinION sequencer and generation of the raw sequencing data. The raw sequencing data was then processed by bioinformatics tools to generate and identify assembled viral consensus sequences, which in turn were submitted, along with its corresponding metadata, to the public sequence database GISAID.

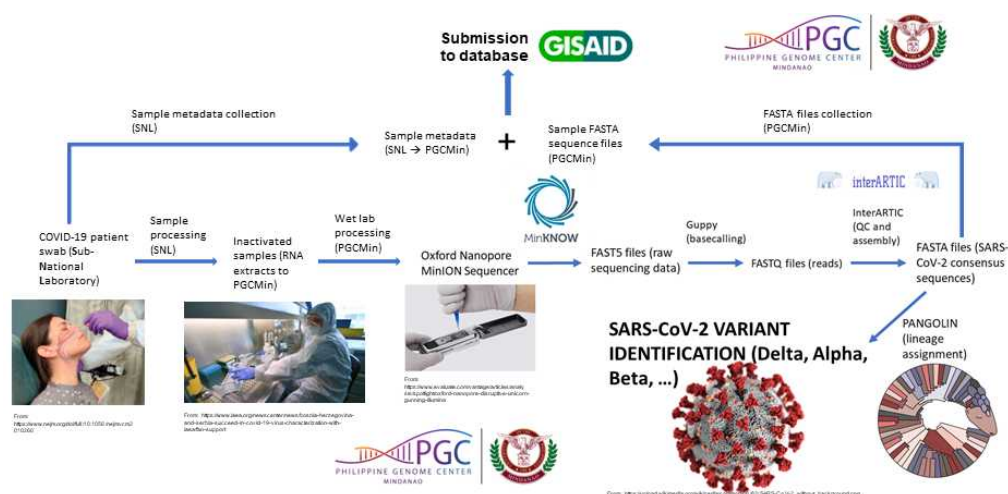


Figure 1. PGC Mindanao SARS-CoV-2 whole genome sequencing workflow.

The subsections below details the steps of the PGC Mindanao workflow.

## 2.1 Patient Sample Collection:

The SARS-CoV-2 viral samples were obtained on May 2021 to August 2021 from COVID-19 patients using the standard oropharyngeal/nasopharyngeal swabbing procedure by the Sub-National Laboratories (or SNLs) located in Cagayan de Oro City, Davao de Oro, Cotabato City, and Marawi City. The SNLs stored the patient swab samples in viral transport medium (VTM) and subsequently extracted for RNA. The resulting RNA extracts were securely transported by the SNLs to PGC Mindanao.

## 2.2 Whole-Genome Sequencing of Samples:

The RNA extracts were received and inspected by the PGC Mindanao wet lab team. The extracts were then processed for sequencing library preparation. The generated libraries were subsequently loaded into the [Oxford Nanopore MinION sequencer](#), which was plugged to a desktop computer. The sequencing was then run by the software [MinKNOW](#) installed within this computer. Sequencing commenced by using the MinKNOW software that was installed in the desktop computer, which also generated the raw sequence data.

Figure 2 shows the loading of the sequence library to the Nanopore MinION sequencing device.

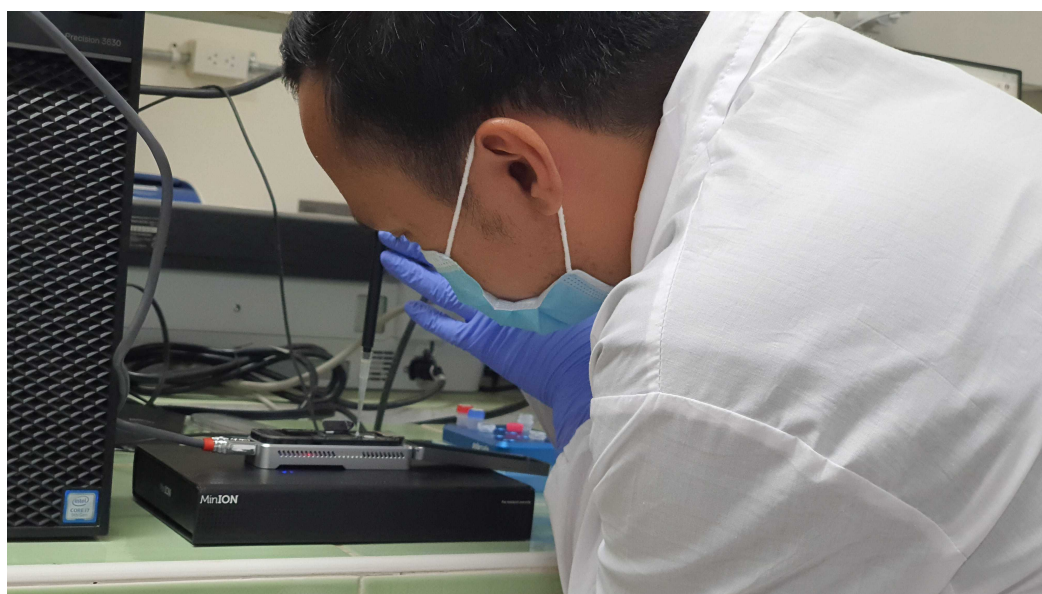


Figure 2. Loading of prepared sequence library to the Nanopore MinION sequencing device.

## 2.3 Generation, Quality Control, and Processing of Raw Sequence Data:

The raw sequence data that was generated by the MinKNOW software was in FAST5 file format. The generated FAST5 files were converted to FASTQ read files using the [Guppy basecaller](#) software. The FASTQ read files were inputted into the [interARTIC workflow](#).

The interARTIC workflow initially involves demultiplexing of FASTQ read files using by [Porechop](#), then quality trimming and filtering by [align\\_trim](#). The remaining reads were aligned to the SARS-CoV-2 reference genome by [minimap2](#), and subsequently assembled using [mpileup](#), which is one of the utilities found in [bcftools](#). Lastly, variant calling was then performed by [medaka](#).

Resulting consensus sequence files (FASTA file format), variant call files, and other intermediate output data files were generated by the workflow.

Figure 3 shows the computer workstation that was used during a Nanopore MinION sequencing run.

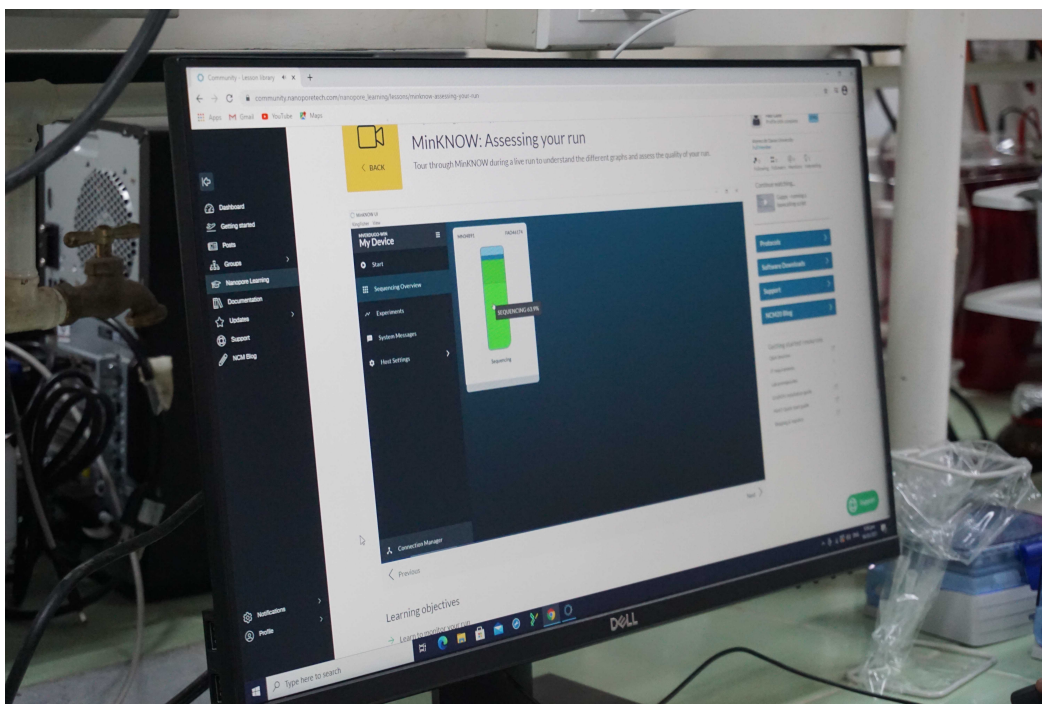


Figure 3. Workstation used during a Nanopore MinION sequencing run

## 2.4 Quality Assessment and Control of Assembled Sequence Data:

The resulting consensus sequences were subjected to quality assessment by submission to



[Nextclade Web](#), a bioinformatics tool that performs alignment, mutation calling, clade assignment, phylogenetic placement and quality control checks. This tool assesses various metrics of the sequence, including number of nucleotide gaps (N's), number of non-N ambiguous bases, and locations of N runs in the genome. Section 2.6 details the use of Nextclade Web.

Percent genome coverage and average genome depth were calculated using in-house scripts and [samtools](#), respectively. The results of the quality assessment were collected and collated in a spreadsheet using various in-house developed Linux BASH scripts. This information was used to perform quality control, such as filtering which sequences to submit to the GISAID database.

After the consensus sequence FASTA files have been generated, these are assigned to their corresponding relevant nomenclature designation. PGC Mindanao primarily employed the PANGO lineage assignment. Section 2.5 details the PANGOLIN workflow used.

## 2.5 PANGOLIN workflow:

The generated consensus sequences (FASTA file format) were uploaded to the [web version of PANGOLIN](#) to determine lineage assignment of the sequences.

Figure 4 shows the initial PANGOLIN web interface page. This contains the area where to upload the sequence(s) to be analyzed (e.g. lineage assignment). Figure 5, on the other hand, shows the results page of the PANGOLIN run on the sequenced samples, which mainly shows the lineage assignment per sequenced sample. Figure 6 shows the bar plot of counts of the variants according to PANGO lineages, as gathered from one of the sequencing runs previously performed by PGC Mindanao. In this sequencing run, it was shown that 10 samples were assigned to the B.1.1.7 lineage, 10 samples to the B.1.351 lineage, and 2 samples to the B.1.1.28 lineage.

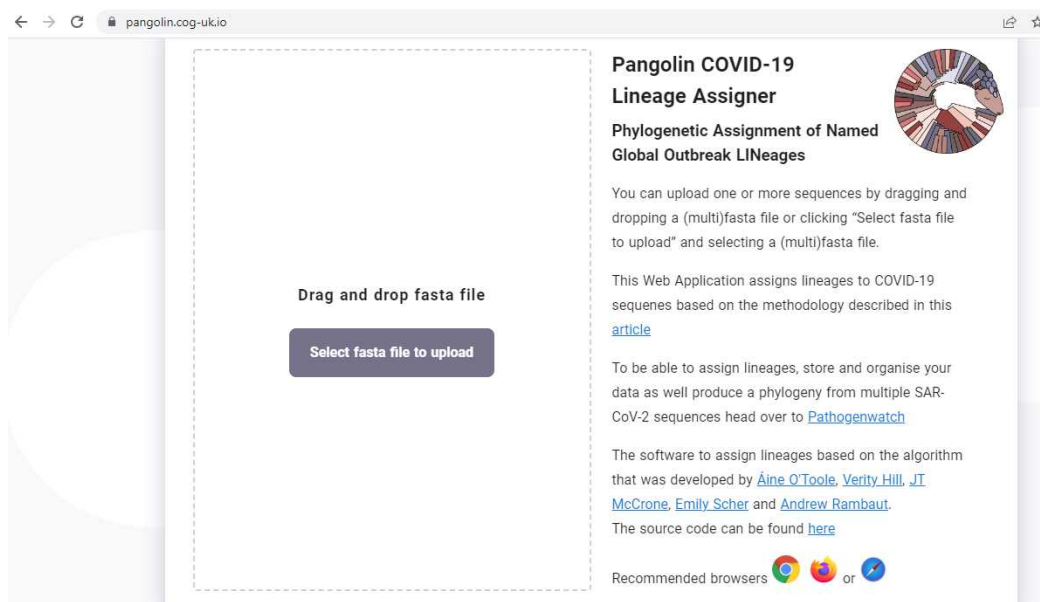


Figure 4 . Initial PANGOLIN web interface page

<a href="#">Retry Failed Sequences</a> <a href="#">Reset entries</a> <a href="#">Upload another file</a> <a href="#">Help</a>			
File name	Sequence name	Lineage	Assignment Conflict
+ FAILED (Click warning icon for more info) 1 sequence			
- ANALYSED (Click tick icon for more info) 22 sequences			
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample02_ARTIC/medaka MN908947.3	B.1.351	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample03_ARTIC/medaka MN908947.3	B.1.1.7	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample04_ARTIC/medaka MN908947.3	B.1.351	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample05_ARTIC/medaka MN908947.3	B.1.1.7	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample06_ARTIC/medaka MN908947.3	B.1.1.7	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample07_ARTIC/medaka MN908947.3	B.1.1.7	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample08_ARTIC/medaka MN908947.3	B.1.1.7	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample09_ARTIC/medaka MN908947.3	B.1.1.7	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample10_ARTIC/medaka MN908947.3	B.1.1.7	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample11_ARTIC/medaka MN908947.3	B.1.1.7	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample12_ARTIC/medaka MN908947.3	B.1.351	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample13_ARTIC/medaka MN908947.3	B.1.1.7	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample14_ARTIC/medaka MN908947.3	B.1.351	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample15_ARTIC/medaka MN908947.3	B.1.351	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample16_ARTIC/medaka MN908947.3	B.1.1.7	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample17_ARTIC/medaka MN908947.3	B.1.1.28	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample18_ARTIC/medaka MN908947.3	B.1.1.28	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample19_ARTIC/medaka MN908947.3	B.1.351	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample20_ARTIC/medaka MN908947.3	B.1.351	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample21_ARTIC/medaka MN908947.3	B.1.351	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample22_ARTIC/medaka MN908947.3	B.1.351	0.0
✓ all_seqc_clinbatch2_final.fasta	SARS_CoV_2_clinical_batch_2_sample23_ARTIC/medaka MN908947.3	B.1.351	0.0

Figure 5. Results page of the PANGOLIN run on sequenced samples

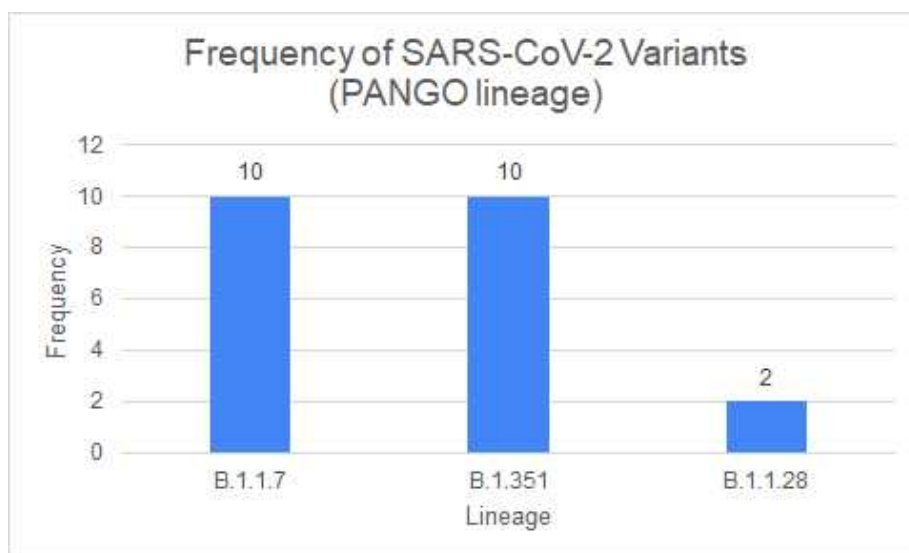


Figure 6. Frequency of SARS-CoV-2 variants from a sequencing run performed by PGC Mindanao, by PANGO lineage

## 2.6 Nextclade workflow:

[Nextclade Web](#) was used to assign sequences to a nomenclature designation. The sequences were uploaded to their online server for analysis and subsequent assignment.

Figure 7 shows the initial Nextclade web interface page. This contains the area where to upload the sequence(s) to be analyzed. Figure 8 shows the results page of the PANGOLIN run on the sequenced samples, which mainly shows the quality assessment metrics per sequenced sample, including mutation data, number of Ns, and number of frameshifts. Figure 9 shows the bar plot of counts of the variants according to Nextstrain clades, as gathered from one of the sequencing runs previously performed by PGC Mindanao. In this sequencing run, it was shown that 10 samples were assigned to the 20I clade (Alpha, V1), 10 samples to the 20H clade (Beta, V2), and 2 samples to the 20B clade.



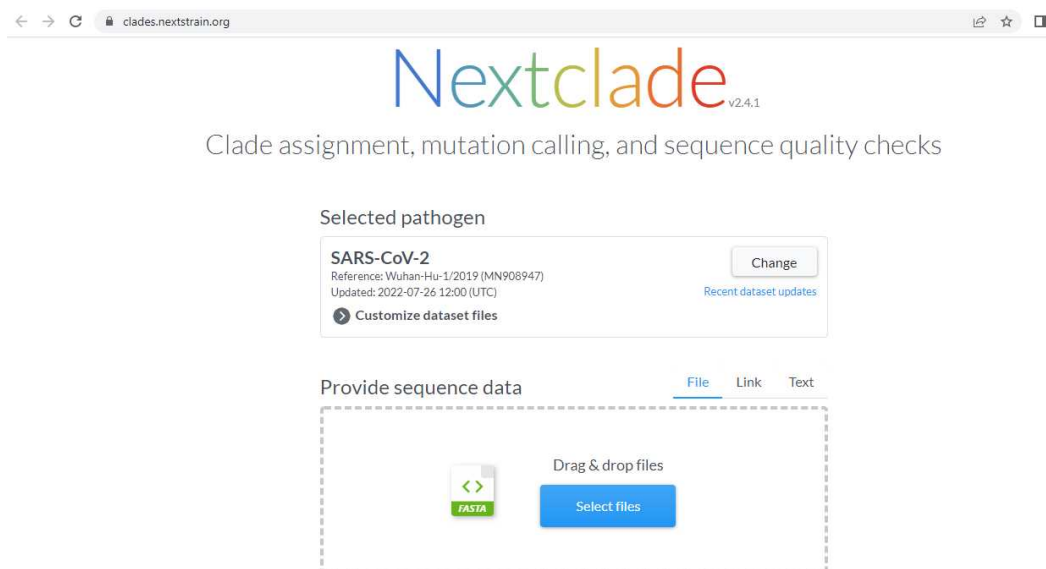


Figure 7. Initial Nextclade web interface page

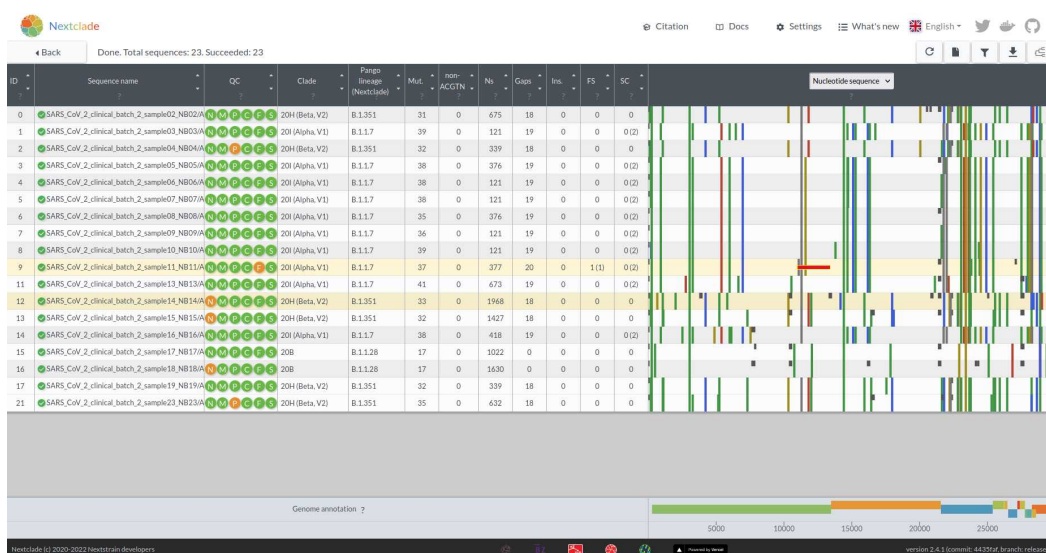


Figure 8. Results page of the Nextclade run on sequenced samples

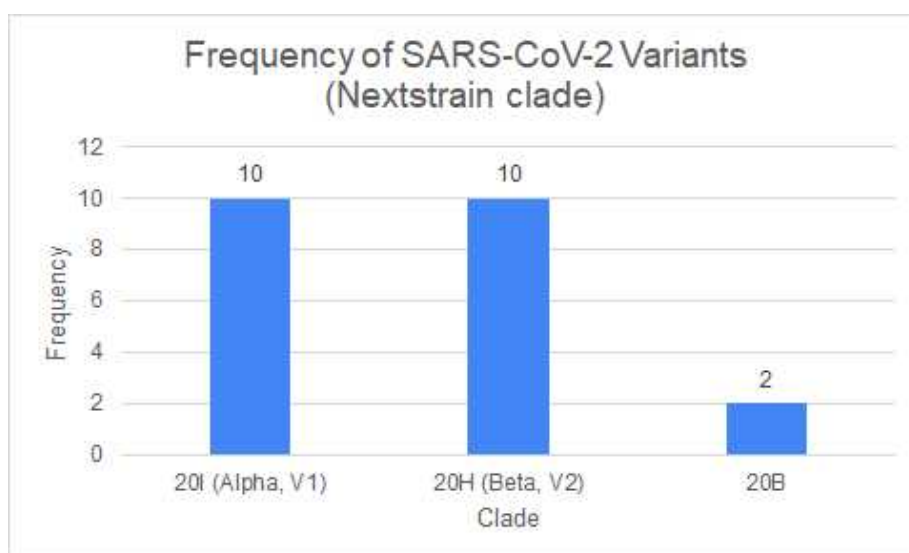


Figure 9. Frequency of SARS-CoV-2 variants from a sequencing run performed by PGC Mindanao, by Nextstrain clade

## 2.7 GISAID workflow:

The project previously entailed submission of sequences to the [GISAID database](#). Details of the submission of sequences and associated metadata to GISAID can be found in this [link/protocol/section](#). Figure 10 shows the landing page for the GISAID EpiCoV web interface. When the submitted sequences are approved for release to the public GISAID databases, these are automatically assigned to the relevant designation according to GISAID. Figure 11 shows the search page where released sequences may be viewed. Figure 12 shows an example of a GISAID entry containing the details of a sample, its sequence, and other associated metadata.

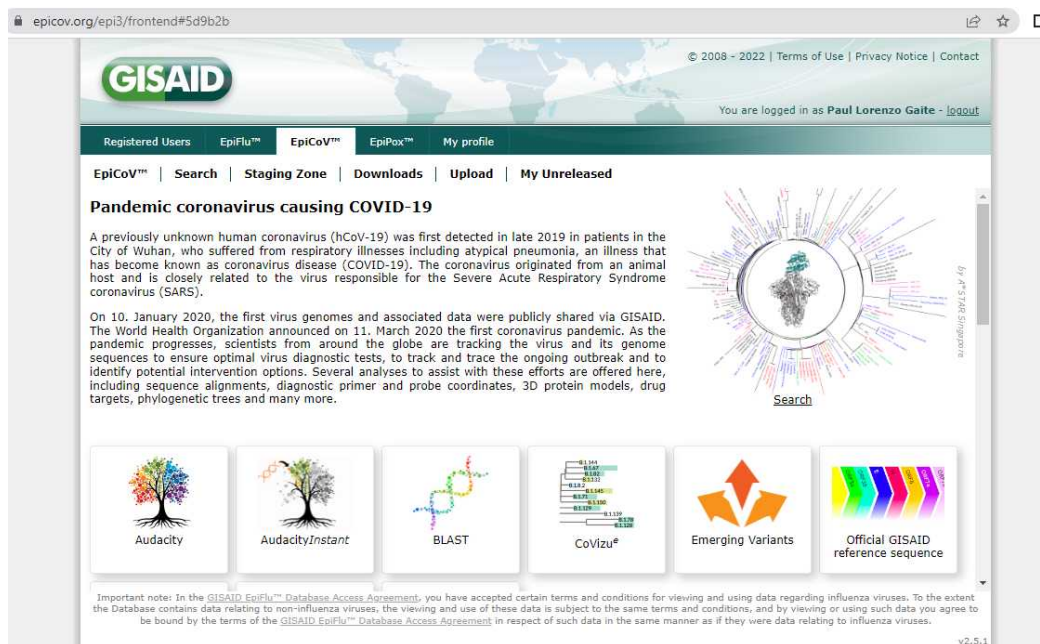


Figure 10. Initial GISAID EpiCoV web interface

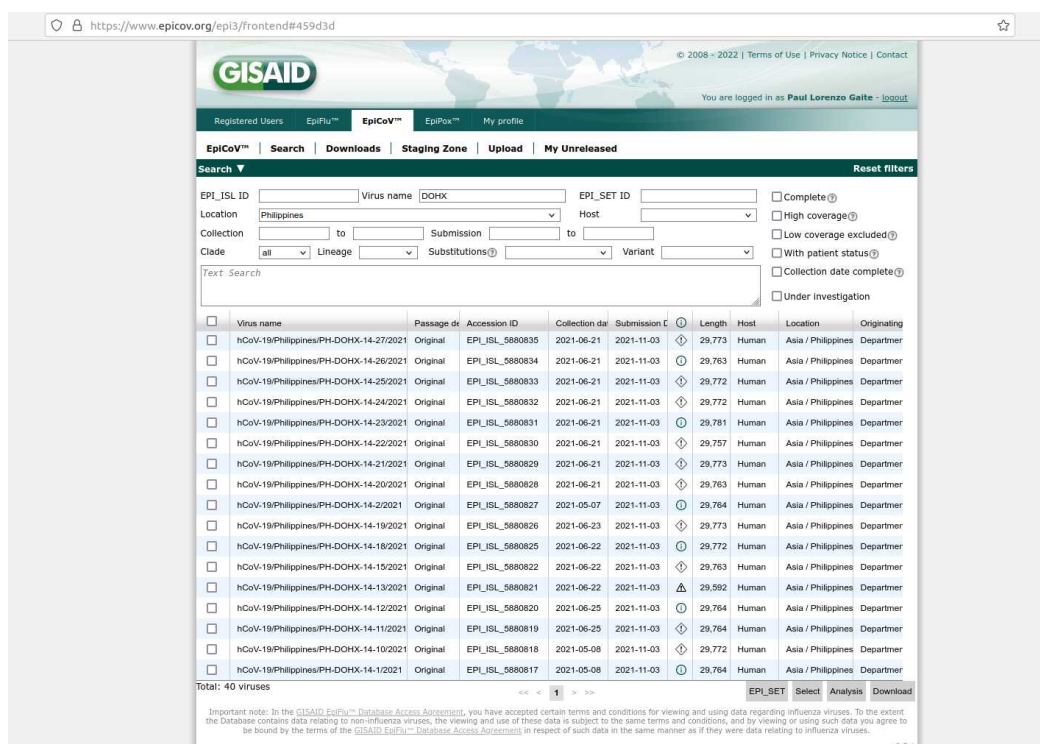


Figure 11. Search page for GISAID entries

Virus detail	
Virus name:	hCoV-19/Philippines/PH-APMC-16-11/2021
Accession ID:	EPI_ISL_10581881
Type:	betacoronavirus
Clade:	GH
Pango Lineage:	B.1.351 (Pango v.4.1.3 PLEARN-v1.15.1), Beta (B.1.351-like) (Scorpio)
AA Substitutions:	Spike A27S, Spike A243del, Spike A701V, Spike D80A, Spike D215G, Spike D614G, Spike E484K, Spike K417N, Spike L242del, Spike L244del, Spike N501Y, Spike N658Y, E P71L, N T205I, NS3 Q57H, NS3 S171L, NSP2 T85I, NSP3 K837N, NSP5 K90R, NSP6 F108del, NSP6 G107del, NSP6 S106del, NSP12 L829F, NSP12 P323L, NSP13 V98F
Variant:	VOC Beta GH/501Y.V2 (B.1.351+B.1.351.2+B.1.351.3) first detected in South Africa
Passage details/history:	Original
Sample information	
Collection date:	2021-08-13
Location:	Asia / Philippines / Region X (Northern Mindanao)
Host:	Human
Additional location information:	
Gender:	Female
Patient age:	53
Patient status:	hospitalized
Specimen source:	Oropharyngeal / Nasopharyngeal swab
Additional host information:	
Sampling strategy:	
Outbreak:	
Last vaccinated:	
Treatment:	
Sequencing technology:	Nanopore MinION Mk1B
Assembly method:	interARTIC
Coverage:	309x
Comment:	Gap of 18 nucleotides when compared to the reference WIV04 sequence.
Institute information	
Originating lab:	Amai Pakpak Medical Center
Address:	Amai Pakpak Avenue, Marawi City, 9700 Lanao del Sur
Sample ID given by the originating laboratory:	
Submitting lab:	Philippine Genome Center Mindanao
Address:	Tugbok District, Brgy. Mintal, Davao City, Philippines
Sample ID given by the submitting laboratory:	
Authors:	Lyre Anni E. Murao, Michael G. Bacus, Joan T. Acaso, April Mae Numeron, Ritchie Mae T. Gamot, Giovanna Fae R. Oguis, Paul Lorenzo A. Gaité, Kahlil Corazo, Ineke Knot, Homer Sajonia II, Norlaila Ampay, Mustafah M. Maradial, Aslia Mira-ato Mapandi
Submitter information	
Submitter:	Gaité, Paul Lorenzo
Submission Date:	2022-03-02
Address:	

Figure 12. Screenshot of the metadata page from a public GISAID entry

### 3 BugSeq Workflow

[BugSeq](#) is a cloud-based bioinformatics workflow that is capable of pre-processing and assembling raw sequence data, thereby generating whole genome sequences, such as from SARS-CoV-2 sequencing. Briefly, it is a website that allows upload of raw sequencing data (either in FASTQ or FAST5 file format). The workflow then performs quality control, mapping, assembly, and variant calling on the uploaded data. Reports are ultimately generated automatically by the workflow.

### 3.1 FASTQ Sequence File Upload thru BugSeq Interface:

The workflow can be accessed through a web browser, such as Google Chrome or Mozilla Firefox. Users must first register for an account before being able to use the actual workflow. The account will grant access to a webpage making up the front-end of the BugSeq workflow (Figure 13).

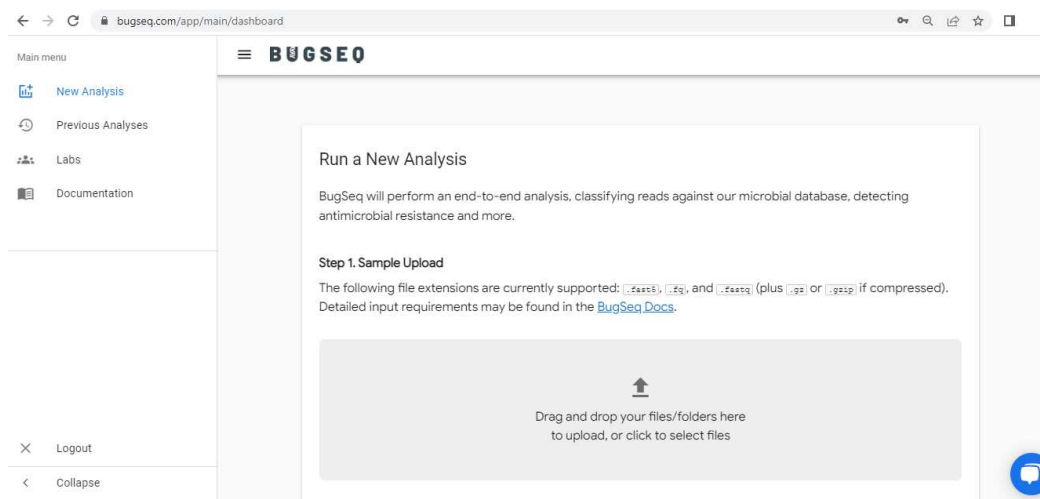


Figure 13. Landing page after logging in to BugSeq account

Users must input parameters such as sequencing platform used and descriptions of the sample. The users would upload the FASTQ (or FAST5) files in the interface.

### 3.2 Quality Control and Processing of FASTQ Data:

The FASTQ read files are processed at the backend of the workflow. Briefly, the read files were demultiplexed by [QCat](#), followed by quality evaluation by [QUAST](#) and [FastQC](#). The reads were then adapter and quality trimmed and filtered by [fastp](#). The reads were subsequently aligned to the SARS-CoV-2 reference genome by [minimap2](#), and then assembled. The assemblies were then variant called. Reads were also subjected to taxonomic binning by [BugSplit](#).

### 3.3 Generation of Results, Output Data (FASTA files), and Reports:

There are various results generated by the BugSeq workflow, such as output data files (especially FASTA files) and reports in webpage (Figure 14) and PDF formats. An interactive report summary of the workflow is one of the results and is in a webpage format (Figure 15). Previous runs on the BugSeq account may also be reviewed (Figure 16). After the consensus

sequence FASTA files have been generated, these are assigned to their corresponding relevant nomenclature designation. For the BugSeq workflow, the PANGO lineage assignment system was used (Figure 17).

Numerous results, such as taxonomic classification of individual read sequences (Figure 18), number of contigs found for each assembly (Figure 19), histogram of mean sequence quality value across each base position (Figure 20), number of reads with certain mean sequence quality (Phred score) (Figure 21), the percentage of base calls at each position for which an N was called (Figure 22), sequence length distribution (Figure 23), adapter content (Figure 24), overall status check per quality metric (Figure 25), filtering statistics of sampled reads (Figure 26), flowcell quality control summary stats (Figure 27), read counts categorized by read quality (Phred score) (Figure 28), number of active pores in the flowcell over time during the sequencing run (Figure 29), and cumulative yield (in gigabases) plot (Figure 30), are also generated by the BugSeq workflow.

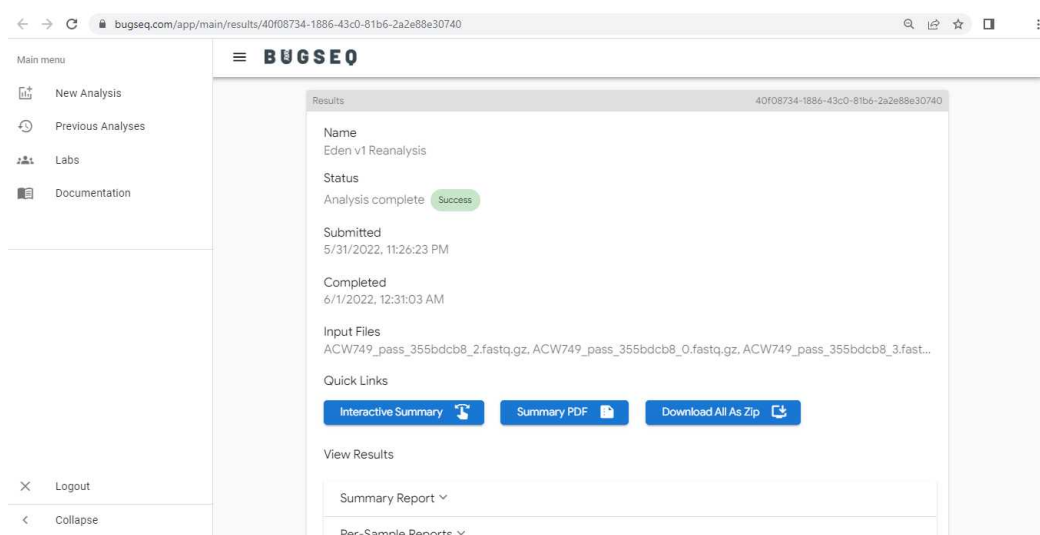


Figure 14. Webpage-based graphical Interface of the output reports



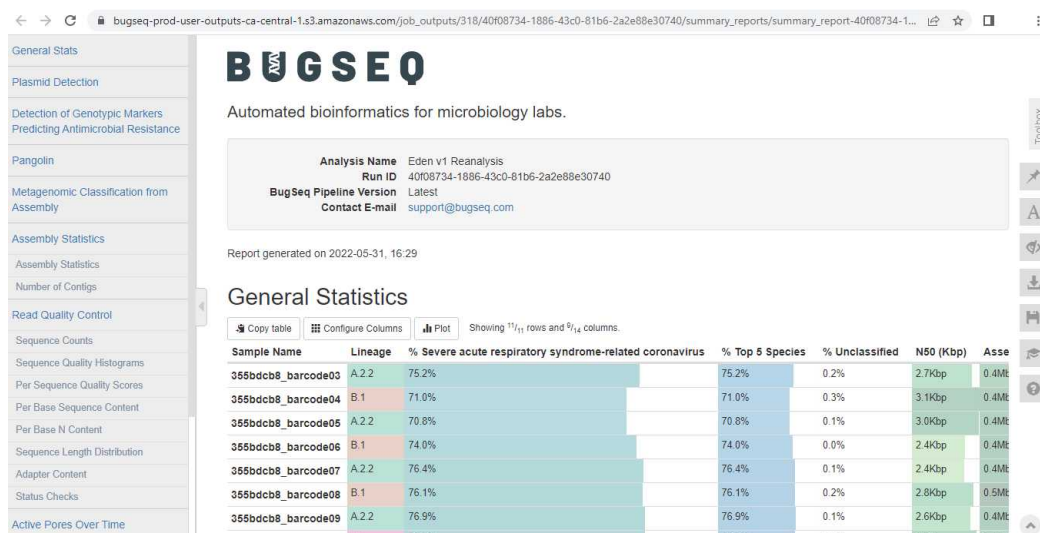


Figure 15. Interactive summary report of the output

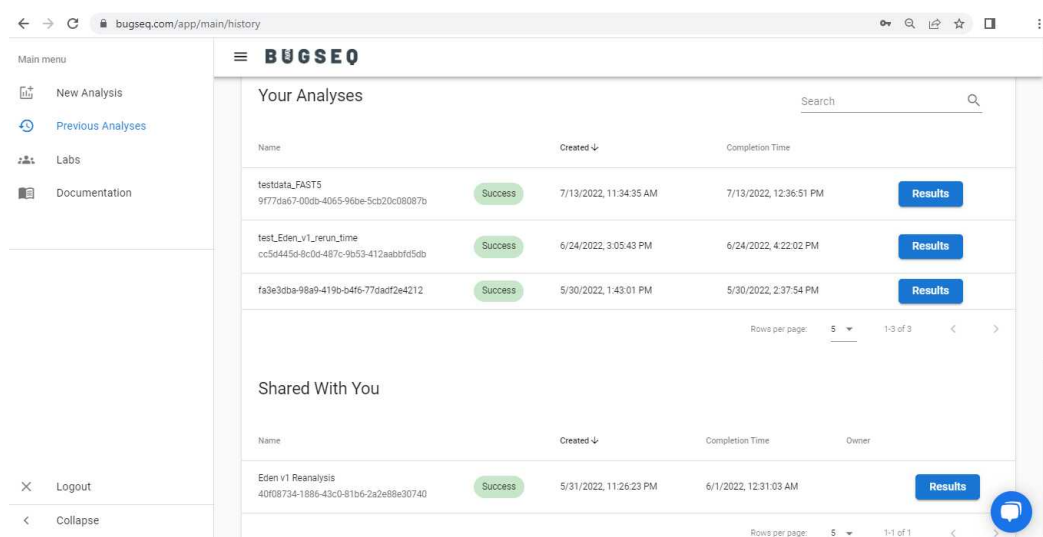


Figure 16. Overview webpage for previous runs

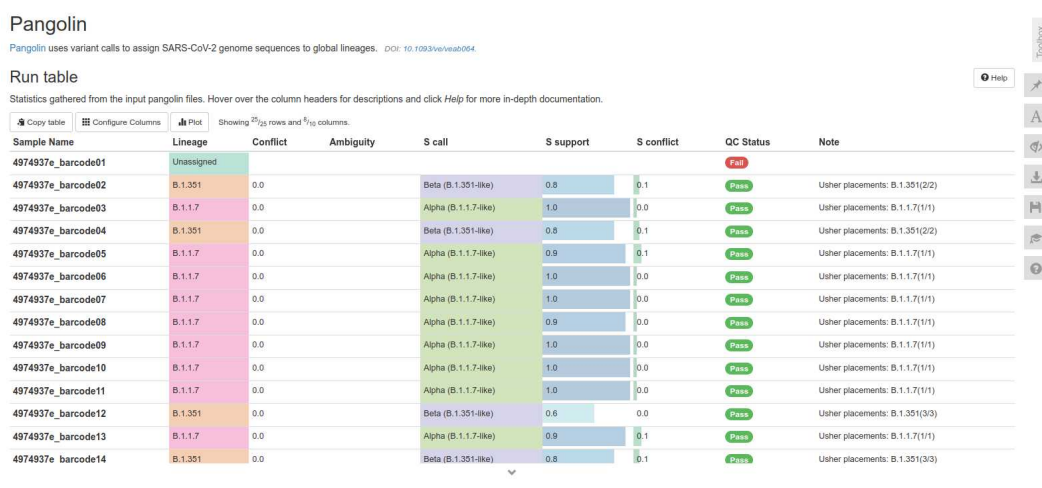


Figure 17. PANGOLIN assignment portion of the report for BugSeq bioinformatics run

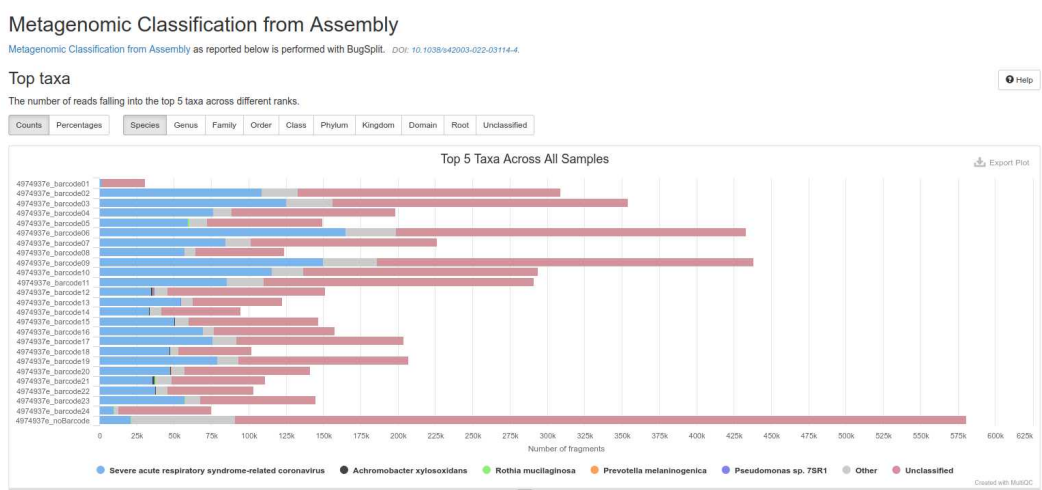


Figure 18. Taxonomic classification of individual read sequences

### Number of Contigs

This plot shows the number of contigs found for each assembly, broken down by length.

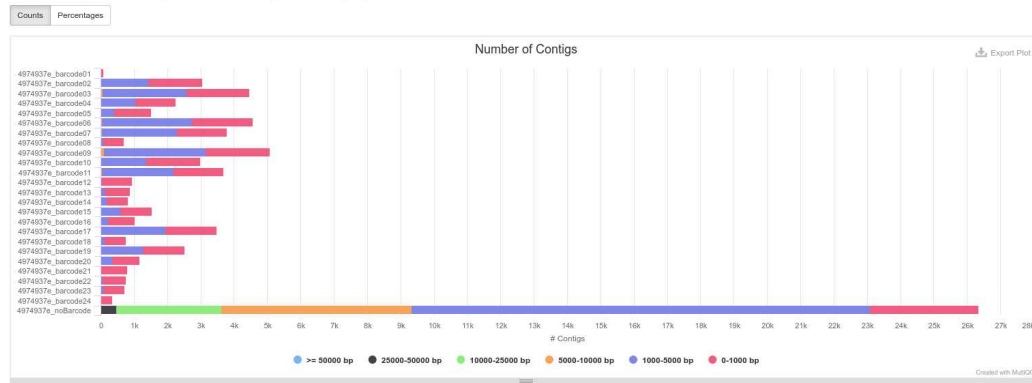


Figure 19. Number of contigs found for each assembly

### Sequence Quality Histograms

The mean quality value across each base position in the read.

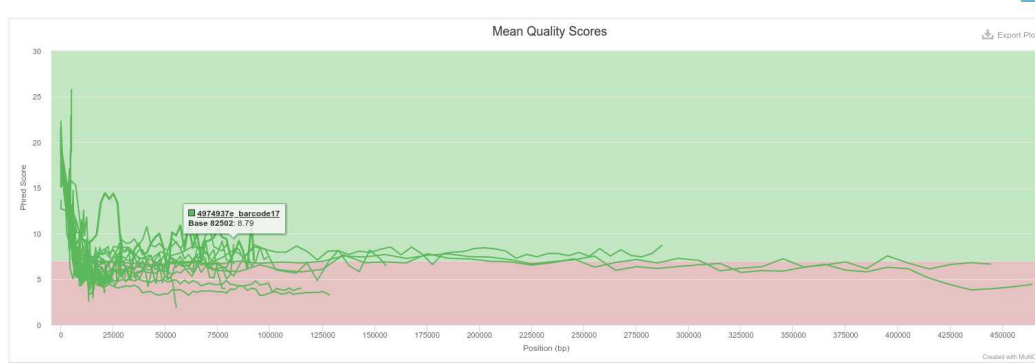


Figure 20. Histogram of mean sequence quality value across each base position



Figure 21. Number of reads with certain mean sequence quality (Phred score)

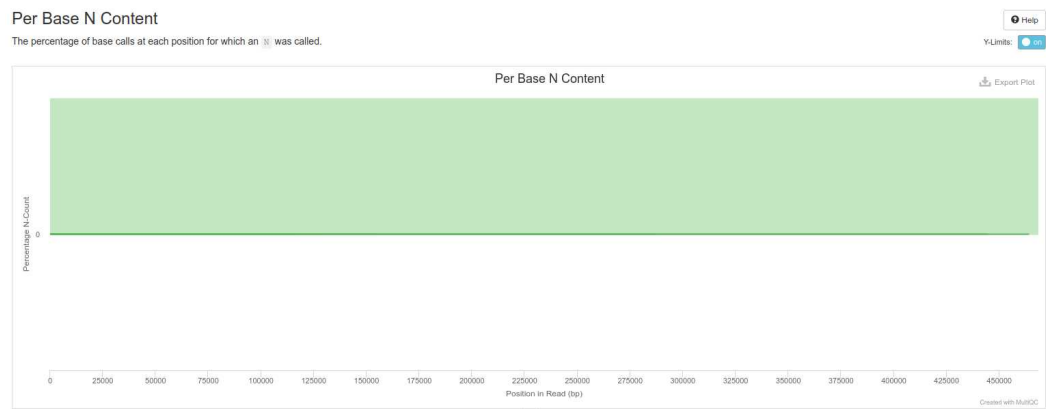


Figure 22. The percentage of base calls at each position for which an N was called

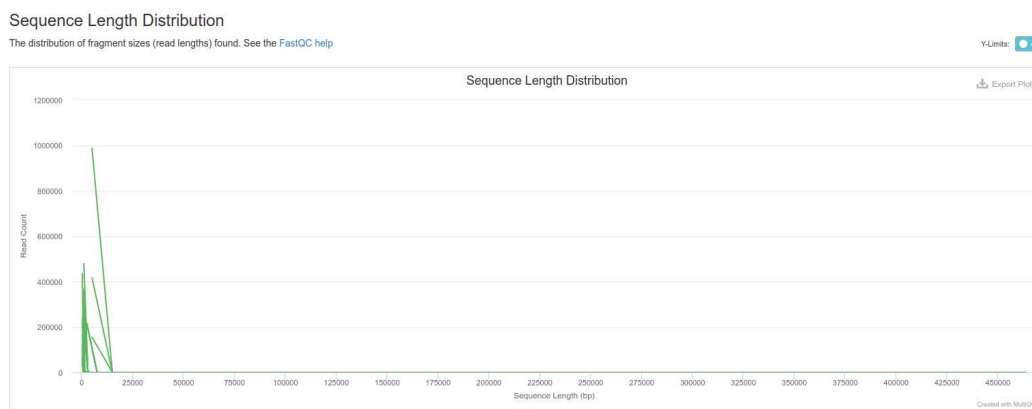


Figure 23. Sequence length distribution among reads



Figure 24. The cumulative percentage count of the proportion within the library that has detected each of the adapter sequences at each position



Figure 25. Overall status check per quality metric

Read Preprocessing

Read Preprocessing trims, filters and corrects reads before further processing. Thresholds for trimming and filtering are based on sequencing platform and experiment type. DOI: 10.1093/bioinformatics/bty560

Filtered Reads

Filtering statistics of sampled reads.

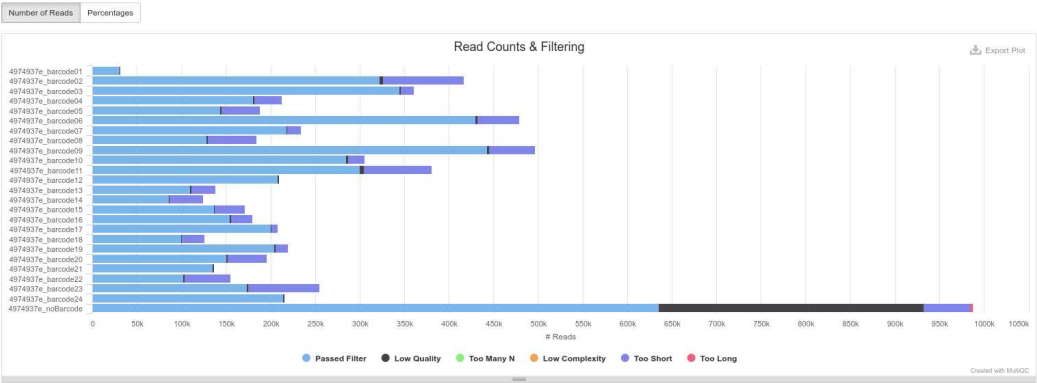


Figure 26. Filtering statistics of sampled reads

Flowcell Quality Control

Flowcell Quality Control assesses the performance of the flowcell for a specific sequencing run. Reads from all barcodes on the sequencing run are aggregated if possible. DOI: 10.1093/bioinformatics/bty149

Seq summary stats

NanoStat statistics from albacore or guppy summary files.

Copy table Configure Columns Plot Showing 1/1 rows and 6/6 columns.

Sample Name	Active channels	Median length	Read N50	Median Qual	# Reads (K)	Total Bases (Mb)
4974937e	447	471 bp	304 bp	10.8	7 585.9	3 630.2

Figure 27. Flowcell quality control summary stats





Figure 28. Read counts categorized by read quality (Phred score)

### Active Pores Over Time

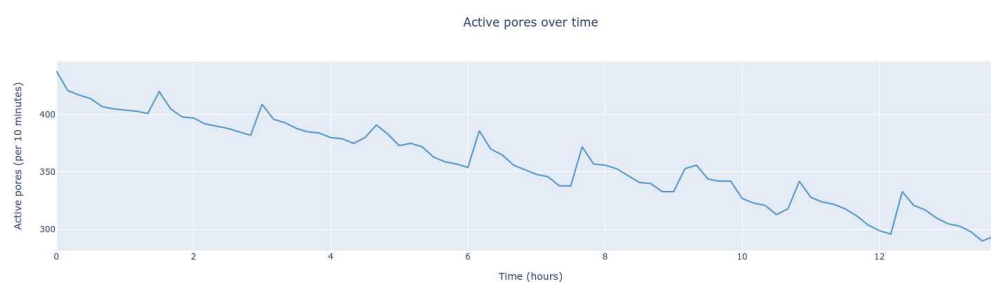


Figure 29. Number of active pores in the flowcell over time during the sequencing run

### Cumulative Yield Plot

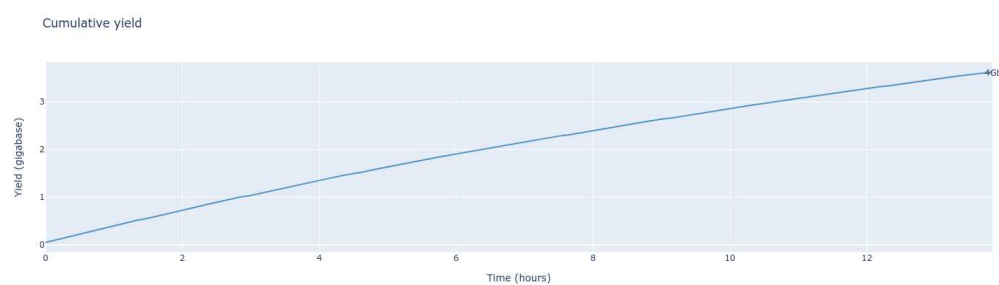


Figure 30. Cumulative yield (in gigabases) plot

#### 4 Comparison of results from PGC Mindanao and BugSeq workflows:

Part of the deliverables of this subgrant is to compare the results of both PGC Mindanao and BugSeq workflows (Figure 31). Both workflows have mostly similar steps and methods, with some differences. Results from similar workflow steps were compared.

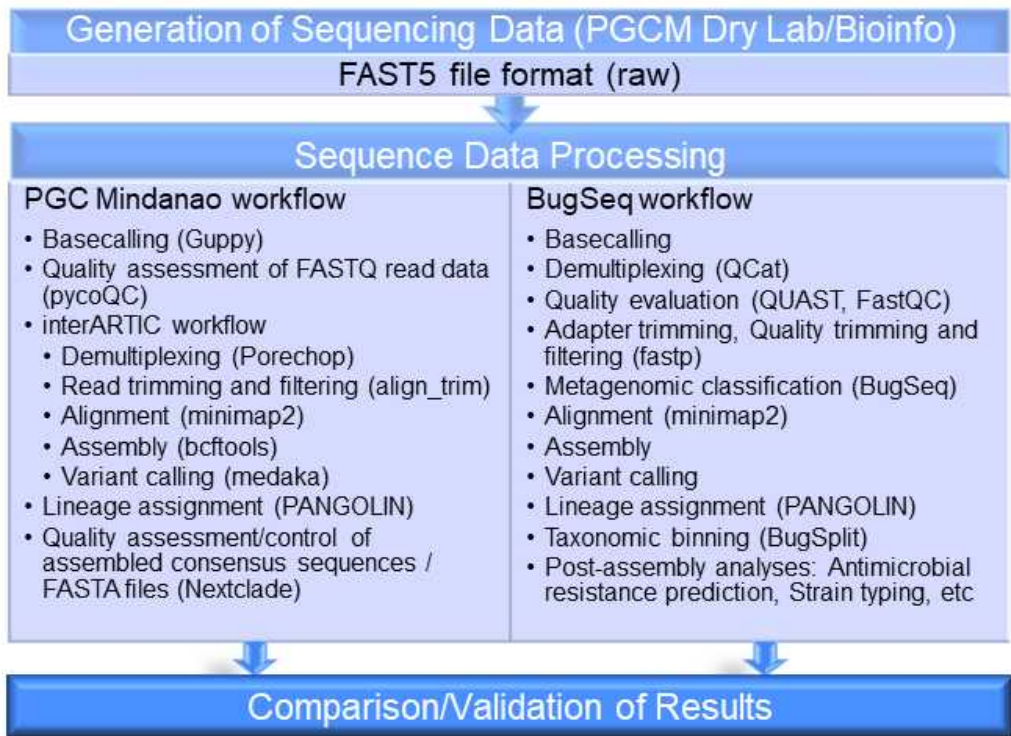


Figure 31. Comparison of PGC Mindanao and BugSeq bioinformatics workflows

As shown in Figure 31 both workflows are generally similar, especially in the alignment algorithms used (minimap2) and lineage assignment workflows (PANGOLIN). However, there were a number of steps that were not done in the BugSeq workflow such as Nextclade. Conversely, there were many steps that were not done in the PGC Mindanao workflow such as metagenomic classification by BugSeq, taxonomic binning by BugSplit, and post-assembly analyses such as antimicrobial resistance prediction and strain typing.

Results of read counts of a certain sequence length were compared between the two workflows, and were found to be comparable (Figure 32). Assembled sequences generated from testing both workflows were compared using pairwise sequence alignment (Needleman-Wunsch algorithm) and were found to be comparable (Table 1). PANGO lineage results from testing both workflows were also compared and generated identical results (Table 2).

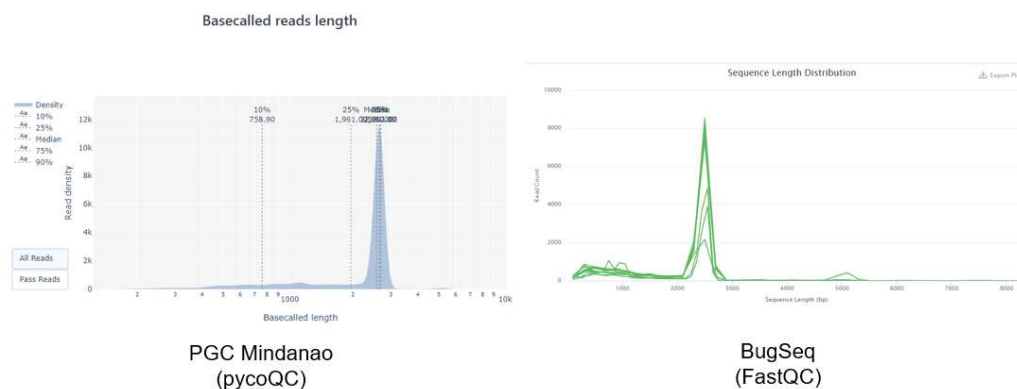


Figure 32. Comparison of results from test run of the two workflows: Read sequence length distribution

Sample (Barcode #)	Score	% identity
03	59181	100.00
04	59181	100.00
05	59181	100.00
06	59176	100.00
07	59181	100.00
08	59181	100.00
09	59181	100.00
10	59206	100.00
11	40796	98.00
12	59172	100.00

Table 1. Comparison of results from test run of the two workflows: Pairwise sequence alignments of assembled sequences (Needleman-Wunsch Algorithm)

Sample (Barcode #)	PGC Mindanao	BugSeq
03	A.2.2	A.2.2
04	B.1	B.1
05	A.2.2	A.2.2
06	B.1	B.1
07	A.2.2	A.2.2
08	B.1	B.1
09	A.2.2	A.2.2
10	A.1	A.1
11	A.2.2	A.2.2
12	B.28	B.28

Table 2. Comparison of results from test run of the two workflows: PANGOLIN (lineage assignment)

In terms of processing time, the BugSeq workflow is able to generate results (in the form of automatically-generated reports presented above) from FASTQ files in approximately 1 hour compared to approximately 1 day by the manual report generation (e.g. entailing much human intervention per step, the reports are in terms of the direct output from the tools) in the PGC Mindanao workflow.

From the aforementioned tests, it can be gathered that both workflows generated comparable and similar results. However, it can be said that the BugSeq workflow is advantageous in terms of speed as shown in the comparison of processing times of both workflows.