medRxiv preprint doi: https://doi.org/10.1101/2021.03.04.21252557; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Detection and genome sequencing of SARS-CoV-2 belonging to the B.1.1.7 lineage in the Philippines

Francis A. Tablizo[1], Cynthia P. Saloma[2,7]*, Marc Jerrone R. Castro[1], Kenneth M. Kim[1], Maria Sofia L. Yangzon[1], Carlo M. Lapid[1], Benedict A. Maralit[3], Marc Edsel C. Ayes[4], Jan Michael C. Yap[1], Jo-Hannah S. Llames[3], Shiela Mae M. Araiza[3], Kris P. Punayan[3], Irish Coleen A. Asin[3], Candice Francheska B. Tambaoan[3], Asia Louisa U. Chong[3], Karol Sophia Agape R. Padilla[3], Rianna Patricia S. Cruz[1], El King D. Morado[1], Joshua Gregor A. Dizon[1], Eva Maria Cutiongco-de la Paz[2,7], Alethea R. de Guzman[5], Razel Nikka M. Hao[6], Arianne A. Zamora[5], Devon Ray Pacial[5], Juan Antonio R. Magalang[5], Marissa Alejandria[7], Celia Carlos[7], Anna Ong-Lim[7], Edsel Maurice Salvaña[7], John Q. Wong[7], Jaime C. Montoya[7], and Maria Rosario Singh-Vergeire[7]

* To whom correspondence should be addressed. Email address: cpsaloma@up.edu.ph

[1] Core Facility for Bioinformatics, Philippine Genome Center, University of the Philippines System

[2] Philippine Genome Center, University of the Philippines System

[3] DNA Sequencing Core Facility, Philippine Genome Center, University of the Philippines System

[4] Clinical Genomics Laboratory, Philippine Genome Center, University of the Philippines System

[5] Epidemiology Bureau, Department of Health, Philippines

[6] Disease Prevention and Control Bureau, Department of Health, Philippines

[7] Inter-Agency Task Force on Emerging Infectious Diseases (IATF) Task Force on COVID-19 Variants, Department of Health, Philippines

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.04.21252557; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

ABSTRACT

We report the sequencing and detection of 39 Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) samples containing lineage-defining mutations specific to viruses belonging to the B.1.1.7 lineage (UK variant) in the Philippines.

ANNOUNCEMENT

The Coronavirus Disease 2019 (COVID-19) is an infectious disease that has gained pandemic status from the World Health Organization, with millions of cases and deaths recorded worldwide. This global health crisis is caused by viruses referred to as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), a member of the genus *Betacoronavirus* (Coronaviridae) together with the causative agents of the first SARS outbreak in 2003 and the Middle East Respiratory Syndrome (MERS) in 2012.

In this study, we present the genome sequences of 39 cases of COVID-19 in the Philippines caused by viruses belonging to SARS-CoV-2 lineage B.1.1.7, more commonly known as the United Kingdom (UK) variant. This particular SARS-CoV-2 variant was initially identified in the United Kingdom and has been reported to cause a surge of COVID-19 infections in the said country (Kirby, 2021). Initial studies also suggest that the B.1.1.7 viruses appear to have replicative advantage (Grabowski et al., 2021) and are more transmissible (Leung et al., 2021).

A total of 39 Philippine SARS-CoV-2 samples were classified under the B.1.1.7 lineage. **Table 1** shows the primary consensus assembly metrics for these samples. The average depth of coverage across all the sequences is 1,128x, with 33 of the samples carrying all 17 hallmark mutations of the B.1.1.7 lineage as listed in the PANGO Lineages report for the B.1.1.7 variant of concern (https://cov-lineages.org/global_report_B.1.1.7.html).

medRxiv preprint doi: https://doi.org/10.1101/2021.03.04.21252557; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The detection of B.1.1.7 from returning overseas Filipino workers and in the community highlights the need for genomic surveillance at the country's ports of entry and the general population to monitor the importation and local transmission of emerging variants of concern that may impact the public health response to the SARS-CoV-2 pandemic in the Philippines.

METHODS

In order to detect the entry of B.1.1.7 infection in the Philippines, nasopharyngeal swabs were collected between December 10, 2020 and January 31, 2021 from COVID-19 cases detected in returning overseas Filipinos, as well as from local case clusters mainly from the Cordillera Administrative Region of the country, among others. The collected swab samples were subjected to RNA extraction using the QIAamp Viral RNA Mini Kit, the product of which were then used as templates for the amplicon-based Illumina COVIDSeq Test (https://www.illumina.com/products/by-type/ivd-products/covidseq.html) sequencing workflow. Afterwards, the resulting sequence reads were mapped to the reference SARS-CoV-2 genome (NCBI Accession Number NC_045512.2) using Minimap2 (Li, 2018) version 2.17-r941, with the `–x sr` parameter for short accurate genomic reads alignment.

Primer clipping and quality trimming, intrahost variant calling, removal of reads associated with mismatched primer indices, and consensus sequence assembly were all done following the suggested workflow of iVar (Grubaugh et al., 2019) version 1.2.2, using default parameters. The consensus variants were identified by comparing the resulting assemblies with the reference sequence using MUMmer (Kurtz et al., 2004) as implemented in the RATT software (Otto et al., 2011). Lastly, SARS-CoV-2 lineage

medRxiv preprint doi: https://doi.org/10.1101/2021.03.04.21252557; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

classifications were assigned using Pangolin version 2.3.2 (github.com/cov-lineages/pangolin; Rambaut et al., 2020).

PUBLIC AVAILABILITY AND ACCESSION NUMBERS

The consensus sequence assemblies reported in this study are deposited in the EpiCoV™ database of the GISAID. The accession codes are shown in **Table 1**.

ACKNOWLEDGEMENT

This project was supported by a Genomics Biosurveillance grant from the Philippine Department of Health, and a Department of Science and Technology – Philippine Council for Health Research and Development grant to BM and the University of the Philippines. The authors would also like to thank the contributing institutions to the Philippine Genomic Biosurveillance Network.

The protocols used in this study were reviewed and approved by the Single Joint Research Ethics Board of the Department of Health, with approval code SJREB-2021-11, as part of a larger research program entitled “A retrospective study on the national genomic surveillance of COVID-19 transmission in the Philippines by SARS-CoV-2 genome sequencing and bioinformatics analysis”.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.04.21252557; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

REFERENCES

Grabowski F., Preibisch G., Kochanczyk M., and Lipniacki T. (2021). SARS-CoV-2 Variant Under Investigation 202012/01 has more than twofold replicative advantage. medRXiV. DOI: https://doi.org/10.1101/2020.12.28.20248906.

Grubaugh N., Gangavarapu K., Quick J., et al. (2019). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biology. 20(8). DOI:https://doi.org/10.1186/s13059-018-1618-7.

Kirby T. (2021). New variant of SARS-COV-2 in UK causes surge of COVID-19. Lancet Respiratory Medicine. DOI:https://doi.org/10.1016/S2213-2600(21)00005-9.

Kurtz S., Phillippy A., Delcher A., Smoot M., Shumway M., Antonescu C., and Salzberg S. (2004). Versatile and open software for comparing large genomes. Genome Biology. 5:R12.

Leung K., Shum M., Leung G., Lam T., and Wu J. (2021). Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. EuroSurveillance. 26(1):pii=2002106. DOI:https://doi.org/10.2807/1560-7917.ES.2020.26.1.2002106.

Li H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 34:3094-3100. doi:10.1093/bioinformatics/bty191.

Otto T., Dillon G., Degrave W., and Berriman M. (2011). RATT: rapid annotation transfer tool. Nucleic Acids Research. 39(9):e57. DOI:https://doi.org/10.1093/nar/gkq1268.

Rambaut, A., Holmes, E.C., O'Toole, Á. et al. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nature Microbiology. 5:1403–140. DOI:https://doi.org/10.1038/s41564-020-0770-5.

Shu, Y., and McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data – from vision to reality. EuroSurveillance. 22(13). DOI:10.2807/1560-7917.ES.2017.22.13.30494.

Table 1. Primary consensus sequence assembly metrics.

Sample Code	GISAID Accession Code	Collection Date	% GC Content	Sample Source	Mean Coverage Depth	# Consensus SNPs	# B.1.1.7 SNPs	% N	Length
PH-PGC-00315	EPI_ISL_1081793	29-Dec-20	37.32	ROF	1,201.27	52	17	1.76	29,884
PH-PGC-00317	EPI_ISL_1081794	29-Dec-20	37.48	ROF	1,194.12	50	17	1.30	29,884
PH-PGC-00401	EPI_ISL_1081795	10-Dec-20	37.22	ROF	996.07	51	17	2.12	29,884
PH-PGC-00986	EPI_ISL_1081796	7-Jan-21	37.14	ROF	1,164.12	50	17	2.26	29,884
PH-PGC-02005	EPI_ISL_1081797	4-Jan-21	31.86	CAR	508.19	44	14	16.82	29,884
PH-PGC-02008	EPI_ISL_1081798	3-Jan-21	35.45	CAR	740.89	48	17	6.87	29,884
PH-PGC-02009	EPI_ISL_1081799	3-Jan-21	34.31	CAR	689.44	45	14	10.19	29,884
PH-PGC-02033	EPI_ISL_1081800	5-Jan-21	37.38	CAR	1,362.04	49	17	1.52	29,884
PH-PGC-02127	EPI_ISL_1081801	7-Jan-21	33.83	CAR	973.77	45	14	11.36	29,884
PH-PGC-02131	EPI_ISL_1081802	7-Jan-21	36.53	CAR	1,326.01	49	17	3.77	29,885
PH-PGC-02133	EPI_ISL_1081803	7-Jan-21	34.19	CAR	892.30	49	16	10.20	29,884
PH-PGC-02152	EPI_ISL_1081804	9-Jan-21	36.16	CAR	1,086.44	38	16	4.80	29,884
PH-PGC-02181	EPI_ISL_1081805	8-Jan-21	36.40	CAR	882.19	50	17	4.18	29,884
PH-PGC-02183	EPI_ISL_1081806	8-Jan-21	37.06	CAR	1,210.23	51	17	2.42	29,884
PH-PGC-02184	EPI_ISL_1081807	8-Jan-21	35.20	CAR	891.29	48	16	7.54	29,884
PH-PGC-02185	EPI_ISL_1081808	8-Jan-21	37.33	CAR	1,131.33	49	17	1.69	29,884
PH-PGC-02225	EPI_ISL_1081809	8-Jan-21	37.13	CAR	1,303.29	51	17	2.14	29,884
PH-PGC-02408	EPI_ISL_1081810	7-Jan-21	37.92	ROF	1,652.89	51	17	0.10	29,884
PH-PGC-02434	EPI_ISL_1081811	12-Jan-21	37.70	ROF	1,168.26	53	17	0.78	29,884
PH-PGC-02630	EPI_ISL_1081812	16-Jan-21	37.83	ROF	1,307.23	48	17	0.41	29,884
PH-PGC-02725	EPI_ISL_1081813	14-Jan-21	37.16	ROF	1,140.03	52	17	2.04	29,884
PH-PGC-02730	EPI_ISL_1081814	16-Jan-21	37.87	ROF	1,620.34	52	17	0.13	29,884

ROF – Returning Overseas Filipino
CAR – Cordillera Administrative Region
NCR – National Capital Region
NM – Northern Mindanao
% N – Percentage of ambiguous base calls ("N" content) in the consensus sequence assembly. High % N values lead to lower % GC content.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.04.21252557; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Sample Code	NCBI Accession Code	Collection Date	% GC Content	Sample Source	Mean Coverage Depth	# Consensus SNPs	# B.1.1.7 SNPs	% N	Length
PH-PGC-02732	EPI_ISL_1081815	17-Jan-21	37.89	ROF	724.58	52	17	0.25	29,884
PH-PGC-02733	EPI_ISL_1081816	17-Jan-21	37.88	ROF	1,499.29	51	17	0.15	29,884
PH-PGC-02740	EPI_ISL_1081817	19-Jan-21	32.63	ROF	769.66	30	12	14.07	29,894
PH-PGC-02745	EPI_ISL_1081818	19-Jan-21	30.64	ROF	679.99	32	12	19.71	29,894
PH-PGC-02756	EPI_ISL_1081819	15-Jan-21	37.01	CAR	1,037.94	50	17	2.23	29,884
PH-PGC-02770	EPI_ISL_1081820	15-Jan-21	35.34	ROF	971.34	44	14	6.34	29,885
PH-PGC-02793	EPI_ISL_1081821	19-Jan-21	37.65	ROF	1,290.90	50	17	0.90	29,884
PH-PGC-02812	EPI_ISL_1081822	24-Jan-21	37.91	CAR	1,654.09	50	17	0.18	29,884
PH-PGC-02826	EPI_ISL_1081823	21-Jan-21	37.85	CAR	1,346.54	52	17	0.31	29,884
PH-PGC-02845	EPI_ISL_1081824	13-Jan-21	37.74	CAR	1,454.34	51	17	0.58	29,884
PH-PGC-02851	EPI_ISL_1081825	11-Jan-21	37.87	CAR	1,563.14	49	17	0.28	29,884
PH-PGC-02886	EPI_ISL_1081826	16-Jan-21	37.67	CAR	1,411.43	51	17	0.86	29,884
PH-PGC-03690	EPI_ISL_1081829	22-Dec-20	19.13	Calabarzon	421.38	20	5	50.35	29,900
PH-PGC-03846	EPI_ISL_1081830	24-Jan-21	37.87	ROF	1,458.88	49	17	0.16	29,884
PH-PGC-03939	EPI_ISL_1081831	31-Jan-21	37.89	ROF	2,082.45	50	16	0.27	29,885
PH-PGC-03978	EPI_ISL_1081832	25-Jan-21	37.25	NCR	990.68	53	16	2.01	29,884
PH-PGC-03995	EPI_ISL_1081841	25-Jan-21	23.29	NM	206.76	22	7	37.77	29,896

ROF – Returning Overseas Filipino
CAR – Cordillera Administrative Region
NCR – National Capital Region
NM – Northern Mindanao
% N – Percentage of ambiguous base calls ("N" content) in the consensus sequence assembly. High % N values lead to lower % GC content.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.04.21252557; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.