Rapid SARS-CoV-2 Whole-genome Sequencing and Analysis for Informed Public Health Decision-making in the Netherlands Replication and Extension

Presented by:

De Ama, Syd
Dizon, Franz
Guaro, Sidney

CSC 771M

February 10, 2021

Presentation Outline

Replication of Phylogenetic Analysis in Netherlands Paper

In this part of the report, we replicate the Phylogenetic Analysis part of Munnick's O., et al. (2020, Sep 26) paper. We skipped the Sequence Data Analysis part because the sequences are already available in the GISAID (https://www.gisaid.org/). Consequently, we only performed the Phylogenetic Analysis part. We also have variances in the tools used because of our computational unavailability.

Data Acquisition

In December 2020, we collected all the sequences and Netherlands only sequences from GISAID (https://www.gisaid.org/) with a submission date of less than or equal to March 17, 2020. The sequences collected are 873 and 189 respectively. The 189 sequences from Netherlands validates our sequences as it has the same number as indicated in Munnink's O., et al. (2020, Sep 26) paper.

Sequences Preprocessing

Similar to Munnink's O., et al. (2020, Sep 26), we filtered the 873 sequences collected from GISAID, which exclude sequences with > 10% NNNs, where the NNNs are known to be the sequence error reads. This is achieved by creating a small python script with the help of a biopython library developed by Cock PA., et al. (2009). The said python script is shown below.

```
from Bio import SeqIO
f_file = "inputFile.fasta"
count = 0
f_sequences = SeqIO.parse(open(f_file), 'fasta')

with open("outputFile.fasta","w") as f:
        for seq in f_sequences:
                seq_seq = seq.seq
                seq_length = len(seq_seq)
                if(seq_seq.count('N')/seq_length<.10):
                        SeqIO.write([seq],f,'fasta')
```

After applying the filter, there are 843 sequences retained in total. These sequences will then be used in Multiple Sequence Alignment (MSA).

Multiple Sequence Alignment

The MSA on the collected sequences was performed using MAFFT (Multiple Alignment using Fast Fourier Transform) software. We used MAFFT instead of MUSCLE which has been used by Munnink’s O., et al. (2020, Sep 26) due to time and memory constraints, MAFFT is a more lightweight program with multithreading support and memory saving option for computers without a huge amount of RAM (Katoh and Toh, 2010).

The command used are shown below:

```
mafft --retree 2 --thread -1 --inputorder “inputFile.fasta” > “outputFile.afa”
```

The parameter “--retree 2” is a default parameter which sets the multiple alignment strategy to use the FFT-NS-2 progressive method. FFT-NS-2 is stated to be capable of aligning a number of sequences up to 5000 on a standard desktop computer (Katoh, et al., 2010). The parameter “--thread -1” is used to automatically set the optimal number of cores to be used. The last parameter used for configuration is the ”--inputorder” which is also a default parameter, it indicates that the order of of sequences in the output file, is the same as the order of sequences in the input file.

After the MSA is generated, we manually checked the sequences for discrepancies.

Tree Construction

In this replication, the Maximum Likelihood (ML) tree was constructed using IQ-TREE (Nguyen, L. et al. 2014) for the aligned sequences output from MSA. We used the ultrafast bootstrap option with 1,000 replicates and GTR+F+I+64 model, the same with the paper of Munnick’s O., et al. (2020, Sep 26).

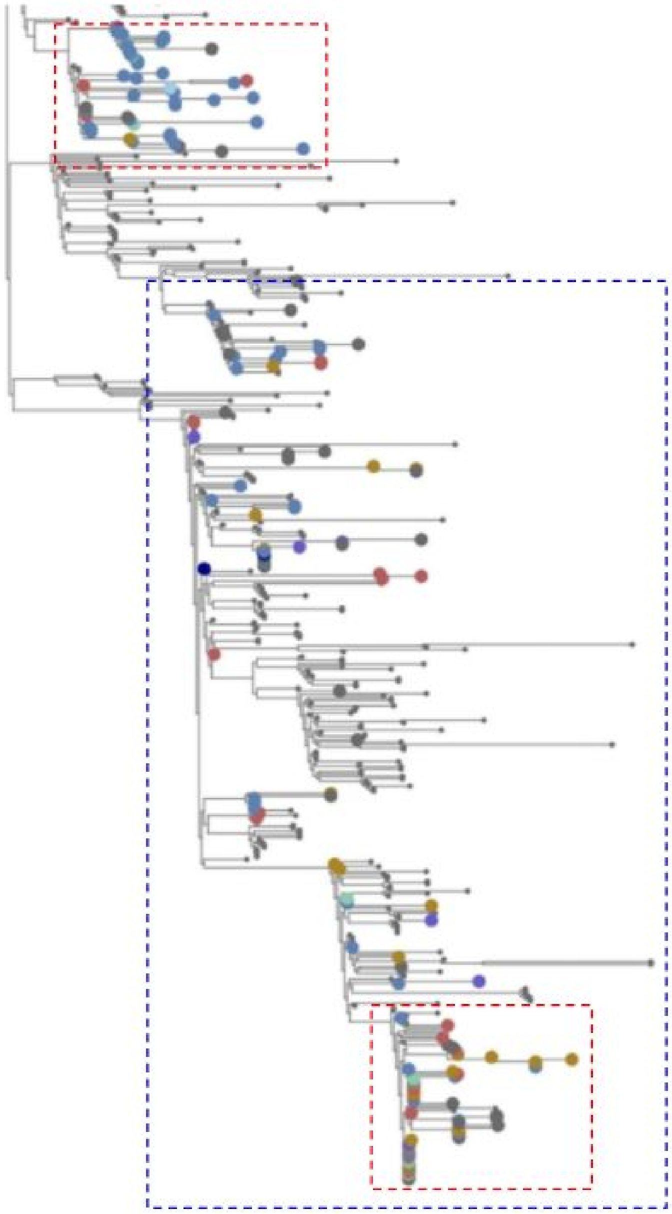
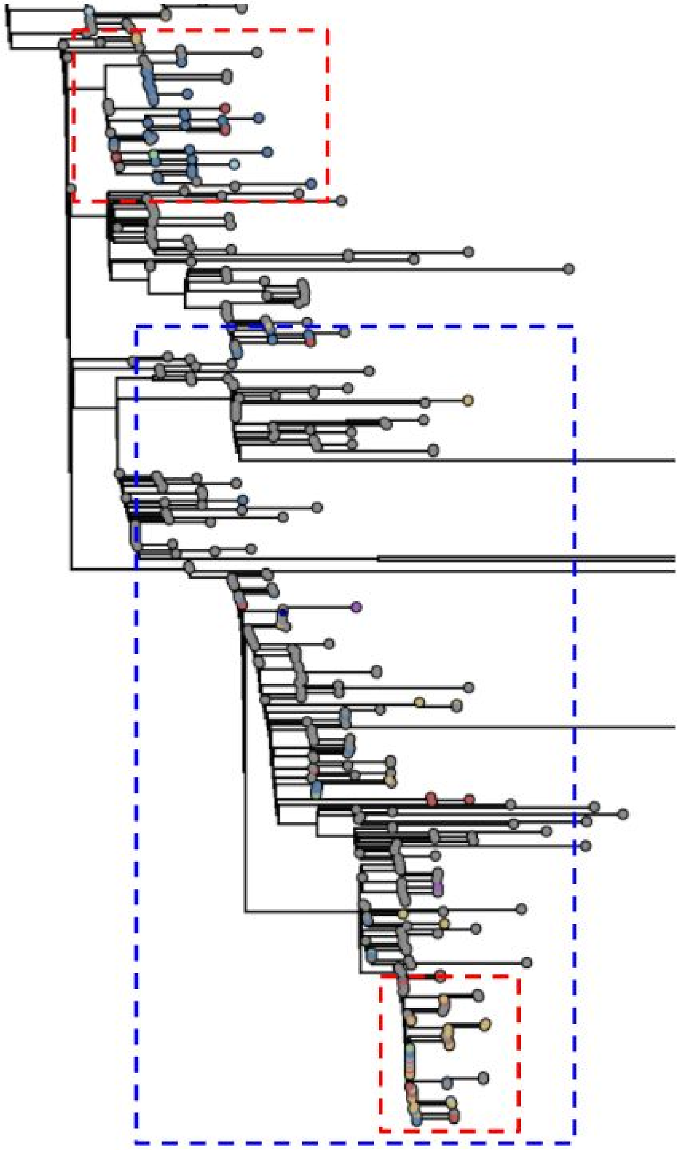
There are several files that are being produced by the IQ-TREE, one of which has a file extension of .treefile which is what we used to visualize the tree. The said file has a file format of newick file, which is then imported to our custom python scripts which uses baltic library for visualizing the tree (Dudas, 2016).

The command used are shown below:

```
.\iqtree2 -s netherlands_world_mafft.afa -bb 1000 -m GTR+F+I+64 -nt AUTO
```

The parameter -s specifies the name of the alignment file. The parameter -bb sets the number of replicates used for bootstrapping. The parameter -m specifies the model to be used. And lastly, the -nt parameter specifies the number of threads used.

Discussion

Munnink’s, O., et al. (2020) phylo. tree Figure 1 (Whole tree shown on appendix A)	Our generated phylogenetic tree Figure 2 (Whole tree shown on appendix B)
	
🟦 Drenthe 🟦 Flevoland 🟩 Gelderland 🟦 Limburg 🟦 Noord-Brabant 🟪 Noord-Holland 🟩 Overijssel 🟥 Utrecht 🟨 Zuid-Holland	

It can be observed that the phylogenetic tree for the Netherlands SARS-Cov-2 cases generated by Munnick's O., et al. (2020), shown in figure 1, was quite similar to the one we have created, which is shown in figure 2. The first major increase of cases in the Netherlands was likely due to multiple introductions of the virus, following the spring holidays where people travel to neighboring countries such as Northern Italy for its ski resorts. This caused several co-circulating viruses, resulting in cases appearing on different clusters in the phylogenetic tree, which can be seen on both figures 1 and 2 indicated by the blue-dashed box. Other cases which add to the diversity of the viruses, which further supports the evidence of the existence of the co-circulating viruses, can be observed on appendix A and B, outside of the green box. On the other-hand, clustering of cases which suggests an increasing number of local transmission, can also be detected on both phylogenetic trees. The upper-red box for both trees indicates the increasing number of local transmission within the province of Noord-Brabant, while the bottom-red box indicates the increasing number of inter-province transmission.

Given this analysis, together with the epidemiological data (e.g. date of exposure, travel history, etc.), the Netherlands government decided to increase the movement restrictions and implement stricter preventive measures for the whole country to prevent further spread of SARS-CoV-2.

Phylogenetic Analysis on Philippine SARS-CoV-2 Sequences Compared to World

The dissemination of SARS-CoV-2 in the world led the countries to control the transmission of diseases by implementing travel bans and community lockdowns. In the Philippines, with a number of foreign visitors and workers, the first two cases recorded were of a Chinese couple entering the country in January 2020. As of February 08, 2021, the COVID-19 cases in the Philippines reached up to 537,310. According to Tablizo, et al (2020), the sequenced Philippine isolates combined with GISAID data, can be divided into three main categories based on sampling dates and potential sources of infection: (1) Samples obtained in the early stages of the pandemic in January are closely related to isolates from Wuhan, China; (2) Samples from March, primarily related to the outbreak of the M/V Diamond Princess Cruise Ship; and (3) Samples clustered with European isolates in June. The clusters should still hold true in the case of the GISAID sequences since there are no recent updates in regards to the Philippine sequences.

In this part of the report, we extend Munnink's, O., et al. (2020, Sep 26) paper by utilizing the phylogenetic analysis methods stated in the study and applying it to the Philippine SARS-CoV-2 sequences. We will utilize the available sequences from the GISAID as it is the only known source which we can acquire the sequences publicly. In this extension, we present the MSA and tree construction of Philippine SARS-CoV-2 genomes in order to analyze how the virus got into the Philippines or from the Philippines, how the virus spread.

Data Acquisition

In January 2021, we investigated how to collect all sequences from GISAID. The option to download all sequences was not available during that time. The only functionally available was the option to download 10,000 sequences at a time from the interface thus we explored the option of web-scraping the GISAID. In mid-January 2021, we downloaded Philippine SARS-CoV-2 sequences for initial pipeline implementation and testing. In the third week of January 2021, we discovered an option that allows us to download all the sequences at once. This option involves communicating with the GISAID support team and acquiring their approval. Fortunately, we obtained the option and on February 03, 2021, the sequences were collected from GISAID.

The sequences collected were aligned and needed to be preprocess in order to perform the sampling selection. The total number of full alignment sequences is 422,925. The sequences are based on 456,943 submissions to EpiCoV. Both duplicate and low-quality sequences are filtered by GISAID where (>5% NNNNs) have been removed, using only complete sequences (length >29,000 bp).

Sequences Preprocessing

The full alignment sequences are 422,925 in total. This will make the MSA and Tree Construction computationally expensive and burdensome on the professionals who analyze phylogenetic trees. To relieve the resources, we first removed the indels (-) in the sequences. Next, we filtered the sequences excluding sequences with > 10% NNNs. This is done in relation to Munnink's, O. et al. (2020, Sep 26) paper. Afterwards, we removed the Philippine sequences to avoid the duplication when adding the whole Philippine sequences.

After the filtration, we sampled the sequences using two methods. The first sampling method is the 10 sequences per country method. This sampling method assumes that the sequences are ordered in descending matter according to date. In the filtered sequences, the sequences order is ascending so we reversed the order and performed the sampling. In the 10 sequences per country sampling, the first 4 sequences, at maximum, from the first month are retained and followed by 4 sequences from the previous month, at maximum, and so on until the total number of sequences of a country is equal to 10. If a month has less than 4 sequences, it will proceed to the previous month (e.g. 2021-01: 1, 2020-12: 4, 2020-11: 4, 2020-10: 1). In this method, 1,195 samples are selected. The second sampling method is the 1 per month method. This sampling method retains the sequences one per month at maximum (e.g. 2021-01: 1, 2020-12: 1, 2020-11: 1, …, 2019-12: 1). In this method, 1,595 samples are selected.

Since the primary goal of this study is to analyze Philippine SARS-CoV-2 mutation and spread, after the sampled sequences are selected, the Philippines sequences are added to the sampled sequences. As a result, the total number of sequences on both sampling methods are 1,231 and 1,631 respectively.

Multiple Sequence Alignment

The MSA on the collected sequences was performed using MAFFT software. We used MAFFT for similar reasons stated in the replication. We also used the similar command stated in the replication and manually checked the sequences for discrepancies (See 10_per_country_msa_visualization.pdf and 1_per_country_msa_visualization.pdf files for the visualization of MSA).

Tree Construction

In the paper extension, the ML trees were constructed using IQ-TREE (Nguyen, L. et al. 2014) for both the multiple sequence aligned outputs. We used the ultrafast bootstrap option with 1,000 replicates for both the aligned sampled sequences. For the 1 per month sampling, the best model is GTR+F+R3 and for the 10 per country, the best model is TIM+F+R3.

The command used are shown below:

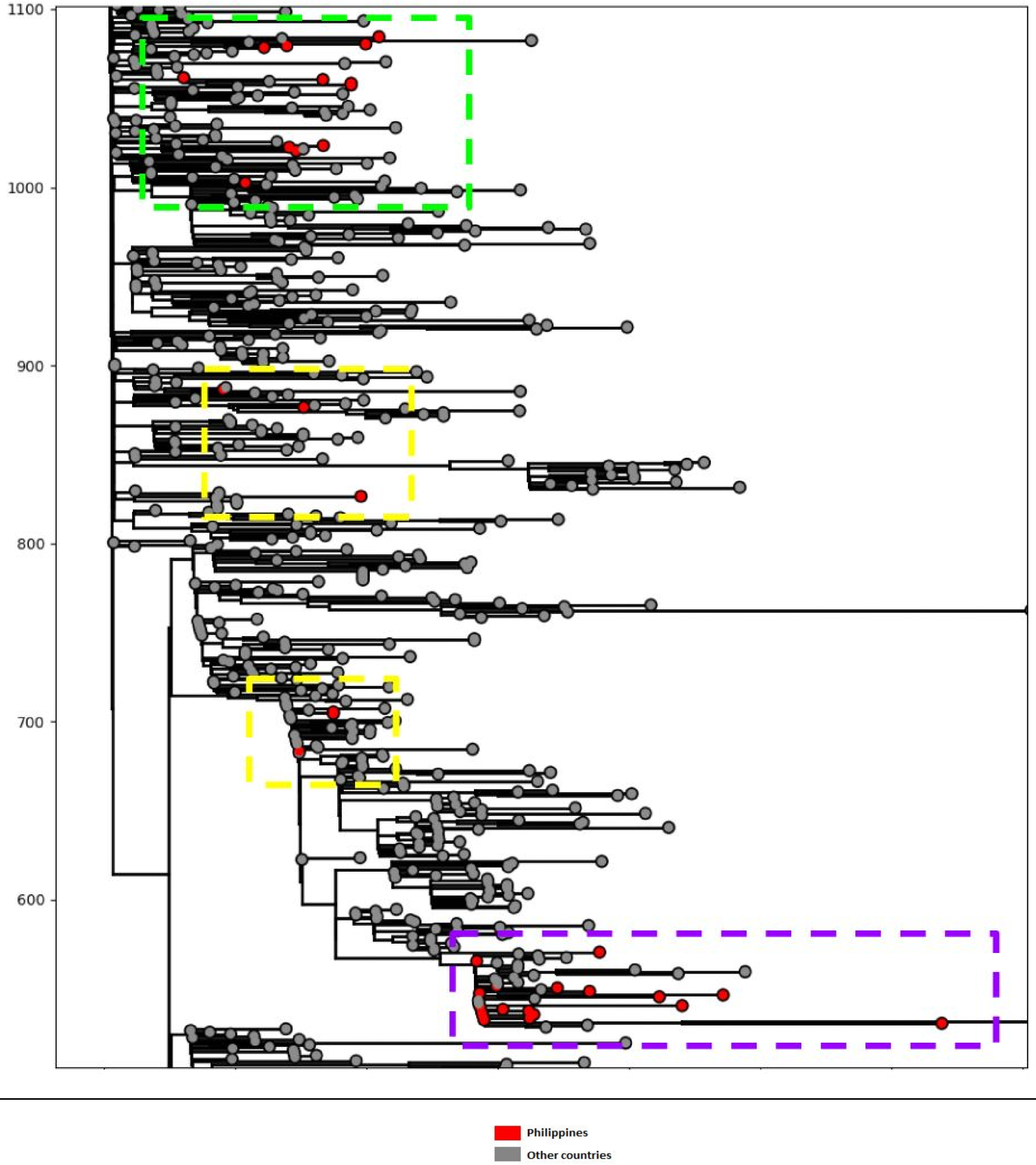
```
.\iqtree2 -s algined_sampled_sequences.afa -bb 1000 -nt AUTO
```

Between the two phylogenetic trees, we select the 10 per country tree in which we will perform our analysis on. We chose the 10 per country tree because it provides a more general clustering on different countries, which may further help us determine the origins of the sequence residing in the Philippines by looking at which cluster it branched out. Similar to the replication part, the output of the IQTREE was visualized using our custom python script that uses the baltic library (Dudas, 2016).

Discussion

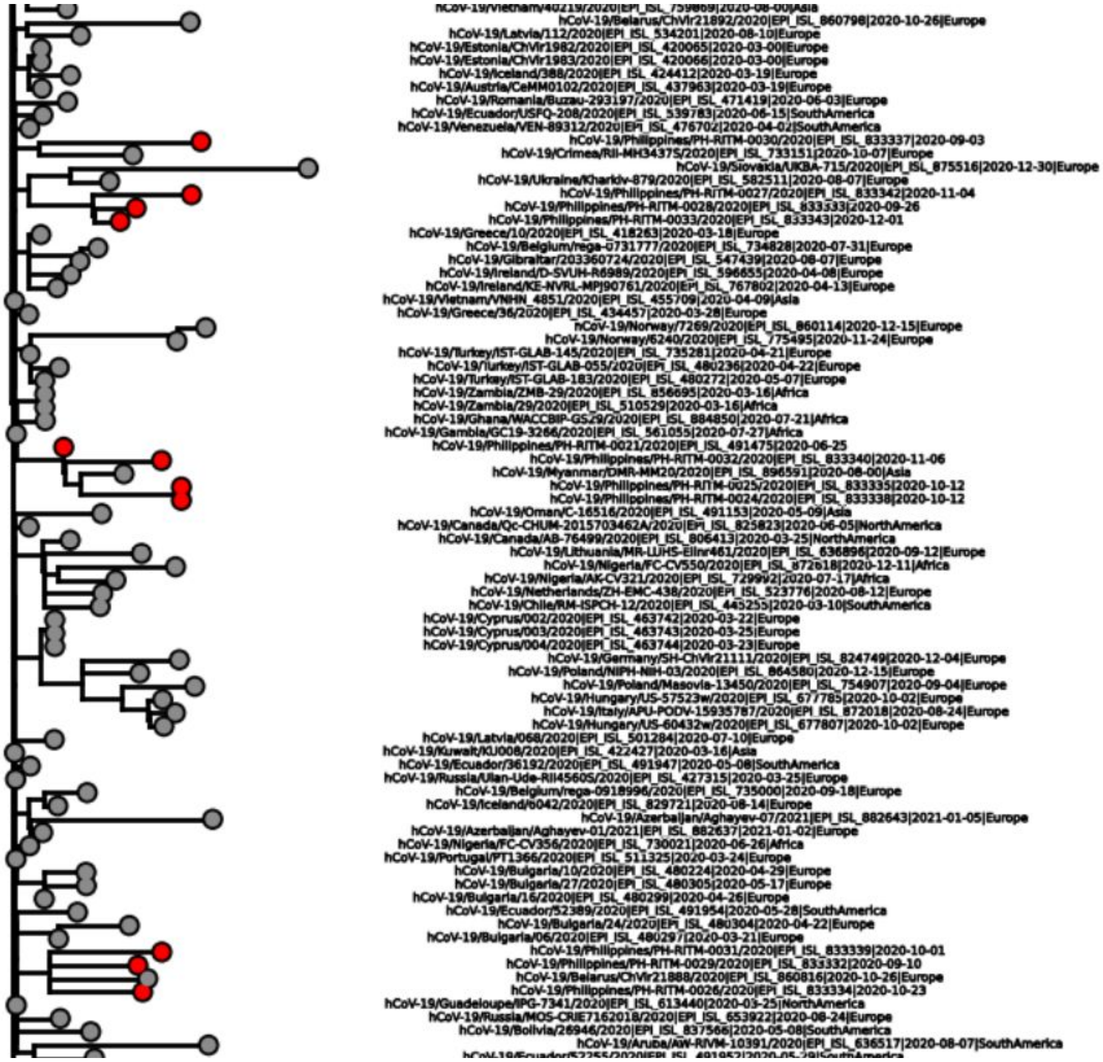
Generated phylogenetic tree for Philippines analysis

Figure 3 (Whole tree shown on appendix C)



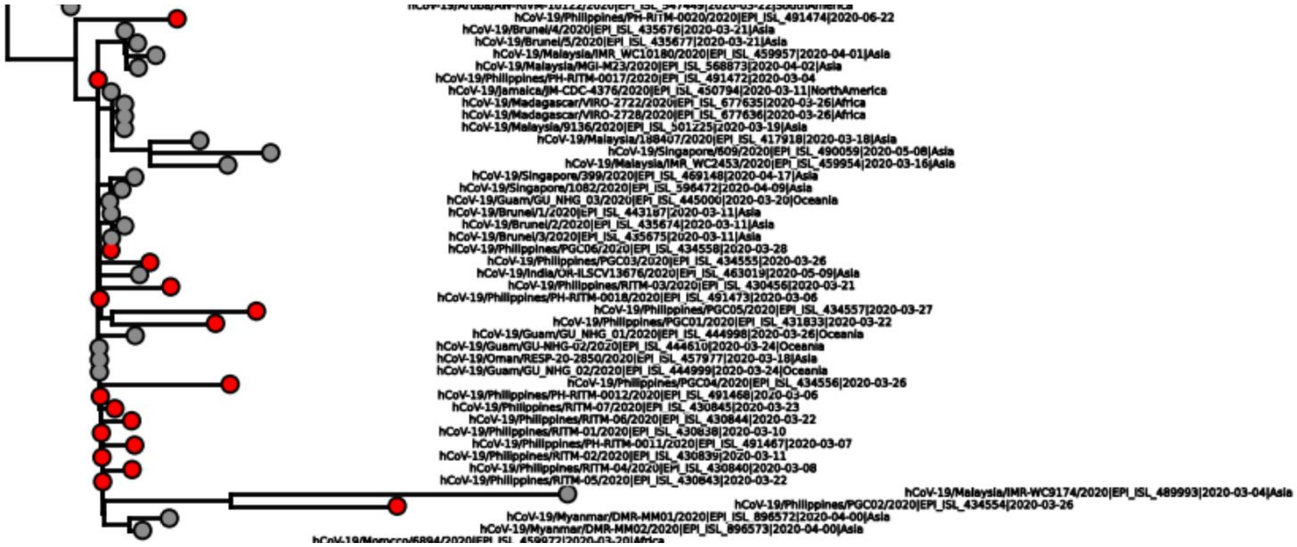
Phylogenetic tree shown on figure 3 is the one that will be used for the analysis of SARS-Cov-2 cases in the Philippines. It can be observed that there exists two forming clusters, highlighted by the red-dashed box and purple-dashed box, which suggests that an ongoing local transmission has occurred in the country. There are also other cases that are located on various parts of the tree, highlighted by the yellow-dashed box, which suggests that there may be new sequences that have been introduced from other countries, and may introduce further local transmission of cases if no action is taken.

Figure 4. Zoom-in of green-dashed box from figure 3


hCoV-19/Vietnam/40219/2020|EPI_ISL_759869|2020-08-00|Asia
hCoV-19/Belarus/ChVir21892/2020|EPI_ISL_860798|2020-10-26|Europe
hCoV-19/Latvia/112/2020|EPI_ISL_534201|2020-08-10|Europe
hCoV-19/Estonia/ChVir1982/2020|EPI_ISL_420065|2020-03-00|Europe
hCoV-19/Estonia/ChVir1983/2020|EPI_ISL_420066|2020-03-00|Europe
hCoV-19/Iceland/388/2020|EPI_ISL_424412|2020-03-19|Europe
hCoV-19/Austria/CeMM0102/2020|EPI_ISL_437963|2020-03-19|Europe
hCoV-19/Romania/Buzau-293197/2020|EPI_ISL_471419|2020-06-03|Europe
hCoV-19/Ecuador/USFQ-208/2020|EPI_ISL_539783|2020-06-15|SouthAmerica
hCoV-19/Venezuela/VEN-89312/2020|EPI_ISL_476702|2020-04-02|SouthAmerica
hCoV-19/Philippines/PH-RITM-0030/2020|EPI_ISL_833337|2020-09-03
hCoV-19/Crimea/RII-MH34375/2020|EPI_ISL_733151|2020-10-07|Europe
hCoV-19/Slovakia/UKBA-715/2020|EPI_ISL_875516|2020-12-30|Europe
hCoV-19/Ukraine/Kharkiv-879/2020|EPI_ISL_582511|2020-08-07|Europe
hCoV-19/Philippines/PH-RITM-0027/2020|EPI_ISL_833342|2020-11-04
hCoV-19/Philippines/PH-RITM-0028/2020|EPI_ISL_833333|2020-09-26
hCoV-19/Philippines/PH-RITM-0033/2020|EPI_ISL_833343|2020-12-01
hCoV-19/Greece/10/2020|EPI_ISL_418263|2020-03-18|Europe
hCoV-19/Belgium/rega-0731777/2020|EPI_ISL_734828|2020-07-31|Europe
hCoV-19/Gibraltar/203360724/2020|EPI_ISL_547439|2020-08-07|Europe
hCoV-19/Ireland/D-SVUH-R6989/2020|EPI_ISL_596655|2020-04-08|Europe
hCoV-19/Ireland/KE-NVRL-MPJ90761/2020|EPI_ISL_767802|2020-04-13|Europe
hCoV-19/Vietnam/VNHN_4851/2020|EPI_ISL_455709|2020-04-09|Asia
hCoV-19/Greece/36/2020|EPI_ISL_434457|2020-03-28|Europe
hCoV-19/Norway/7269/2020|EPI_ISL_860114|2020-12-15|Europe
hCoV-19/Norway/6240/2020|EPI_ISL_775495|2020-11-24|Europe
hCoV-19/Turkey/IST-GLAB-145/2020|EPI_ISL_735281|2020-04-21|Europe
hCoV-19/Turkey/IST-GLAB-055/2020|EPI_ISL_480236|2020-04-22|Europe
hCoV-19/Turkey/IST-GLAB-183/2020|EPI_ISL_480272|2020-05-07|Europe
hCoV-19/Zambia/ZMB-29/2020|EPI_ISL_856695|2020-03-16|Africa
hCoV-19/Zambia/29/2020|EPI_ISL_510529|2020-03-16|Africa
hCoV-19/Ghana/WACCBIP-GS29/2020|EPI_ISL_884850|2020-07-21|Africa
hCoV-19/Gambia/GC19-3266/2020|EPI_ISL_561055|2020-07-27|Africa
hCoV-19/Philippines/PH-RITM-0021/2020|EPI_ISL_491475|2020-06-25
hCoV-19/Philippines/PH-RITM-0032/2020|EPI_ISL_833340|2020-11-06
hCoV-19/Myanmar/DMR-MM20/2020|EPI_ISL_896591|2020-08-00|Asia
hCoV-19/Philippines/PH-RITM-0025/2020|EPI_ISL_833335|2020-10-12
hCoV-19/Philippines/PH-RITM-0024/2020|EPI_ISL_833338|2020-10-12
hCoV-19/Oman/C-16516/2020|EPI_ISL_491153|2020-05-09|Asia
hCoV-19/Canada/Qc-CHUM-2015703462A/2020|EPI_ISL_825823|2020-06-05|NorthAmerica
hCoV-19/Canada/AB-76499/2020|EPI_ISL_806413|2020-03-25|NorthAmerica
hCoV-19/Lithuania/MR-LUHS-Elinr461/2020|EPI_ISL_636896|2020-09-12|Europe
hCoV-19/Nigeria/FC-CV550/2020|EPI_ISL_872618|2020-12-11|Africa
hCoV-19/Nigeria/AK-CV321/2020|EPI_ISL_729992|2020-07-17|Africa
hCoV-19/Netherlands/ZH-EMC-438/2020|EPI_ISL_523776|2020-08-12|Europe
hCoV-19/Chile/RM-ISPCH-12/2020|EPI_ISL_445255|2020-03-10|SouthAmerica
hCoV-19/Cyprus/002/2020|EPI_ISL_463742|2020-03-22|Europe
hCoV-19/Cyprus/003/2020|EPI_ISL_463743|2020-03-25|Europe
hCoV-19/Cyprus/004/2020|EPI_ISL_463744|2020-03-23|Europe
hCoV-19/Germany/SH-ChVir21111/2020|EPI_ISL_824749|2020-12-04|Europe
hCoV-19/Poland/NIPH-NIH-03/2020|EPI_ISL_864580|2020-12-15|Europe
hCoV-19/Poland/Masovia-13450/2020|EPI_ISL_754907|2020-09-04|Europe
hCoV-19/Hungary/US-57523w/2020|EPI_ISL_677785|2020-10-02|Europe
hCoV-19/Italy/APU-PODV-15935787/2020|EPI_ISL_872018|2020-08-24|Europe
hCoV-19/Hungary/US-60432w/2020|EPI_ISL_677807|2020-10-02|Europe
hCoV-19/Latvia/068/2020|EPI_ISL_501284|2020-07-10|Europe
hCoV-19/Kuwait/KU008/2020|EPI_ISL_422427|2020-03-16|Asia
hCoV-19/Ecuador/36192/2020|EPI_ISL_491947|2020-05-08|SouthAmerica
hCoV-19/Russia/Ulan-Ude-RII4560S/2020|EPI_ISL_427315|2020-03-25|Europe
hCoV-19/Belgium/rega-0918996/2020|EPI_ISL_735000|2020-09-18|Europe
hCoV-19/Iceland/6042/2020|EPI_ISL_829721|2020-08-14|Europe
hCoV-19/Azerbaijan/Aghayev-07/2021|EPI_ISL_882643|2021-01-05|Europe
hCoV-19/Azerbaijan/Aghayev-01/2021|EPI_ISL_882637|2021-01-02|Europe
hCoV-19/Nigeria/FC-CV356/2020|EPI_ISL_730021|2020-06-26|Africa
hCoV-19/Portugal/PT1366/2020|EPI_ISL_511325|2020-03-24|Europe
hCoV-19/Bulgaria/10/2020|EPI_ISL_480224|2020-04-29|Europe
hCoV-19/Bulgaria/27/2020|EPI_ISL_480305|2020-05-17|Europe
hCoV-19/Bulgaria/16/2020|EPI_ISL_480299|2020-04-26|Europe
hCoV-19/Ecuador/52389/2020|EPI_ISL_491954|2020-05-28|SouthAmerica
hCoV-19/Bulgaria/24/2020|EPI_ISL_480304|2020-04-22|Europe
hCoV-19/Bulgaria/06/2020|EPI_ISL_480297|2020-03-21|Europe
hCoV-19/Philippines/PH-RITM-0031/2020|EPI_ISL_833339|2020-10-01
hCoV-19/Philippines/PH-RITM-0029/2020|EPI_ISL_833332|2020-09-10
hCoV-19/Belarus/ChVir21888/2020|EPI_ISL_860816|2020-10-26|Europe
hCoV-19/Philippines/PH-RITM-0026/2020|EPI_ISL_833334|2020-10-23
hCoV-19/Guadeloupe/IPG-7341/2020|EPI_ISL_613440|2020-03-25|NorthAmerica
hCoV-19/Russia/MOS-CRIE7162018/2020|EPI_ISL_653922|2020-08-24|Europe
hCoV-19/Bolivia/26946/2020|EPI_ISL_837566|2020-05-08|SouthAmerica
hCoV-19/Aruba/AW-RIVM-10391/2020|EPI_ISL_636517|2020-08-07|SouthAmerica
hCoV-19/Ecuador/52255/2020|EPI_ISL_491952|2020-05-29|SouthAmerica

In the zoom-in view of the green-dashed box shown on figure 4, it suggests that cases in the Philippines that can be found from this cluster may have originated from Europe, and it also indicates that a possible local transmission on the country with this type of strain have already occurred. This is because most of the cases that can be found in this subtree, originated from Europe.

Figure 5. Zoom-in of purple-dashed box from figure 3


hCoV-19/Aruba/AW-RIVM-10122/2020|EPI_ISL_547449|2020-03-22|SouthAmerica
hCoV-19/Philippines/PH-RITM-0020/2020|EPI_ISL_491474|2020-06-22
hCoV-19/Brunei/4/2020|EPI_ISL_435676|2020-03-21|Asia
hCoV-19/Brunei/5/2020|EPI_ISL_435677|2020-03-21|Asia
hCoV-19/Malaysia/IMR_WC10180/2020|EPI_ISL_459957|2020-04-01|Asia
hCoV-19/Malaysia/MGI-M23/2020|EPI_ISL_568873|2020-04-02|Asia
hCoV-19/Philippines/PH-RITM-0017/2020|EPI_ISL_491472|2020-03-04
hCoV-19/Jamaica/JM-CDC-4376/2020|EPI_ISL_450794|2020-03-11|NorthAmerica
hCoV-19/Madagascar/VIRO-2722/2020|EPI_ISL_677635|2020-03-26|Africa
hCoV-19/Madagascar/VIRO-2728/2020|EPI_ISL_677636|2020-03-26|Africa
hCoV-19/Malaysia/9136/2020|EPI_ISL_501225|2020-03-19|Asia
hCoV-19/Malaysia/188407/2020|EPI_ISL_417918|2020-03-18|Asia
hCoV-19/Singapore/609/2020|EPI_ISL_490059|2020-05-08|Asia
hCoV-19/Malaysia/IMR_WC2453/2020|EPI_ISL_459954|2020-03-16|Asia
hCoV-19/Singapore/399/2020|EPI_ISL_469148|2020-04-17|Asia
hCoV-19/Singapore/1082/2020|EPI_ISL_596472|2020-04-09|Asia
hCoV-19/Guam/GU_NHG_03/2020|EPI_ISL_445000|2020-03-20|Oceania
hCoV-19/Brunei/1/2020|EPI_ISL_443187|2020-03-11|Asia
hCoV-19/Brunei/2/2020|EPI_ISL_435674|2020-03-11|Asia
hCoV-19/Brunei/3/2020|EPI_ISL_435675|2020-03-11|Asia
hCoV-19/Philippines/PGC06/2020|EPI_ISL_434558|2020-03-28
hCoV-19/Philippines/PGC03/2020|EPI_ISL_434555|2020-03-26
hCoV-19/India/OR-ILSCV13676/2020|EPI_ISL_463019|2020-05-09|Asia
hCoV-19/Philippines/RITM-03/2020|EPI_ISL_430456|2020-03-21
hCoV-19/Philippines/PH-RITM-0018/2020|EPI_ISL_491473|2020-03-06
hCoV-19/Philippines/PGC05/2020|EPI_ISL_434557|2020-03-27
hCoV-19/Philippines/PGC01/2020|EPI_ISL_431833|2020-03-22
hCoV-19/Guam/GU_NHG_01/2020|EPI_ISL_444998|2020-03-26|Oceania
hCoV-19/Guam/GU-NHG-02/2020|EPI_ISL_444610|2020-03-24|Oceania
hCoV-19/Oman/RESP-20-2850/2020|EPI_ISL_457977|2020-03-18|Asia
hCoV-19/Guam/GU_NHG_02/2020|EPI_ISL_444999|2020-03-24|Oceania
hCoV-19/Philippines/PGC04/2020|EPI_ISL_434556|2020-03-26
hCoV-19/Philippines/PH-RITM-0012/2020|EPI_ISL_491468|2020-03-06
hCoV-19/Philippines/RITM-07/2020|EPI_ISL_430845|2020-03-23
hCoV-19/Philippines/RITM-06/2020|EPI_ISL_430844|2020-03-22
hCoV-19/Philippines/RITM-01/2020|EPI_ISL_430838|2020-03-10
hCoV-19/Philippines/PH-RITM-0011/2020|EPI_ISL_491467|2020-03-07
hCoV-19/Philippines/RITM-02/2020|EPI_ISL_430839|2020-03-11
hCoV-19/Philippines/RITM-04/2020|EPI_ISL_430840|2020-03-08
hCoV-19/Philippines/RITM-05/2020|EPI_ISL_430843|2020-03-22
hCoV-19/Malaysia/IMR-WC9174/2020|EPI_ISL_489993|2020-03-04|Asia
hCoV-19/Philippines/PGC02/2020|EPI_ISL_434554|2020-03-26
hCoV-19/Myanmar/DMR-MM01/2020|EPI_ISL_896572|2020-04-00|Asia
hCoV-19/Myanmar/DMR-MM02/2020|EPI_ISL_896573|2020-04-00|Asia
hCoV-19/Morocco/6894/2020|EPI_ISL_459972|2020-03-20|Africa

In the zoom-in view of the purple-dashed box shown on figure 5, it suggests that cases in the Philippines that can be found from this cluster may have originated from the neighboring countries like Malaysia, Singapore, Brunei, and others. Furthermore, since the cases here are observably denser, it suggests that a relatively large amount of local transmision took place with respect to the green-dashed box. This may be one of the reasons why the Philippine government decided to implement lockdowns and cancellation of flights.

Figure 6. Zoom-ins for yellow-dashed boxes from figure 3

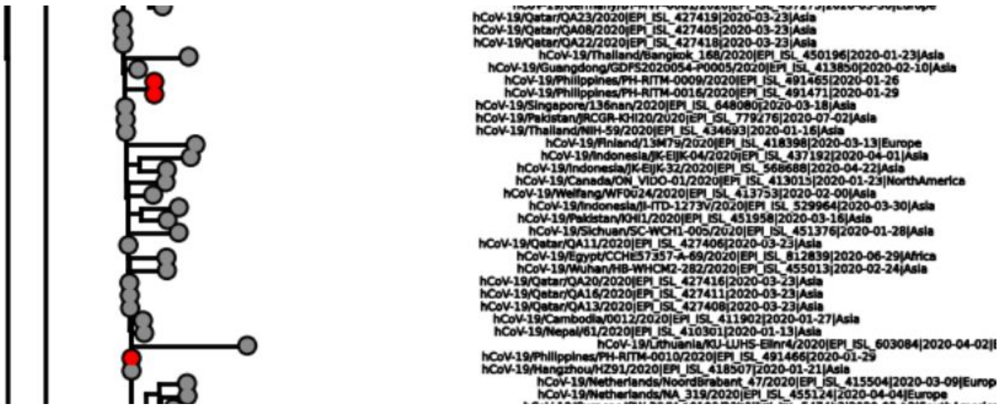

hCoV-19/Qatar/QA23/2020|EPI_ISL_427419|2020-03-23|Asia
hCoV-19/Qatar/QA08/2020|EPI_ISL_427405|2020-03-23|Asia
hCoV-19/Qatar/QA22/2020|EPI_ISL_427418|2020-03-23|Asia
hCoV-19/Thailand/Bangkok_168/2020|EPI_ISL_450196|2020-01-23|Asia
hCoV-19/Guangdong/GDFS2020054-P0005/2020|EPI_ISL_413850|2020-02-10|Asia
hCoV-19/Philippines/PH-RITM-0009/2020|EPI_ISL_491465|2020-01-26
hCoV-19/Philippines/PH-RITM-0016/2020|EPI_ISL_491471|2020-01-29
hCoV-19/Singapore/136nan/2020|EPI_ISL_648080|2020-03-18|Asia
hCoV-19/Pakistan/JRCGR-KHI20/2020|EPI_ISL_779276|2020-07-02|Asia
hCoV-19/Thailand/NIH-59/2020|EPI_ISL_434693|2020-01-16|Asia
hCoV-19/Finland/13M79/2020|EPI_ISL_418398|2020-03-13|Europe
hCoV-19/Indonesia/JK-EIJK-04/2020|EPI_ISL_437192|2020-04-01|Asia
hCoV-19/Indonesia/JK-EIJK-32/2020|EPI_ISL_568688|2020-04-22|Asia
hCoV-19/Canada/ON_VIDO-01/2020|EPI_ISL_413015|2020-01-23|NorthAmerica
hCoV-19/Weifang/WF0024/2020|EPI_ISL_413753|2020-02-00|Asia
hCoV-19/Indonesia/JI-ITD-1273V/2020|EPI_ISL_529964|2020-03-30|Asia
hCoV-19/Pakistan/KHI1/2020|EPI_ISL_451958|2020-03-16|Asia
hCoV-19/Sichuan/SC-WCH1-005/2020|EPI_ISL_451376|2020-01-28|Asia
hCoV-19/Qatar/QA11/2020|EPI_ISL_427406|2020-03-23|Asia
hCoV-19/Egypt/CCHE57357-A-69/2020|EPI_ISL_812839|2020-06-29|Africa
hCoV-19/Wuhan/HB-WHCM2-282/2020|EPI_ISL_455013|2020-02-24|Asia
hCoV-19/Qatar/QA20/2020|EPI_ISL_427416|2020-03-23|Asia
hCoV-19/Qatar/QA16/2020|EPI_ISL_427411|2020-03-23|Asia
hCoV-19/Qatar/QA13/2020|EPI_ISL_427408|2020-03-23|Asia
hCoV-19/Cambodia/0012/2020|EPI_ISL_411902|2020-01-27|Asia
hCoV-19/Nepal/61/2020|EPI_ISL_410301|2020-01-13|Asia
hCoV-19/Lithuania/KU-LUHS-Eilnr4/2020|EPI_ISL_603084|2020-04-02|
hCoV-19/Philippines/PH-RITM-0010/2020|EPI_ISL_491466|2020-01-29
hCoV-19/Hangzhou/HZ91/2020|EPI_ISL_418507|2020-01-21|Asia
hCoV-19/Netherlands/NoordBrabant_47/2020|EPI_ISL_415504|2020-03-09|Europ
hCoV-19/Netherlands/NA_319/2020|EPI_ISL_455124|2020-04-04|Europe

Figure 6.A
Bottom
Yellow Box

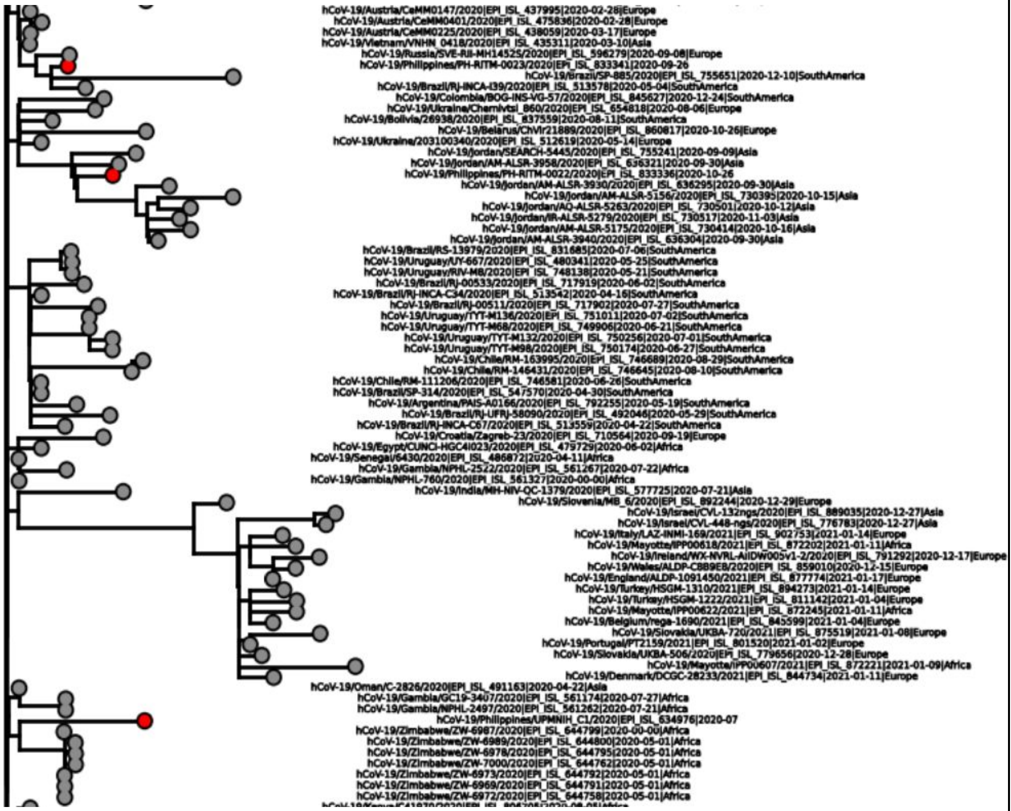

hCoV-19/Austria/CeMM0147/2020|EPI_ISL_437995|2020-02-28|Europe
hCoV-19/Austria/CeMM0401/2020|EPI_ISL_475836|2020-02-28|Europe
hCoV-19/Austria/CeMM0225/2020|EPI_ISL_438059|2020-03-17|Europe
hCoV-19/Vietnam/VNHN_0418/2020|EPI_ISL_435311|2020-03-10|Asia
hCoV-19/Russia/SVE-RII-MH1452S/2020|EPI_ISL_596279|2020-09-08|Europe
hCoV-19/Philippines/PH-RITM-0023/2020|EPI_ISL_833341|2020-09-26
hCoV-19/Brazil/SP-885/2020|EPI_ISL_755651|2020-12-10|SouthAmerica
hCoV-19/Brazil/RJ-INCA-I39/2020|EPI_ISL_513578|2020-05-04|SouthAmerica
hCoV-19/Colombia/BOG-INS-VG-57/2020|EPI_ISL_845627|2020-12-24|SouthAmerica
hCoV-19/Ukraine/Chernivtsi_860/2020|EPI_ISL_654818|2020-08-06|Europe
hCoV-19/Bolivia/26938/2020|EPI_ISL_837559|2020-08-11|SouthAmerica
hCoV-19/Belarus/ChVir21889/2020|EPI_ISL_860817|2020-10-26|Europe
hCoV-19/Ukraine/203100340/2020|EPI_ISL_512619|2020-05-14|Europe
hCoV-19/Jordan/SEARCH-5445/2020|EPI_ISL_755241|2020-09-09|Asia
hCoV-19/Jordan/AM-ALSR-3958/2020|EPI_ISL_636321|2020-09-30|Asia
hCoV-19/Philippines/PH-RITM-0022/2020|EPI_ISL_833336|2020-10-26
hCoV-19/Jordan/AM-ALSR-3930/2020|EPI_ISL_636295|2020-09-30|Asia
hCoV-19/Jordan/AM-ALSR-5156/2020|EPI_ISL_730395|2020-10-15|Asia
hCoV-19/Jordan/AQ-ALSR-5263/2020|EPI_ISL_730501|2020-10-12|Asia
hCoV-19/Jordan/IR-ALSR-5279/2020|EPI_ISL_730517|2020-11-03|Asia
hCoV-19/Jordan/AM-ALSR-5175/2020|EPI_ISL_730414|2020-10-16|Asia
hCoV-19/Jordan/AM-ALSR-3940/2020|EPI_ISL_636304|2020-09-30|Asia
hCoV-19/Brazil/RS-13979/2020|EPI_ISL_831685|2020-07-06|SouthAmerica
hCoV-19/Uruguay/UY-667/2020|EPI_ISL_480341|2020-05-25|SouthAmerica
hCoV-19/Uruguay/RIV-M8/2020|EPI_ISL_748138|2020-05-21|SouthAmerica
hCoV-19/Brazil/RJ-00533/2020|EPI_ISL_717919|2020-06-02|SouthAmerica
hCoV-19/Brazil/RJ-INCA-C34/2020|EPI_ISL_513542|2020-04-16|SouthAmerica
hCoV-19/Brazil/RJ-00511/2020|EPI_ISL_717902|2020-07-27|SouthAmerica
hCoV-19/Uruguay/TYT-M136/2020|EPI_ISL_751011|2020-07-02|SouthAmerica
hCoV-19/Uruguay/TYT-M68/2020|EPI_ISL_749906|2020-06-21|SouthAmerica
hCoV-19/Uruguay/TYT-M132/2020|EPI_ISL_750256|2020-07-01|SouthAmerica
hCoV-19/Uruguay/TYT-M98/2020|EPI_ISL_750174|2020-06-27|SouthAmerica
hCoV-19/Chile/RM-163995/2020|EPI_ISL_746689|2020-08-29|SouthAmerica
hCoV-19/Chile/RM-146431/2020|EPI_ISL_746645|2020-08-10|SouthAmerica
hCoV-19/Chile/RM-111206/2020|EPI_ISL_746581|2020-06-26|SouthAmerica
hCoV-19/Brazil/SP-314/2020|EPI_ISL_547570|2020-04-30|SouthAmerica
hCoV-19/Argentina/PAIS-A0166/2020|EPI_ISL_792255|2020-05-19|SouthAmerica
hCoV-19/Brazil/RJ-UFRJ-58090/2020|EPI_ISL_492046|2020-05-29|SouthAmerica
hCoV-19/Brazil/RJ-INCA-C67/2020|EPI_ISL_513559|2020-04-22|SouthAmerica
hCoV-19/Croatia/Zagreb-23/2020|EPI_ISL_710564|2020-09-19|Europe
hCoV-19/Egypt/CUNCI-HGC4I023/2020|EPI_ISL_479729|2020-06-02|Africa
hCoV-19/Senegal/6430/2020|EPI_ISL_486872|2020-04-11|Africa
hCoV-19/Gambia/NPHL-2522/2020|EPI_ISL_561267|2020-07-22|Africa
hCoV-19/Gambia/NPHL-760/2020|EPI_ISL_561327|2020-00-00|Africa
hCoV-19/India/MH-NIV-QC-1379/2020|EPI_ISL_577725|2020-07-21|Asia
hCoV-19/Slovenia/MB_6/2020|EPI_ISL_892244|2020-12-29|Europe
hCoV-19/Israel/CVL-132ngs/2020|EPI_ISL_889035|2020-12-27|Asia
hCoV-19/Israel/CVL-448-ngs/2020|EPI_ISL_776783|2020-12-27|Asia
hCoV-19/Italy/LAZ-INMI-169/2021|EPI_ISL_902753|2021-01-14|Europe
hCoV-19/Mayotte/IPP00618/2021|EPI_ISL_872202|2021-01-11|Africa
hCoV-19/Ireland/WX-NVRL-AIIDW005v1-2/2020|EPI_ISL_791292|2020-12-17|Europe
hCoV-19/Wales/ALDP-CB89EB/2020|EPI_ISL_859010|2020-12-15|Europe
hCoV-19/England/ALDP-1091450/2021|EPI_ISL_877774|2021-01-17|Europe
hCoV-19/Turkey/HSGM-1310/2021|EPI_ISL_894273|2021-01-14|Europe
hCoV-19/Turkey/HSGM-1222/2021|EPI_ISL_811142|2021-01-04|Europe
hCoV-19/Mayotte/IPP00622/2021|EPI_ISL_872245|2021-01-11|Africa
hCoV-19/Belgium/rega-1690/2021|EPI_ISL_845599|2021-01-04|Europe
hCoV-19/Slovakia/UKBA-720/2021|EPI_ISL_875519|2021-01-08|Europe
hCoV-19/Portugal/PT2159/2021|EPI_ISL_801520|2021-01-02|Europe
hCoV-19/Slovakia/UKBA-506/2020|EPI_ISL_779656|2020-12-28|Europe
hCoV-19/Mayotte/IPP00607/2021|EPI_ISL_872221|2021-01-09|Africa
hCoV-19/Denmark/DCGC-28233/2021|EPI_ISL_844734|2021-01-11|Europe
hCoV-19/Oman/C-2826/2020|EPI_ISL_491163|2020-04-22|Asia
hCoV-19/Gambia/GC19-3407/2020|EPI_ISL_561174|2020-07-27|Africa
hCoV-19/Gambia/NPHL-2497/2020|EPI_ISL_561262|2020-07-21|Africa
hCoV-19/Philippines/UPMNIH_C1/2020|EPI_ISL_634976|2020-07
hCoV-19/Zimbabwe/ZW-6987/2020|EPI_ISL_644799|2020-00-00|Africa
hCoV-19/Zimbabwe/ZW-6989/2020|EPI_ISL_644800|2020-05-01|Africa
hCoV-19/Zimbabwe/ZW-6978/2020|EPI_ISL_644795|2020-05-01|Africa
hCoV-19/Zimbabwe/ZW-7000/2020|EPI_ISL_644762|2020-05-01|Africa
hCoV-19/Zimbabwe/ZW-6973/2020|EPI_ISL_644792|2020-05-01|Africa
hCoV-19/Zimbabwe/ZW-6969/2020|EPI_ISL_644791|2020-05-01|Africa
hCoV-19/Zimbabwe/ZW-6972/2020|EPI_ISL_644758|2020-05-01|Africa

Figure 6.B
Upper
Yellow Box

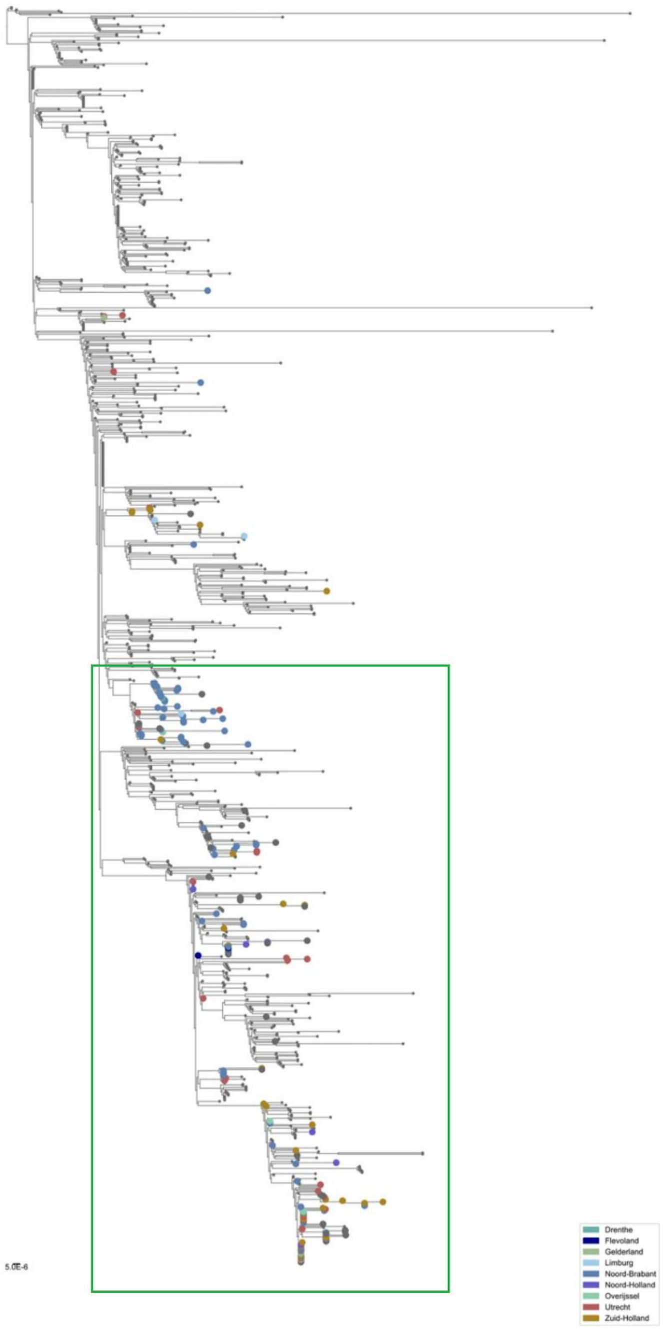
In the zoom-in views of the yellow-dashed boxes shown on figure 6, it suggests that there are also different strains of viruses that are circulating in the Philippines. One possible reason that we can think for this is due to an increased number of Overseas-Filipino-Workers (OFW) that are going back home in the country, specially the neighboring cases that can be found under the subtree originated from countries like Qatar, Jordan, and Singapore, which are known countries to have a number of OFWs.

References

- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2014, November 3). *IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies*. OUP Academic. https://academic.oup.com/mbe/article/32/1/268/2925592?login=true.
- Munnink, B. B. O., Nieuwenhuijse, D. F., Stein, M., O'Toole, Á., Haverkate, M., Mollers, M., … Koopmans, M. (2020, July 16). *Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands*. Nature News. https://www.nature.com/articles/s41591-020-0997-y.
- Tablizo, F. A., Lapid, C. M., Maralit, B. A., Yap, J. M. C., Destura, R. V., Alejandria, M. A., ... & Saloma, C. P. (2020). ANALYSIS OF SARS-COV-2 GENOME SEQUENCES FROM THE PHILIPPINES: GENETIC SURVEILLANCE AND TRANSMISSION DYNAMICS. medRxiv.
- Katoh, Misawa, Kuma, Miyata 2002 (*Nucleic Acids Res.* **30**:3059-3066)
  MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.
  (describes the FFT-NS-1, FFT-NS-2 and FFT-NS-i strategies)
- Katoh, Toh 2010 (*Bioinformatics* **26**:1899-1900)
  Parallelization of the MAFFT multiple sequence alignment program.
  (describes the multithread version)
- Dudas, G. (2016). *baltic.* Retrieved from Github: https://github.com/evogytis/baltic
- Cock PA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B and de Hoon MJL (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25, 1422-1423
- ETE Toolkit. (n.d.). Analysis and Visualization of (phylogenetic) trees. Accessed February 09, 2021, from http://etetoolkit.org/

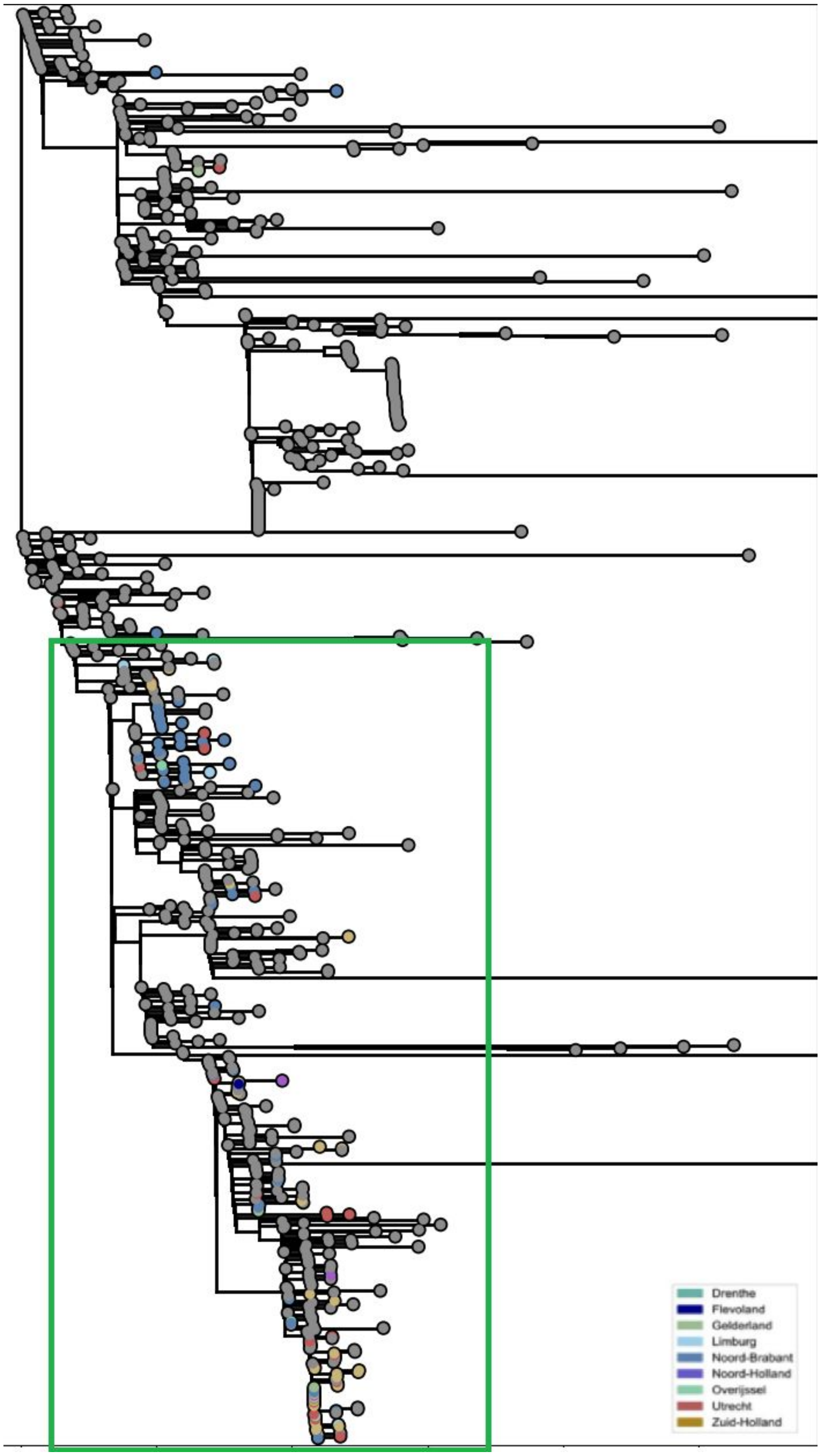
Appendices

A. Munnink’s, O., et al. (2020) whole generated phylogenetic tree using Maximum likelihood for sequences gathered on March 22, 2020



* Green box is what is shown on figure 1.

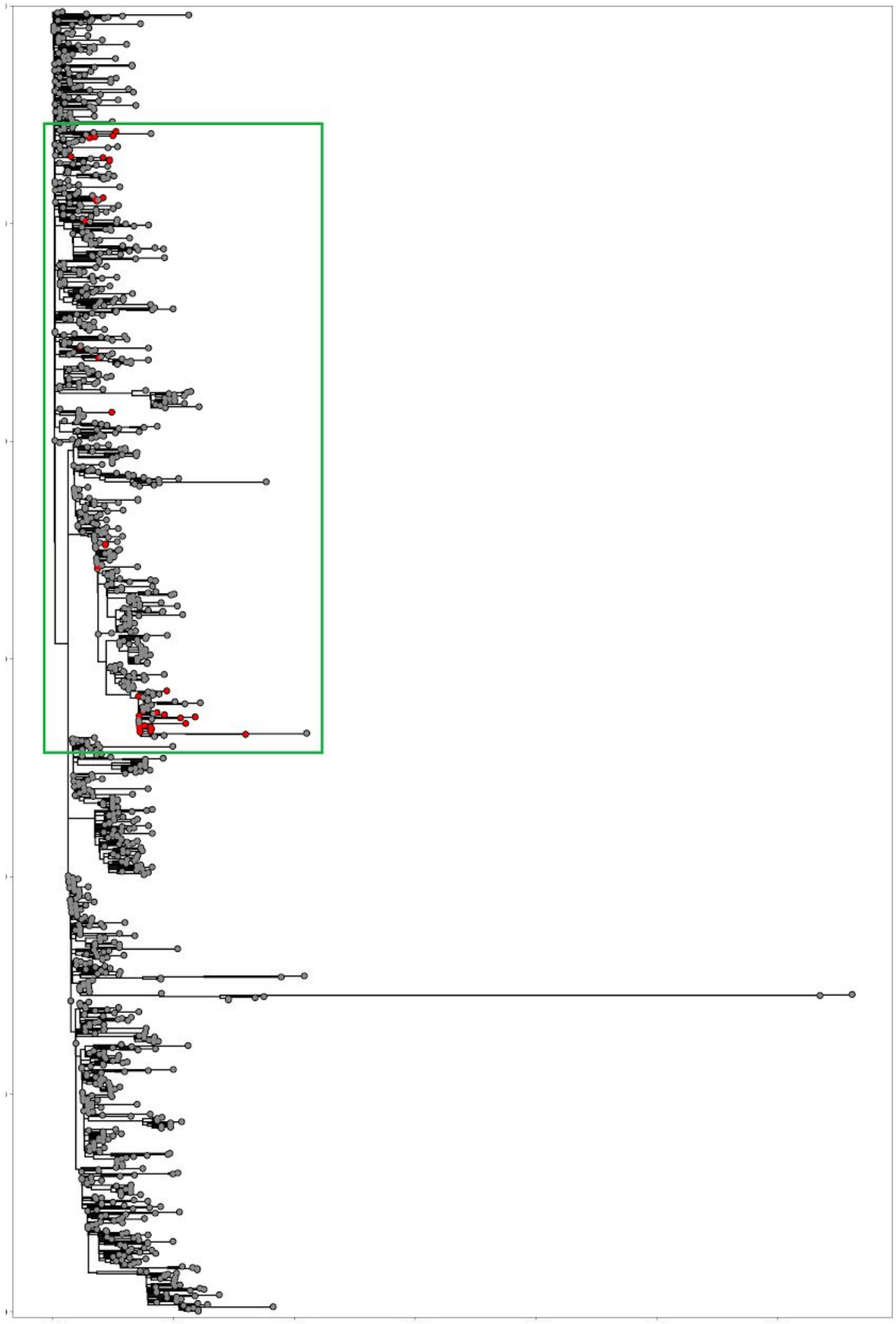
B. Our generated phylogenetic tree for sequences of Covid-19 cases from March 22, 2020 and earlier.



* Green box is what is shown on figure 2.

* Extremely long branches are cut, since it would collapse the view of the tree.

C. Phylogenetic tree using Maximum likelihood for sequences gathered 10 per country and 36 for Philippines



* The red points are the Philippine sequences and the gray points are sequences from other countries.

* Green box is what is shown on figure 3.