Developing a bioinformatics pipeline to detect Delta and Omicron co-infection cases in the Philippines during SARS-CoV-2 genomic surveillance

Adeliza Mae L. Realingo[1,*], Francisco Gerardo M. Polotan[1,*], Miguel Francisco B. Abulencia[1], Roslind Anne R. Pantoni[1], Jessel Babe G. Capin[1], Gerald Ivan Sotelo[1], Maria Carmen A. Corpuz[1], Neil Tristan M. Yabut[1], Saul M. Rojas[1], Amalea Dulcene Nicolasora[1], Ma. Angelica Tujan[1], Karen Iana Tomas[1], Ardiane Ysabelle Dolor[1], Czarina Christelle Alyannah Celis[1], Stephen Paul Ortia[1], Ezekiel Melo[1], Chelsea Mae Reyes[1], Elijah Miguel P. Flores[1], Anne Pauline A. Alpino[1], Aldwin Kim A. Penales[1], Kathlene Mae C. Medina[1], Joanna Ina Manalo[1], Lei Lanna Dancel[1], Timothy John R. Dizon[1], Sandeep Kasaragod[2,3], Katie Hampson[2], Joseph Hughes[3,*], Kirstyn Brunker[2,3,*]

[1] Advanced Molecular Technologies Laboratory, Research Institute for Tropical Medicine - Department of Health, Philippines
[2] School of Biodiversity, One Health & Veterinary Medicine, University of Glasgow, UK
[3] MRC-University of Glasgow Centre for Virus Research, UK

*These authors contributed equally to this work

ABSTRACT

Co-infection with two SARS-CoV-2 variants is rarely detected, but may have clinical implications and is hypothesized to be a precursor to SARS-CoV-2 variant recombination. It is therefore important to rapidly identify co-infection cases. Here, we report two cases of co-infection detected during routine genomic surveillance. Both samples were initially suspected to have been cross-contaminated due to the unusual number of private mutations identified by Nextclade and positions flagged for nucleotide mixtures by Bammix. To rule out cross-contamination, we decontaminated the workspaces and resequenced the samples. However, the anomalies remained and the samples were therefore investigated for co-infection. To validate this hypothesis, we created a bioinformatics pipeline to determine the lineages present in the samples, performed *de novo* assembly, and investigated the alternative allele fraction of mutations mapping to both variants. Freyja was used and showed an estimated lineage abundance of 49.81% and 48.09% for Delta and Omicron, respectively, in the first sample. For the second sample, Delta had 82.76% abundance and Omicron had 4.99% abundance, with the

presence of 12.24% abundance for other variants. Our findings support that the first sample is a case of co-infection of Delta and Omicron variants while the second sample is a cryptic co-infection of both variants, potentially undergoing recombination events. This study highlights the importance of developing a more stringent bioinformatics workflow to identify co-infection cases in genomic surveillance.

INTRODUCTION

Cases of COVID-19 in the Philippines have surpassed 4 million. Since the detection of the first case in March 2020, the country has experienced distinct epidemic waves marked by surges in COVID-19-related hospitalizations and deaths. The emergence of new variants has underpinned such epidemiological patterns globally, emphasizing the critical importance of genomic surveillance. The Philippines has emerged as a leading contributor to SARS-CoV-2 sequencing in Southeast Asia, demonstrating the country's commitment to strengthening its capacity for genomic surveillance of infectious diseases. This effort has contributed to an enhanced understanding of transmission patterns and the association of distinct SARS-CoV-2 lineages with subsequent phases of the epidemic in the Philippines (Tablizo et al, 2020; Li et al, 2022, Tablizo et al, 2022). Different waves of SARS-CoV-2 variants of concern (VOC) associated with increased transmissibility and immune escape have been identified. The Delta variant was first detected in May 2021 causing a surge in cases and prompted lockdowns and community quarantine throughout the country. Then, in December 2021, the Omicron variant was initially detected and soon dominated, displacing Delta. Among the currently circulating variants, Omicron is the most dominant, constituting 52.58% of the total samples sequenced in the country to date. Co-circulation of lineages possibly allowed recombination events, producing recombinant lineages that have been circulating more recently.

It has been hypothesized that co-infection with different variants may be precursors to recombination events. Co-infection cases are rarely reported but such cases can have important clinical and epidemiological significance and may contribute to increased disease severity and duration in patients and viral immune escape (Bolze et al, 2022; Shiraz & Tripathi, 2022). Epidemiological evidence in Andalusian genomic surveillance indicates a correlation between an increase in co-infection cases and a higher number of circulating recombinant variants. The

unexpectedly high number of co-infection cases led to the emergence of new recombinant variants (Perez-Florido et al., 2023). Detection of co-infection is probably underestimated (Combes et al, 2022), and it is, therefore, crucial to detect these cases during genomic surveillance as early as possible for clinical management and treatment strategies, elucidation of transmission dynamics and appropriate public health measures, and surveillance of genomic diversity and evolution.

To accurately detect co-infection cases, it is essential to consider several factors during the sequencing process. Primer amplification bias of some genomic regions of specific variants, contamination during the sequencing process (which might be the actual cause of nucleotide mixtures in the sample), and insufficient bioinformatics methods to systematically detect co-infections can all influence the ability to distinguish true co-infection cases from within-sample noise (Bal et al., 2022). The sample needs to be sufficiently sequenced at high depth to identify and characterize co-infection of different VOCs (Bolze et al., 2022). As more and more variants come into the picture, both sequencing procedures and bioinformatics tools need critical quality control measures to determine the presence of any co-infection, recombination, and contamination in the sequence data. There are limited tools for identifying and describing SARS-CoV-2 co-infections of the known variants.

To date, there have been no officially documented cases of co-infection in the Philippines. However, there is speculation that cases of co-infection in the Philippines have been occurring as early as March 2022 (Pangilinan et al, 2023). Here, we investigate two samples flagged as possible Delta and Omicron co-infections, due to the inconsistencies in lineage assignment and the presence of unusual private mutations, by routine genomic surveillance efforts at the Research Institute for Tropical Medicine (RITM). We develop a robust pipeline to diagnose co-infection cases and monitor their prevalence as part of ongoing genomic surveillance efforts.

RESULTS

As of writing, out of the 3,184 sequences processed by RITM for routine SARS-CoV-2 genomic surveillance, two samples were flagged for further quality control check due to inconsistencies in their lineage assignment with Pangolin tool and Nextclade (Supplementary

Fig. 1). For sample PH-RITM-1395, the consensus sequence had a coverage of 99.57% and was initially identified as B.1.1.529 (Omicron) by the Pangolin tool, while Scorpio determined the sample to be a probable Omicron. This was discordant with Nextclade's lineage assignment of B.1.617.2 (Delta). Usher placement noted that the sequence has insufficient support to be assigned to a specific lineage. Meanwhile, the sample PH-RITM-4146 had an initial coverage of 94.3% and was assigned by Pangolin as AY.1 (Delta) and by Nextclade as B.1.617.2 (Delta). The Scorpio call also determined the sample as Delta B.1.617.2-like with an additional K417N mutation, characteristic of a Delta plus variant (Table 1).

Table 1. Lineage assignment using PUSHER-1.2, Nextclade v2.14.1, and Scorpio v.0.3.17

Sample		Coverage	Pangolin UShER	Nextclade	Scorpio call	note
PH-RITM-1395	1st Sequencing	99.57%	B.1.1.529 (Omicron)	B.1.617.2 (Delta)	Probable Omicron (Unassigned)	Usher placements: B.1.1.529(1/1); scorpio found insufficient support to assign a specific lineage
	Replicate	91.88%	B.1.1.529 (Omicron)	B.1.617.2 (Delta)	Delta (B.1.617.2-like)	Usher placements: B.1.1.529(1/1); scorpio lineage B.1.617.2 conflicts with inference lineage B.1.1.529
PH-RITM-4146	1st Sequencing	94.3%	AY.1 (Delta)	B.1.617.2 (Delta)	Delta (B.1.617.2-like) +K417N	Usher placements: AY.1(2/4) AY.122(1/4) B.1.617.2(1/4)
	Replicate	99.5%	AY.1 (Delta)	B.1.617.2 (Delta)	Delta (B.1.617.2-like) +K417N	Usher placements: AY.1(2/4) AY.122(1/4) B.1.617.2(1/4)

Upon re-extracting and re-sequencing, we confirmed that the flagged private mutations persisted, thus ruling out the possibility of cross-contamination. The replicate consensus sequence for sample PH-RITM-1395 had a coverage of 91.88%, and its lineage assignment remained inconsistent between Pangolin and Nextclade, which showed B.1.1.529 (Omicron) and B.1.617.2 (Delta), respectively. Usher placements also highlighted a conflict in lineage assignment of B.1.617.2 and B.1.1.529. On the other hand, sample PH-RITM-4146 retained the lineage assignment for Pangolin and Nextclade as AY.1 (Delta) and B.1.617.2 (Delta), respectively. Usher placements indicated the sample to be either AY.1, AY.122, or B.1.617.2. Since we have ruled out cross-contamination as the source of the discrepancies, we conducted a more thorough investigation for possible co-infection of the two VOCs Delta and Omicron.

Multiple sites crossing the threshold of more than 20% mixture of nucleotide bases for both samples were flagged by bammix (Figure 1). We validated the nucleotide mixtures at each position by looking at the aligned sequence reads using Tablet. Notably, after sample

re-extraction and resequencing, the nucleotide mixtures remained in the replicates for both samples.

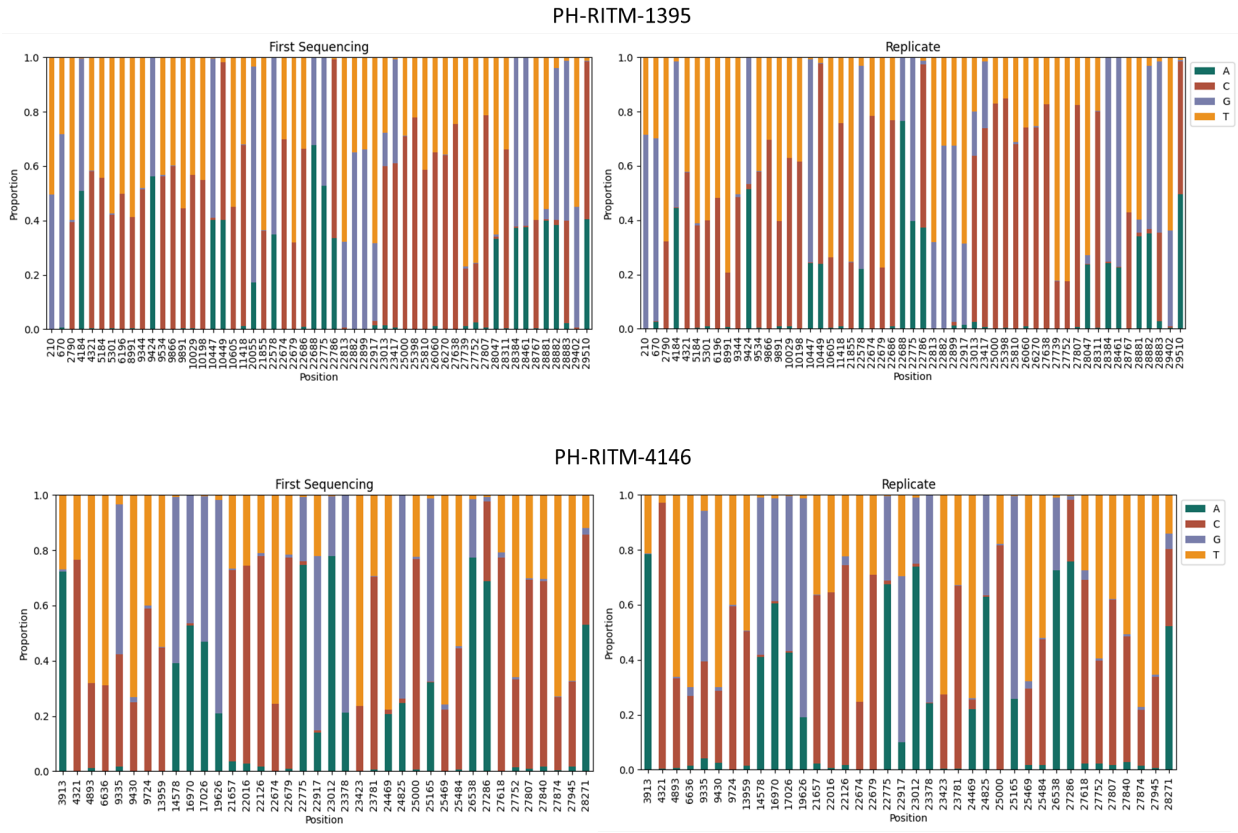

Figure 1. Positions flagged by bammix for nucleotide mixtures for each sample. All sites with single nucleotide variant (SNV) or single N's, or ambiguous base call, in the consensus assembly of samples of PH-RITM-1395 and PH-RITM-4146. Multiple sites containing heterogeneous bases crossing the threshold of >20% mixture in both replicates of each sample are apparent.

Observation of the nucleotide positions flagged by bammix and examination of overlapping SNVs and mutations support our hypothesis that the sample is co-infected with both Delta and Omicron. To confirm this, we further analyzed the sample using VirStrain, Freyja, and investigated the allele fraction of sequencing reads that mapped to mutations unique to each VOC.

Both samples were run against a custom VirStrain database created from the downloadable multi-fasta file of representative sequences from global nextregions in GISAID. For PH-RITM-1395, results from VirStrain showed to be consistent with the lineage assignment of Pangolin tool and Nextclade. VirStrain indicated the most possible strain as B.1.617.2 (Delta) and the other possible strain as BA.2 (Omicron). Meanwhile, for PH-RITM-4146, the most possible strain is AY.4 (Delta) and the other possible strain is BA.3 (Delta), discordant with the results from the lineage assignment tools.

We then used Freyja to test for the relative abundance of lineages present in the samples. For the initial sequence of PH-RITM-1395, the estimated lineage abundance of Delta and Omicron were 49.81% and 48.09%, respectively. The replicate sequence had a slight increase in Delta abundance, 56.83% as compared to Omicron with 42.55%. Other variants were detected by Freyja but at low levels. Meanwhile, the Freyja results for the initial sequence of sample PH-RITM-4146 showed that the lineage abundance of Delta and Omicron were 82.76% and 4.99%, respectively. Other variants, mostly recombinants, were found to comprise 12.24% of the sample. The replicate sequence had 80.89% and 7.23% abundance of Delta and Omicron, respectively, and other variants decreased slightly at 11.87%. These 'Other variants' include the following lineages: XAC and XZ, recombinants of Omicron lineages BA.2 and BA.1; XD, a recombinant of B.1.617.2 (Delta) and BA.1 (Omicron); and XAY, recombinant of AY.45 and BA.4/BA.5 (Figure 2, Supplementary figure 3).

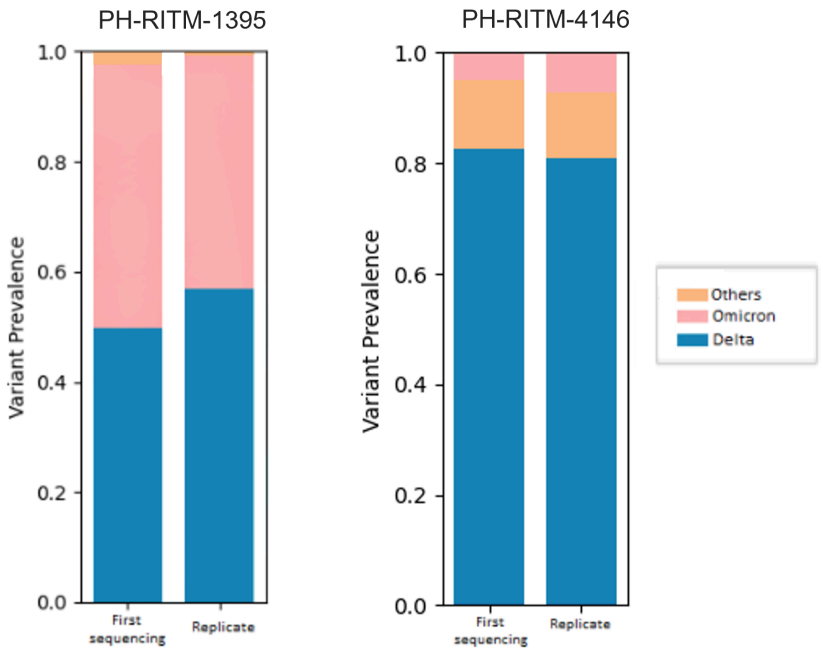

Figure 2. Variant frequency in the sample for the first and replicated sequence run plotted by Freyja tool. For sample PH-RITM-1395, abundances of Delta and Omicron are almost the same for the first and replicate sequence, and other variants were reported but at low levels. On the other hand, sample PH-RITM-4146 is composed more of Delta than that of Omicron, also, it is notable that the prevalence of other variants is greater than Omicron.

From cov-spectrum.org, we listed the established mutations that are shared and distinct to B.1.617.2 (Delta) and B.1.1.529 (Omicron). Delta has 74 lineage-defining mutations, while Omicron has 51. Additionally, there are four amino acid changes that are shared between the two (Supplementary Figure 2). Similar to the co-infection analysis visualizations described by Bolze et al (2022), we utilized custom python scripts illustrating the alternative allele fraction of the mutations across the SARS-CoV-2 genome of both samples. Only high-quality reads were used to create the allele fraction plots. Mutations shared between the two variants have an allele fraction equal to one, as expected, since the allele is found in all of the reads in that position.

Mixture of Delta and Omicron defining mutations is observed in PH-RITM-1395 supporting the hypothesis that this sample is a case of co-infection. The mixed alleles have fractions less than one and showed a pattern for co-infection. In contrast, sample PH-RITM-4146 have Delta and Omicron lineage defining mutations presented in distinct

sections, with allele fractions almost equal to one and a sudden drop of allele fraction in the Spike mutation N679K. This pattern is characteristic of recombinants according to Bolze et al (Figure 3).

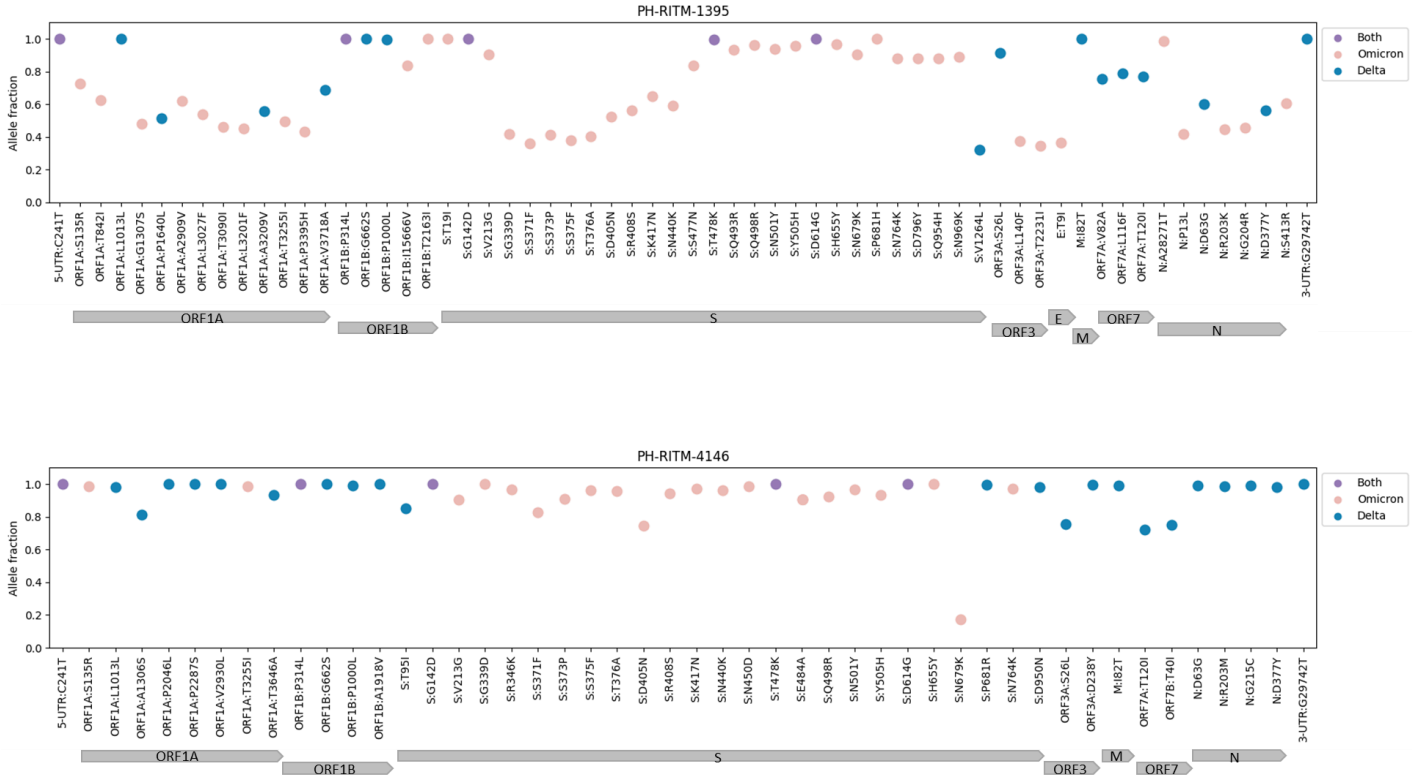

Figure 3. Alternative allele fraction of PH-RITM-1395 and PH-RITM-4146 for each mutation across the SARS-Cov-2 genome. Delta, Omicron, and shared mutations are represented by blue, pink, and purple, respectively. The Alternative Allele Fraction (AAF) plot for PH-RITM-1395 showed mixtures of Delta and Omicron specific mutations with fractions less than one. Meanwhile, the AAF plot of PH-RITM-4146 is characteristic of a recombinant variant due to its mutations appearing in sections rather than mixture.

We then determined the AAF for each amplicon by getting the average allele fraction for Delta and Omicron defining mutation within the range of that amplicon. PH-RITM-1395 shows that there is a mixture of the two VOCs for almost all amplicons supporting the pattern of co-infection. In contrast, PH-RITM-4146 plotted little mixture of the mutations on the amplicons and showed distinct sections within the SARS-CoV-2 genome containing amplicons for either Delta or Omicron only, supporting the pattern of recombinant variants (Figure 4).

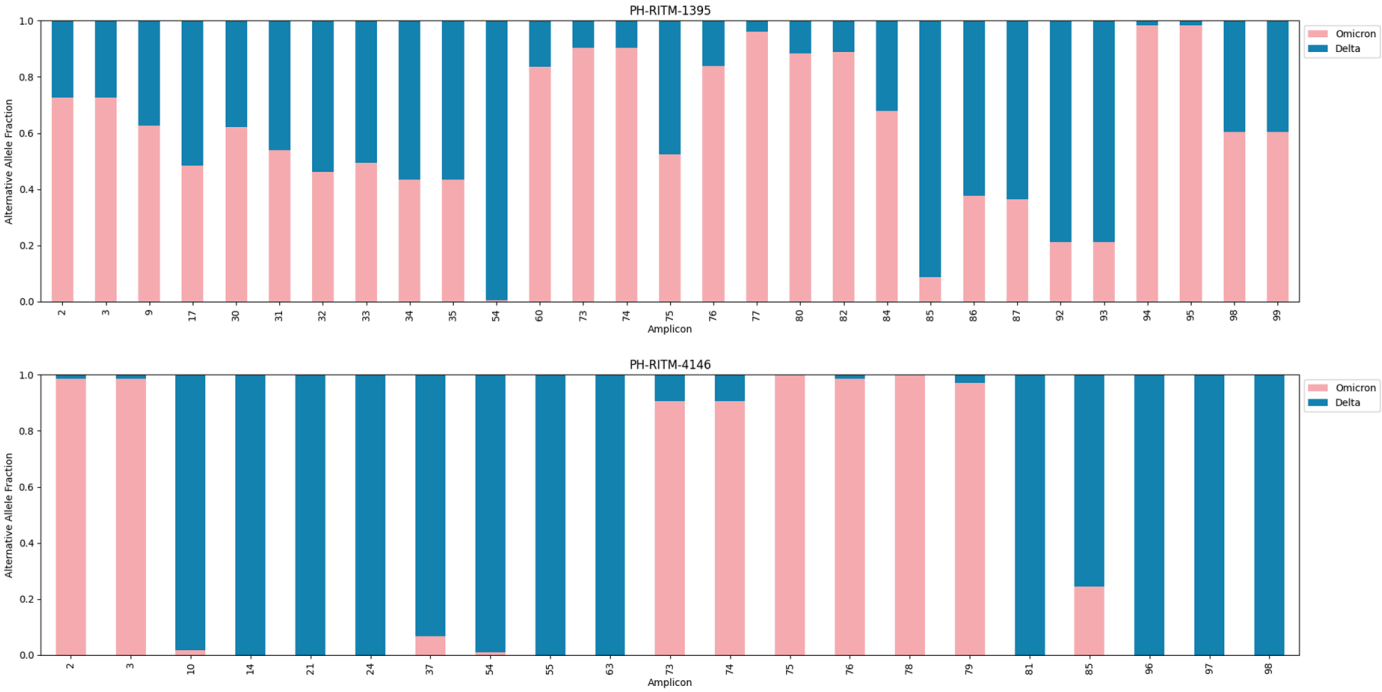

Figure 4. Alternative allele fraction per amplicon. By getting the average AAF of Delta and Omicron lineage defining mutations for each amplicon, we determined that PH-RITM-1395 showed a consistent pattern of co-infection while results for PH-RITM-4146 suggested that the sample is a recombinant.

Our *t*-test analysis is shown in Figure 5. Although there was a slight increase in variant abundance of Delta for the replicate sequence, we determined that there is no significant difference in the proportion of Delta and Omicron for both the first ($p = 0.73$) and replicate ($p = 0.69$) sequences. This supports the claim that PH-RITM-1395 had co-infection of both variants at the time of collection.

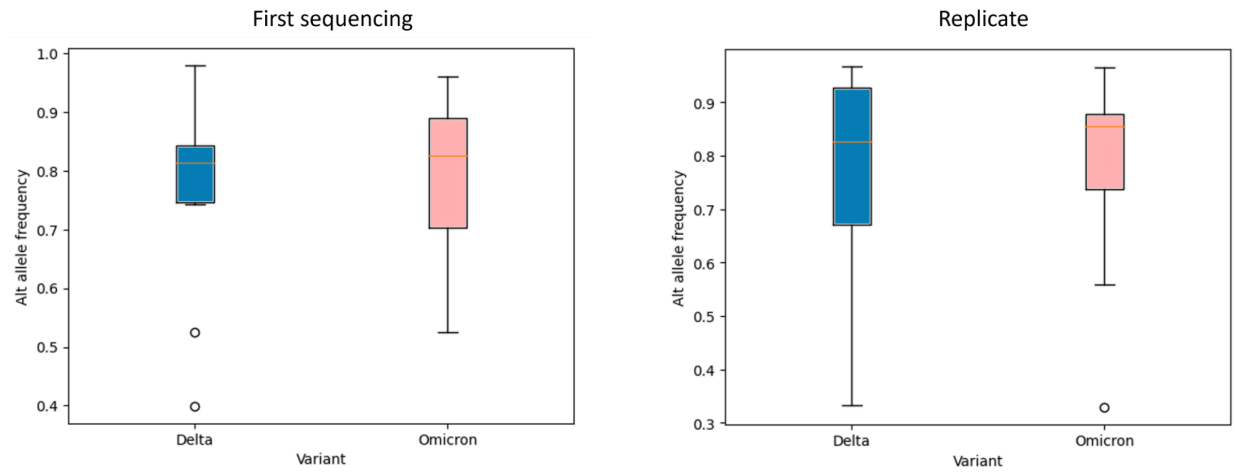

Figure 5. T-test for the read frequency distribution of Delta and Omicrons lineage markers of the first (P=0.73) and replicate (P=0.69) sequences showing no significant difference. Variant prevalence for Delta and Omicron remained almost similar after first and replicate sequencing.

Due to the allele fraction results for PH-RITM-4146, we considered the sample to be a recombinant instead of a co-infection case and subjected it to sc2rf tool. Results showed that the sample is a potential recombinant between Delta and Omicron as shown in Figure 6.

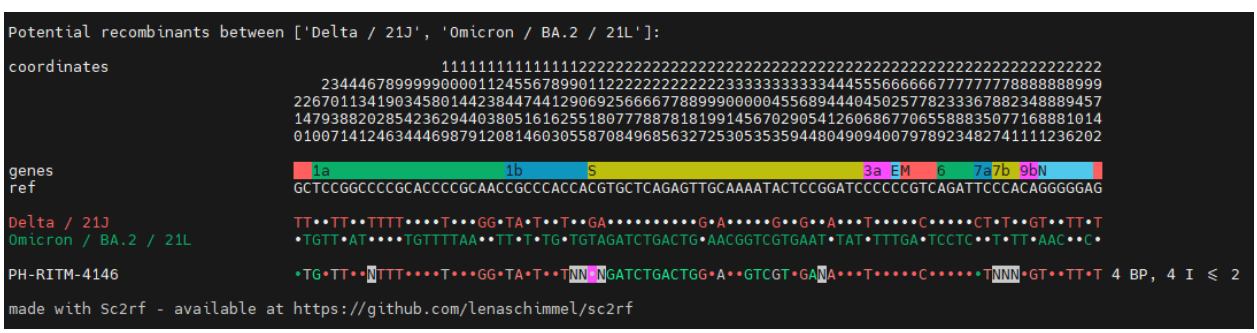

Figure 6. Screenshot of Sc2rf results showing that sample PH-RITM-4146 is a possible recombinant

Results from running custom python scripts and the tool samjdk in jvarkit showed the assigned lineages for the samples (Table 2). For sample PH-RITM-1395, the Pangolin tool assigned the separated consensus sequences as B.1.617.2 (Delta) and BA.2 (Omicron), while Nextclade assigned them as B.1.617.2 (Delta) and BA.2.3 (Omicron). These results are concordant with Virstrain.

Table 2. Lineage assignment after amplicon sorting

Sample	Pangolin UShER	Nextclade	Scorpio call	Note
PH-RITM-1395	B.1.617.2	B.1.617.2	Delta (B.1.617.2-like)	Usher placements: B.1.617.2(1/1)
	BA.2	BA.2.3	Probable Omicron (Unassigned)	Usher placements: BA.2(1/1)
PH-RITM-4146	AY.1	B.1.617.2	Delta (B.1.617.2-like) +K417N	Usher placements: AY.1(2/4) AY.122(1/4) B.1.617.2(1/4)
	B.1.1.529	B.1	Probable Omicron (Unassigned)	Usher placements: B.1.1.529(2/2); scorpio found insufficient support to assign a specific lineage

On the other hand, the sorted sequences of sample PH-RITM-4146 were assigned as AY.1 (Delta) and B.1.1.529 (Omicron) by Pangolin, while Nextclade assigned them as B.1.617.2 (Delta) and B.1 (Omicron). The lineage assignment of the Delta separated reads for PH-RITM-4146, is similar to the initial lineage assignment using Pangolin and Nextclade to be AY.1 and B.1.617.2, respectively.

We developed a pipeline compiling the steps we used to analyze the two samples flagged as possible Delta and Omicron co-infection cases. This pipeline can be accessed at https://github.com/lanadelrea/Katmon. The pipeline can generate a summary report including the results from lineage assignment, flags for nucleotide mixtures, and the alternative allele fraction plots.

DISCUSSION

Anomalies with the routine surveillance results of two samples, PH-RITM-1395 and PH-RITM-4146, were detected which led us to investigate further. Lineage assignment for the two samples was discordant, hence we looked into the possibilities of co-infection. At the time of routine genomic surveillance, there were limitations in detecting co-infection cases; thus, we pivoted to using additional bioinformatics tools for more stringent detection of nucleotide mixtures within the sample and we were able to rule out contamination.

In Australia, Rockett et al. were able to flag heterozygous calls for sampled sequences from epidemiologically unrelated patients using their bioinformatic quality control system (Rockett et al., 2022), similar to the findings that prompted further investigation in our study. In another study in Costa Rica, Molina-Mora et al. (2022), developed a metagenomic pipeline to

identify co-infections of distinct SARS-CoV-2 variants as part of their genomic surveillance efforts. A combination of different approaches was done to detect these co-infections, which may vary from case to case, as shown in our study as well.

This study investigated two cases that were flagged due to an egregiously high number of Nextclade QC flags and nucleotide positions with mixed bases. Other samples from genomic surveillance that appeared to be contaminated could actually be from co-infection cases but did not stand out during QC inspection of sequences and were not investigated closely. Out of more than 7,000 samples processed for genomic surveillance in the Philippines, by RITM and the Sub-national laboratories, it might take months to reprocess all the samples to have an estimate of the total number of co-infections present in the country. Currently, the computing resources at RITM are quite limited but work continues on improving QC procedures. Co-infections between the same lineage or very similar lineages may also not trigger QC flags or look like anomalies, thus not prompting a closer investigation.

In co-infection cases, the presence of both variants in a subset of host cells is a potential environment for recombination to occur (Bolze et al, 2022). Reported cases of co-infection during periods of variant co-circulation include a study from France, where Delta and Omicron co-infection was estimated at 0.18% over a 3-month period (Bal et al., 2022). Another study in France involving 3237 samples screened for VOC-specific SNPs observed 0.2% of samples co-infected with Delta and Omicron (Combes et al., 2022). In addition, a case of dual infection of other VOCs, namely Beta and Delta, was reported in Equatorial Guinea (Hosch et al., 2021). Building on this, our work demonstrates how the persistence and co-circulation of Delta alongside the BA.2 Omicron lineage have facilitated co-infection scenarios, providing the groundwork for the emergence of recombinant lineages.

Our work shows that the persistence and co-circulation of Delta during the dominance of the BA.2 Omicron lineage did result in co-infection, opening the possibility for recombination and emergence of recombinant lineages. A study by Jackson et al. (2021) described multiple independent recombination events involving UK lineages, highlighting the significance of co-infections during periods of high viral prevalence in driving the emergence of new recombinant lineages. In the Philippines, a recent study investigating the emergence of a recombinant lineage XBC described one sample with an allele fraction pattern and an abundance

of within-host mutations consistent with an active co-infection case (Pangilinan et al, 2023). Co-infection is a prerequisite for recombination, implying that co-infection cases must have been occurring in the Philippines potentially as early as March 2022. Genomic surveillance data from the country have also identified XBC sequences from as early as January 11, 2022 (EPI_ISL_16454618), during which Delta was being replaced by Omicron. Phylogenomic methods by Turakhia et al (2022) have identified 589 recombination events in 1.6 million samples, emphasizing the importance of timely detection and analysis of recombinant lineages for understanding genetic variation and viral evolution (Turakhia et al., 2022). This is further supported by our study, where a co-infection case between Omicron and Delta occurred after XBC emerged and was circulating, highlighting how co-infection and novel recombination events likely continue to occur at low levels of detection.

A similar study by Pisano et al. (2022), identified three samples with Delta and Omicron co-infection during the period in Argentina, December 2021 to January 2022, when both variants co-circulated. Cases of co-infection between Omicron and Delta have been found in many countries (Rockett et al, 2022; Bal et al, 2022; Combes et al, 2022; Zhou et al, 2022). Most of these however were from cases during the lineage replacement of Delta by Omicron from late 2021 to early 2022, whereas our case stands out as it was detected near the middle of 2022 in May, a point in time when Omicron had become dominant and Delta had been absent from sequencing runs for 8 months since August 21, 2021. Though detected in May, it is also possible that the patient had a chronic infection that started earlier. This may reflect the high relative transmissibility of both VOCs and how Delta continued to circulate in the background even when Omicron made up most of the infections. The latest Delta variant Philippine sequence available on GISAID, for example, was collected from Davao in January 2023 and submitted by the Philippine Genome Center. A small number of Delta variant sequences (n = 22) were still being collected in 2023 from countries in North America, Europe, Africa, and Asia (Khare et al, 2021).

Only two co-infection cases were detected among 7,259 total samples collected in the country during Delta and Omicron co-circulation from November 2021 to May 2022. Estimates of the prevalence of co-infections have varied, potentially due to a multitude of factors including the varying genetic diversity of SARS-CoV-2 lineages circulating at the time of study,

differences in transmission between different geographical locations, and methodological differences (Bal et al, 2022; Liu et al, 2021; Tonkin-Hill et al, 2021; Dezordi et al, 2022). For example, in Singapore from January to March 2020, a prevalence of 8% was reported among 131 patients in a cohort study (Liu et al, 2021). In France, co-infection with Delta and Omicron (BA.1), and BA.1 and BA.2 Omicron lineages, were estimated to have a lower prevalence of 0.18% and 0.26%, respectively, for the period December 2021 to February 2022 (Bal et al, 2022).

To rule out contamination during sample collection, collecting multiple samples from the same patient was advisable but it was not possible in our case. However, contamination seems highly unlikely to have occurred from the nasopharyngeal and oropharyngeal collection procedures. The repeat extraction and sequencing showed the same results and the negative extraction control showed no reads. The absence of reads in the negative control supports the conclusion that the detected genetic material is specific to the samples and not a result of contamination during the extraction process.

Sample collection was only performed at one point in time, as this was done prior to the development of our pipeline that automatically flags anomalous reads, so we were unable to observe any possible changes in relative abundance in the two variants over time. In a study by Pedro et al. 2021, co-infection by two SARS-CoV-2 lineages was detected in a patient wherein the frequency of the lineage changed over time, from an absolute 20A lineage at the time of diagnosis, to a 3% 20B lineage frequency after nine days, and then to a 100% 20B lineage frequency after two months. This indicates the importance of temporal sampling when it comes to monitoring the evolution of viral populations and might be a point for consideration in improving genomic surveillance.

The two cases presented here stood out as anomalies during routine genomic surveillance, triggering an investigation that confirmed co-infection. Genomic surveillance could be positioned to estimate the prevalence of co-infections and thus the risk of emerging recombinant lineages. Besides the risks of new recombinant variants, certain specific co-infections may sometimes have led to longer or more severe disease (Bal et al, 2022). If done in a timely manner, it may potentially inform diagnostics, patient care, and clinical outcomes.

To address this need, we have developed a bioinformatics pipeline designed to analyze the samples flagged as possible co-infection cases. This pipeline specifically reports on possible Delta and Omicron co-infections. While our current focus is on these variants, ongoing development aims to expand the pipeline's capabilities to include other major variants of concern of SARS-CoV-2.

CONCLUSIONS AND RECOMMENDATIONS INCLUDING FUTURE DIRECTIONS

We have identified and characterized two co-infection cases from a routine SARS-CoV-2 genomic surveillance. The first sample was likely a result of the co-circulation of Omicron with persistent Delta lineage even after Omicron had replaced Delta as the dominant VOC in the country. Meanwhile, the second sample was a probable case of cryptic co-infection of Delta and Omicron, potentially undergoing recombination events. Recombinants were highly circulating at the time of sample collection. We eliminated the possibility that they were a product of cross-contamination by first decontaminating the workspaces and then resequencing the samples. The negative controls were clean after resequencing the sample. This is the first report of co-infection in the country, and we have developed a pipeline to summarize the steps used to analyze the samples efficiently, detecting any possible cases of co-infections. The pipeline can be accessed at https://github.com/lanadelrea/Katmon. It will output a summary html report containing the lineage assignment, flagged positions with nucleotide mixtures, and alternative allele fraction plots. Investigation of recent and past samples would be interesting to look into, especially in less-sampled regions of the country, to determine if the Delta lineage persisted through cryptic local transmission even after its replacement by Omicron. To better estimate co-infection prevalence, a retrospective investigation of past samples can be done and may reveal co-infection cases that were not previously identified. Prospective detection of co-infection cases may be worth pursuing during routine genomic surveillance given that they are precursors to recombination, which can generate new SARS-CoV-2 variants and have been suggested to affect patient outcomes.

MATERIALS AND METHODS

Sample collection, RNA inactivation and extraction

Samples obtained from naso- and oro-pharyngeal swabs of COVID-19 positive patients were endorsed by the Epidemiology Bureau (EB) to the Research Institute for Tropical Medicine (RITM) for sequencing as part of the routine SARS-CoV-2 genomic biosurveillance in the Philippines. RNA was inactivated and extracted using MagMax™ Viral/Pathogen Nucleic Acid Isolation Kit (A42352, ThermoFisher) in the Kingfisher flex purification system. This protocol used 200 uL of clinical specimen and a final elution volume of 80 uL.

As part of the sequencing protocol, the identifiers used for each sequence were assigned in the laboratory during processing and cannot be directly traced to patients from whom the samples were collected. Among all the samples, we further investigated two sequences, PH-RITM-1395 and PH-RITM-4146, as possible co-infection cases because of their inconsistent lineage assignment and unusual number of private mutations flagged by Nextclade. The first sample was collected in May 2022 in Zamboanga City, Zamboanga del Sur while the second sample was collected in April 2023 in Daraga, Albay. Since the specimens were endorsed by EB, no serology or RT-qPCR testing for other respiratory pathogens were performed. As part of the acceptance criteria for sequencing, the Ct value of the specimens were recorded. Low Ct values of 19.4 for N gene and 17.2 for E gene were obtained for the first sample, and Ct values 30.82 for N gene and 19.58 for ORF were obtained for the second sample, indicating suitability for whole-genome sequencing (WGS).

Whole-genome sequencing using Oxford Nanopore Technology (ONT)

Sequencing libraries were prepared using the ARTIC LoCost sequencing protocol (Quick, 2020). The cDNA was prepared by reverse transcription using LunaScript® RT SuperMix Kit (E3010, NEB) and ARTIC primers v4+4.1 were used with Q5® Hot Start High-Fidelity 2X Master Mix (M0494, NEB) to amplify the whole genome in a multiplex PCR reaction, followed by purification of the DNA using 0.9X AMPure XP beads (A63881, Beckman Coulter). Sample concentration was checked using a Qubit 4 fluorometer, with acceptable concentrations of >10 ng/ul. To prepare for ONT whole-genome sequencing the amplicons were library prepared and barcoded using Ultra II reagents (E7546 and E7595, NEB) and ONT Native Barcoding Kit (EXP-NBD196), cleaned up with 0.4X AMPure XP beads, and the adapters

ligated using the NEBNext Quick Ligation Module (E6056, NEB). The final libraries were prepared using the ONT Ligation Sequencing Kit (SQK-LSK109) and loaded into an R9.4.1 flow cell (FLO-MIN106D, ONT). Sample PH-RITM-1395 was sequenced and re-sequenced on a GridION with high accuracy basecalling for 72 hours, while sample PH-RITM-4146 on a MinION Mk1B device with fast basecalling for 48 hours.

Detection of SARS-CoV-2 variant of concern and pango-lineage

Following the RITM routine process for SARS-CoV-2 genomic surveillance, we used the ARTIC bioinformatics standard protocol for SARS-CoV-2 implemented in the ncov2019-artic-nf nextflow workflow (Artic Network, 2020) to generate a consensus sequence and assign a pango-lineage to the samples. Basecalling of raw ONT fast5 data was performed on the GridION machine using guppy software version 5.1.13. Raw fast5 and fastq files were transferred to the RITM high-performance computing (HPC) server. The workflow performs demultiplexing of samples by barcodes, read filtering by length of 400-700 bp, and mapping of basecalled reads to raw fast5 signal data. Fastq reads are then aligned to the SARS-CoV-2 reference genome using minimap2. Primers are trimmed from aligned reads, and sequences of aligned reads falling outside primer boundaries are soft-clipped. Variants are called using nanopolish for each primer pool separately. Variant calls are merged into a single VCF file, and then filtered with the criteria of >20x depth and a minimum of 50% supporting reads from each strand. Regions of the genome covered with <20x depth are masked, and variant calls are introduced to the regions with >20x depth to produce the consensus sequence. Pangolin with default UShER as lineage assignment engine was used to assign the consensus sequence to a SARS-CoV-2 lineage and Nextclade was used for quality control. As part of genomic surveillance, all the sequences having greater than 70% genome coverage, including the samples investigated in this study, were shared publicly to the Global Initiative for Sharing All Influenza Data (GISAID).

Detection of Omicron and Delta co-infection

A close examination of the sequence with anomalous quality control metrics was performed, including inspection of the quality of the raw reads, to the consensus sequence, variant calls, and read-level alignment. FastQC was used to evaluate the quality of basecalled

Fastq reads. The Nextclade web server was used to evaluate the quality of the consensus sequence, including metrics such as percent coverage, depth, unexpected private mutations, frameshifts, and stop codons. The bammix tool was used to visualize and identify all the genome positions with heterogeneous nucleotide bases from aligned reads, i.e. not a single dominant nucleotide base but a mixture. The covSPECTRUM and CoVariants web resources were used to identify the list of single nucleotide variants (SNVs) that were distinct between the mixed lineages suspected to be present in the same sample. Multiple SNVs that were located near enough each other and could be spanned by reads that are 400-700 bp long were inspected in the alignment BAM file using Tablet. Individual reads in Tablet that contained the expected SNVs of one of the two lineages but not the other were identified as supporting evidence that multiple lineages, and not a single recombinant lineage, was present in the sample. Finally, VirStrain and Freyja are tools that are able to disentangle and quantify mixed lineages in a sample. We used VirStrain and Freyja to confirm the findings from manual inspection of aligned reads, assign more specific lineages, and quantify the abundance of each lineage in the sample.

The Cq value for these samples were low implying high source viral load that would be less susceptible to contaminaton. We replicated these results by re-extracting RNA from the collected sample and re-sequencing.

We then looked at the alternative allele fraction to determine the number of reads supporting each VOC. AAF is the ratio between the number of reads supporting an alternative allele or mutation and the total number of reads covering the corresponding position. We illustrated the average AAF of Delta and Omicron per amplicon. Lastly, using jvarkit, we separated the Omicron and Delta reads and created consensus sequences and determined their lineage assignment.

DATA AVAILABILITY

All sequences processed for Philippine SARS-CoV-2 genomic surveillance, including those used in this study, have been uploaded to and are publicly available on the EpiCoV database of the Global Initiative for Sharing All Influenza Data (GISAID). Batches containing the flagged samples analyzed in this study are accessible in GISAID under the compiled EPI

SET (EPI_SET_240703bc), as detailed in Supplementary Table 1, and in Bioproject PRJNA1124409 on the Sequence Read Archive (SRA).

ACKNOWLEDGEMENTS

The authors would also like to acknowledge Aldrin V. Imbag and Fatima Rose S. Guemo for providing and maintaining the high-performance computing resources that enabled the assembly and analysis of data, and Charalyn A. Babida for managing the metadata of all samples during genomic surveillance. We extend our gratitude to the Department of Health Epidemiology Bureau of the Philippines, the COVID-19 Surveillance and Quick Action Unit, the Regional Epidemiology Surveillance Units, and our partner Sub-National Laboratories; Baguio General Hospital and Medical Center (BGHMC), Lung Center of the Philippines (LCP), San Lazaro Hospital (SLH), Vicente Sotto Memorial Medical Center (VSMMC), and Southern Philippines Medical Center (SPMC) for their invaluable support in the nation's biosurveillance efforts.

FUNDING

The work was supported by the UK Research and Innovation Global effort on COVID-19 (MR/V035444/1) and Wellcome (207569/Z/17/Z, 224520/Z/21/Z). J.H. is funded by the Medical Research Council (MC_UU_12014/12) and KB is supported by a Medical Research Council New Investigator Research Grant (MR/X002047/1) and a University of Glasgow Lord Kelvin/Adam Smith Fellowship.

REFERENCES

Bal, A., Simon, B., Destras, G., Chalvignac, R., Semanas, Q., Oblette, A., Quéromès, G., Fanget, R., Regue, H., Morfin, F., Valette, M., Lina, B., & Josset, L. (2022). Detection and prevalence of SARS-CoV-2 co-infections during the Omicron variant circulation in France. *Nature Communications*, *13*(1), 6316. https://doi.org/10.1038/s41467-022-33910-9

Bolze, A., Basler, T., White, S., Dei Rossi, A., Wyman, D., Dai, H., Roychoudhury, P., Greninger, A. L., Hayashibara, K., Beatty, M., Shah, S., Stous, S., McCrone, J. T., Kil,

E., Cassens, T., Tsan, K., Nguyen, J., Ramirez, J., Carter, S., … Luo, S. (2022). Evidence for SARS-CoV-2 Delta and Omicron co-infections and recombination. *Med*, *3*(12), 848-859.e4. https://doi.org/10.1016/j.medj.2022.10.002

Jackson, B., Boni, M.F., Bull, M.J., Colleran, A., Colquhoun, R.M., Darby, A.C., Haldenby, S., Hill, V., Lucaci, A., McCrone, J.T., Nicholls, S.M., O'Toole, Á., Pacchiarini, N., Poplawski, R., Scher, E., Todd, F., Webster, H.J., Whitehead, M., Wierzbicki, C. (2021) COVID-19 Genomics UK (COG-UK) Consortium; Loman NJ, Connor TR, Robertson DL, Pybus OG, Rambaut A. Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic. Cell. Sep 30;184(20):5179-5188.e8. doi: 10.1016/j.cell.2021.08.014. Epub 2021 Aug 17. PMID: 34499854; PMCID: PMC8367733.

Khare, S., et al (2021) GISAID's Role in Pandemic Response. *China CDC Weekly*, 3(49): 1049-1051. doi: 10.46234/ccdcw2021.255 PMCID: 8668406

Molina-Mora, J. A., Cordero-Laurent, E., Calderón-Osorno, M., Chacón-Ramírez, E., & Duarte-Martínez, F. (2022). Metagenomic pipeline for identifying co-infections among distinct SARS-CoV-2 variants of concern: Study cases from Alpha to Omicron. *Scientific Reports*, *12*(1), 9377. https://doi.org/10.1038/s41598-022-13113-4

Pangilinan, E. A. R., Egana, J. M. C., Mantaring, R. J. Q., Telles, A. J. E., Tablizo, F. A., Lapid, C. M., Yangzon, M. S. L., Endozo, J. J. S., Padilla, K. S. A. R., Nipales, J. E., Carandang, L. C. D. L., Enriquez, Z. M. R., Barot, T. A. U., Manlimos, R. A., Mangonon, K. N. P., Plantig, Ma. E. L., Araiza, S. M. M., Llames, J.-H. S., Punayan, K. P., … Saloma, C. P. (2023). *Analysis of SARS-CoV-2 Recombinant Lineages XBC and XBC.1 in the Philippines and Evidence for Delta-Omicron Co-infection as a Potential Origin* [Preprint]. Genetics. https://doi.org/10.1101/2023.04.12.534029

Pedro, N., Silva, C.N., Magalhães, A.C., Cavadas, B., Rocha, A.M., Moreira, A.C., Gomes, M.S., Silva, D., Sobrinho-Simões, J., Ramos, A., et al. (2021) Dynamics of a Dual SARS-CoV-2 Lineage Co-Infection on a Prolonged Viral Shedding COVID-19 Case: Insights into Clinical Severity and Disease Duration. *Microorganisms*. 9(2):300. https://doi.org/10.3390/microorganisms9020300

Perez-Florido, J., Casimiro-Soriguer, C. S., Ortuño, F., Fernandez-Rueda, J. L., Aguado, A., Lara, M., Riazzo, C., Rodriguez-Iglesias, M. A., Camacho-Martinez, P., Merino-Diaz,

L., Pupo-Ledo, I., de Salazar, A., Viñuela, L., Fuentes, A., Chueca, N., The Andalusian COVID-19 Sequencing Initiative, García, F., Dopazo, J., & Lepe, J. A. (2023). Detection of High Level of Co-Infection and the Emergence of Novel SARS CoV-2 Delta-Omicron and Omicron-Omicron Recombinants in the Epidemiological Surveillance of Andalusia. *International Journal of Molecular Sciences*, *24*(3), 2419. https://doi.org/10.3390/ijms24032419

Pisano, M.B., Sicilia, P., Zeballos, M., Lucca, A., Fernandez, F., Castro, G.M., Goya, S., Viegas, M., López, L., Barbás, M.G. and Ré, V.E. (2022) SARS-CoV-2 Genomic Surveillance Enables the Identification of Delta/Omicron Co-Infections in Argentina. *Front.Virol.* 2:910839. doi: 10.3389/fviro.2022.910839

Quick, J. (2020). nCoV-2019 sequencing protocol v3 (LoCost) V.3. Retrieved from dx.doi.org/10.17504/protocols.io.bp2l6n26rgqe/v3

Rockett, R. J., Draper, J., Gall, M., Sim, E. M., Arnott, A., Agius, J. E., Johnson-Mackinnon, J., Fong, W., Martinez, E., Drew, A. P., Lee, C., Ngo, C., Ramsperger, M., Ginn, A. N., Wang, Q., Fennell, M., Ko, D., Hueston, L., Kairaitis, L., … Sintchenko, V. (2022). Co-infection with SARS-CoV-2 Omicron and Delta variants revealed by genomic surveillance. *Nature Communications*, *13*(1), 2745. https://doi.org/10.1038/s41467-022-30518-x

Turakhia, Y., Thornlow, B., Hinrichs, A. *et al.* (2022). Pandemic-scale phylogenomics reveals the SARS-CoV-2 recombination landscape. *Nature* 609, 994–997. https://doi.org/10.1038/s41586-022-05189-9

Zhou, H.-Y., Cheng, Y.-X., Xu, L., Li, J.-Y., Tao, C.-Y., Ji, C.-Y., Han, N., Yang, R., Li, Y., & Wu, A. (2021). *Genomic evidence for divergent co-infections of SARS-CoV-2 lineages* [Preprint]. Microbiology. https://doi.org/10.1101/2021.09.03.458951