

Title

'Vivaldi': An amplicon-based whole genome sequencing method for the four seasonal human coronaviruses 229E, NL63, OC43 & HKU1, alongside SARS-CoV-2'

Authors

C. Patrick McClure^{1,2,6*}, Theocharis Tsoleridis^{1,2,3}, Nadine Holmes^{2,4}, Joseph G. Chappell^{1,2}, Timothy Byaruhanga^{1,2}, Joshua Duncan^{1,2}, Miruna Tofan², Abdul Khater², Louise Berry⁵, Gemma Clark⁵, William L. Irving^{1,2,5,6}, Alexander W. Tarr^{1,2,6}, Jonathan K. Ball^{1,2,6,8}, Stuart Astbury^{6,7,*}, Matt Loose^{2,4,*}

Affiliations

¹ Wolfson Centre for Global Virus Research, University of Nottingham, Nottingham, UK

² School of Life Sciences, University of Nottingham, Nottingham, UK.

³ The Pirbright Institute, Ash Road, Pirbright, Woking, UK

⁴ Deep Seq, School of Life Sciences, University of Nottingham, Nottingham, UK

⁵ Clinical Microbiology, Nottingham University Hospitals NHS Trust, Nottingham, UK

⁶ National Institute for Health Research Nottingham Biomedical Research Centre, University of Nottingham and Nottingham University Hospitals NHS Trust, Nottingham, UK

⁷ Nottingham Digestive Diseases Centre, Translational Medical Sciences, School of Medicine, University of Nottingham, Nottingham, UK.

⁸ Dept of Tropical Disease Biology, Liverpool School of Tropical Medicine, Liverpool, UK

*Correspondence to:

C. Patrick McClure Patrick.mcclure@nottingham.ac.uk

Stuart Astbury stuart.astbury@nottingham.ac.uk

Matt Loose matt.loose@nottingham.ac.uk

Abstract

Prior to the emergence of SARS-CoV-2 in 2019, Alphacoronaviruses 229E and NL63 and Betacoronaviruses OC43 and HKU1 were already established endemic 'common cold' viral infections. Despite their collective contribution towards global respiratory morbidity and mortality and potential to inform the future trajectory of SARS-CoV-2 endemicity, they are infrequently sequenced. We therefore developed a 1200bp amplicon-based whole genome sequencing scheme targeting all four seasonal coronaviruses and SARS-CoV-2.

The 'Vivaldi' method was applied retrospectively and prospectively using Oxford Nanopore Technology to approximately 400 seasonal coronavirus infections diagnosed in Nottingham, UK, from February 2016 to July 2023. We demonstrate that the amplicon multiplex strategy can be applied agnostically to determine complete genomes of five different species from two coronaviral genera. 304 unique seasonal coronavirus genomes of greater than 95% coverage were achieved: 64 for 229E, 85 for NL63, 128 for OC43 and 27 for HKU1. They collectively indicated a dynamic seasonal coronavirus genomic landscape, with co-circulation of multiple variants emerging and declining over the UK winter respiratory infection season, with further geographical distinction when compared to a global dataset. Prolonged infection with concomitant intra-host evolution was also observed for both Alpha- (NL63) and Betacoronaviruses (OC43).

This data represents the largest single cohort of seasonal coronavirus genomes to date and also a novel amplicon scheme for their future global surveillance suitable for widespread and easy adoption in the post-SARS-CoV-2 era of viral genomics.

Introduction

The *Coronaviridae* family belongs to the *Nidovirales* order of positive single-stranded RNA viruses with relatively large genomes (27-33 kb) and four genera (*Alpha-*, *Beta-*, *Gamma-* and *Deltacoronavirus*) within the Orthocoronavirinae subfamily, with Alpha- and Betacoronavirus genera members capable of human infection (1, 2).

Prior to the emergence of SARS-CoV-2 in 2019, four other Coronaviruses – Alphacoronaviruses 229E, NL63 and Betacoronaviruses OC43 and HKU1 – had previously established endemic status (3-6). Collectively these are responsible for a significant proportion of global respiratory infections as part of the informal ‘common cold’ viral grouping (7). Whilst they present a broad spectrum of disease severity in all age groups, infections are typically but not exclusively symptomatically mild (8), which has in part led to their clinical and genomic under-investigation to the detriment of understanding the emergence and future trajectory of SARS-CoV-2 endemicity (7, 9, 10).

To date, six 229E genotypes (1 – 6) and three for NL63 (A-C) (11, 12) have been described for the Alphacoronaviruses, whilst Betacoronaviruses OC43 and HKU1 have been categorised into 10 (A-K) and 3 (A to C) genotypes, respectively (12-16). Genetic diversity is also generated by extensive recombination events within and between genera (11, 15, 17).

Amplicon-based Whole Genome Sequencing (WGS), utilising contiguous overlapping PCR products of various size, has previously been described for a wide range of viruses (18-24), yielding both population-level epidemiological insight and outbreak management with near-realtime capability (19). There is a trade off in the priming strategy between the increased sensitivity of smaller amplicon tiling versus the heightened chance of primer mismatch where greater number of primers are used. Greater diversity in the targeted viral taxon, as observed in established endemic pathogens, creates a more challenging target preferentially necessitating fewer amplicons to span the genome (18, 24). Continuous viral evolution, especially noticeable during the early phases of rapid viral spread following a recent zoonotic spillover can result in primer mismatches, necessitating careful surveillance for amplicon ‘drop-out’ and swift primer redesign (25).

Due to the perceived low medical threat posed by seasonal coronaviruses their inclusion in diagnostic panels and molecular epidemiological surveillance programmes has been overlooked in some instances. Since the emergence of SARS-CoV-2, there is increasing interest in the seasonal coronavirus molecular and clinical epidemiology (10, 26-28).

This study describes an amplicon-based approach to provide whole genome sequences of the four seasonal coronaviruses alongside SARS-CoV-2. We have applied the methodology both retrospectively to a large cohort of archival extracts pre-SARS-CoV-2 emergence and prospectively to contemporary post-pandemic infections. The resulting sequence data provides greater depth of insight into seasonal coronaviral variation, potentially informing future SARS-CoV-2 endemicity. Contemporary viral genomes and future uptake of the methodology will facilitate further understanding of seasonal coronavirus evolution in a post-pandemic immune landscape.

Methods

Samples

Surplus total nucleic acid from anonymised Coronaviral-positive patient respiratory samples diagnosed as part of the routine care pathway at Nottingham University Hospitals NHS Trust (NUH NHST) was stored at -70°C from February 2016 to December 2018 and May 2021 to July 2023 as previously described (29, 30). Prior to May 2021, the AusDiagnostics Respiratory 16-plex clinical diagnostic panel used could discriminate seasonal coronavirus type, but this was not always recorded by laboratory personnel, nor was the absolute quantitative template copy number. Due to potential clinical confusion and focussing of diagnostic resources, seasonal coronaviruses were not routinely

investigated at the onset of the SARS-CoV-2 Pandemic between February 2020 and April 2021, where after no diagnostic distinction was made between 229E, NL63, OC43 & HKU1 with a newly configured AusDiagnostics respiratory multiplex RT-qPCR in use. Available nucleic acid with a recorded seasonal coronavirus type and a diagnostic laboratory quantitation of greater than 10 copies per 10ul RNA were selected from the 2016 to 2018 archive, with the exception of HKU1, where all positives were tested. Subsequently, all available diagnosed seasonal coronavirus positive extracts were investigated from the 2021 to 2023 sub-cohort. Anonymised diagnostic laboratory PCR results were curated and analysed in Excel. NUH NHST approved extended molecular investigation of diagnosed coronavirus positives under clinical audit number 23-078C.

Amplicon scheme primer design

All seasonal coronavirus genomes deposited in GenBank by July 2021 were downloaded, aligned as below, and used to construct maximum likelihood trees, from which 5 distinct lineage sequences were chosen and concatenated with different SARS-CoV-2 variant sequences to generate 5 templates for PrimalScheme (21). PrimalScheme was then instructed to generate primer sequences targeting 1200bp regions of the concatenated template, covering all 4 seasonal coronavirus and SARS-CoV-2 genomes. The outputted 278 primer scheme was manually inspected in MEGA7 for mismatch against the alignment of all available genomes and accordingly edited with degeneracy or redesigned where 3' mismatch was observed in a significant proportion of the global dataset. Amplicon coverage was subsequently reviewed with each sequencing run, redesigning primers where amplicon drop-out or excessively lower coverage was observed. Amplicon balancing was further attempted in a minority of instances by doubling or halving primer concentration where coverage remained significantly low or high respectively. Final selected primer sequences and relative concentrations are listed in supplementary table 1.

cDNA synthesis, PCR & whole genome sequencing

Coronaviral cDNA was prepared with RNA to cDNA EcoDry™ Premix containing random hexamers (Takara Bio) as per the manufacturer's instruction, then processed similarly to previously described (20). Briefly, up to 2.5 µl of cDNA was used as template in a 25 µl PCR reaction assembled with Q5® High-Fidelity 2X Master Mix (New England Biolabs) and either primer set 1 (odd numbered primer pairs) or 2 (even numbered primer pairs) of either targeting individual coronaviral species or all 5 combined at a final concentration of 0.015 µM unless stated otherwise (Supplementary table 1). PCR reactions were thermocycled as follows: 98°C for 30 seconds then 45 cycles of 98°C for 15 seconds and 65°C for 5 minutes. PCR products from reaction 1 and 2 were inspected for specificity and yield on a 2% agarose gel with ethidium bromide, and subsequently combined before Qubit quantification and normalisation to 100ng of DNA in 7.5 µl water. Amplicons were prepared and barcoded using the SQK-LSK109 and EXP-NBD196 kits respectively (Oxford Nanopore), and sequenced on an Oxford Nanopore GridION as previously described (19).

Sequence and phylogenetic analysis

Following basecalling using Guppy (v6.5.7), demultiplexed reads passing the quality threshold (average score >7) were used as the input for the ARTIC pipeline (21). Briefly, reads were filtered to remove those below 700bp and above 1400bp, and aligned to a reference.fasta containing the 4 seasonal coronaviruses and SARS-CoV-2 using Minimap2 (31). A .bed file corresponding to the primers used in the scheme was then used to softmask resulting alignments via the ARTIC align_trim script to ensure variants were not called in primer sites. Finally, the softmasked alignment was used as the input for nanopolish (32) to generate a consensus sequence which was then aligned to the reference using MUSCLE (33). Coverage was based on a 20x individual amplicon threshold with those

dropping below this classified as dropouts. The reference .fasta and scheme .bed file are available on request in advance of submission to GenBank.

Genomes were aligned for the presented analysis using the Geneious Prime 2019.0.4 software with the relevant seasonal coronavirus species genomes deposited in GenBank by December 2023. Maximum-Likelihood trees were generated to visualise the evolutionary relationships between high quality (>95% complete) study and publicly available genomes with IQ-TREE2 using the following models of evolution as suggested by the software's model finder respectively: 229E - TIM+F+R2; NL63 - TIM+F+R3; OC43 - TIM+F+R6; HKU1 - TN+F+R2, with 1000 bootstraps of the Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT). Seasonality was assigned for all sequences with >95% coverage, using the first instance of infection in serially sampled patients only. Core season was defined as October to May (e.g. 16/17 for samples sequenced between October 2016 and May 2017), whilst June to September was considered out of typical season (e.g. 2017 for a sequence generated in June 2017). Snip-it plots from the CIVET tool (<https://github.com/artic-network/civet>) were used to illustrate genetic difference between sequences derived from prolonged infection (34).

Figure visualisation and data availability

Figures were generated in R v4.3.2 using the package ggplot2 v3.5.1. At time of preprint, sequences are in preparation for submission to GenBank and scheme files are available on request.

Results

Seasonal coronavirus diagnosis in a regional UK diagnostic lab

Seasonal coronaviruses represent a significant proportion of the viral respiratory pathogens identified in the Nottingham University Hospitals NHS Trust diagnostic laboratory (NUH NHST). Since their inclusion in the respiratory diagnostic RT-qPCR multiplex in February 2016, through to the study end in July 2023, 1540 positives have been recorded in c. 62,000 tests (data not shown), representing 2.5% of all samples tested and 5.75% of those with a diagnosed viral infection.

Pre-Pandemic, broadly similar numbers of coronaviral positives were identified in each yearly infection season, most abundantly in one of the first 3 months of the year (Figure 1), with a peak seen in March of 2016, February of 2017 and January of both 2018 and 2019. However, infections were recorded all year round, but are significantly decreased to a background level outside of the wider coronaviral season of October to May. Post-pandemic, considerably less samples were submitted for seasonal coronavirus testing due to reconfigured diagnostic practice, reducing discrimination of seasonality. Nonetheless, 2023 appeared to indicate seasonal corona infections peaking in January as previously.

Whilst seasonal coronavirus type was only noted for approximately half (798) of recorded positives (1531), with further periodic unevenness (e.g. mid July onward no discrimination made in laboratory records), OC43 (312) and NL63 (308) were indicated in approximately equal numbers. In contrast 229E was observed much more infrequently with 80 positives and HKU1 the rarest of types with just 33 recorded infections. With typing determined by this sequencing study, all four seasonal coronaviruses were again detected post-pandemic, with OC43 the most abundant and in line with more extensive seasonal coronavirus epidemiological studies reported in the UK (7) and USA (10) over similar time periods.

A method to sequence Seasonal Coronaviruses

To interrogate the molecular epidemiology of the seasonal coronaviruses detected at NUH NHST, a novel 1200bp amplicon sequencing scheme to generate near-complete genomes was developed and applied to available typed and untyped archived nucleic acid extracts.

In total, 402 unique samples were sequenced across 700 reactions, summarised in Table 1 and Figure 2:

Table 1: Summary of Coronaviral sequencing investigations and genomic coverage achieved

All sequencing reactions attempted							
		Genome coverage					
Coronavirus species	total	>90	%	>95	%	>99	%
229E	115	99	86.09	88	76.52	72	62.61
NL63	233	178	76.39	167	71.67	85	36.48
OC43	265	207	78.11	166	62.64	106	40.00
HKU1	78	68	87.18	54	69.23	26	33.33
SARS-CoV-2	9	7	77.78	6	66.67	5	55.56
Grand total	700	559	81.11	481	69.35	294	45.60
Unique samples with final scheme iteration							
		Genome coverage					
Coronavirus species	total	>90	%	>95	%	>99	%
229E	71	69	97.18	64	90.14	58	81.69
NL63	100	86	86.00	85	85.00	65	65.00
OC43	185	149	80.54	128	69.19	94	50.81
HKU1	38	29	76.32	27	71.05	25	65.79
SARS-CoV-2	8	7	87.50	6	75.00	5	62.50
Grand total	402	340	85.51	310	78.08	247	65.16

The relative higher prevalence of OC43 and NL63 was reflected in both attempted and ultimately unique sequenced genomes. High quality (>95%) genomes were achieved for 69% (OC43) to 90% (229E) of coronaviral samples, whilst complete amplicon coverage (and near complete genomes) were achieved in 51 to 82% of cases. However, all 702 sequenced samples generated at least 25% of the genome, facilitating conclusive typing, although further samples were initially attempted by PCR, but failed to generate any observable amplicons and were thus not progressed to sequencing (data not shown).

Complete genome sequences could be retrieved over the wide dynamic range of recorded diagnostic lab template copy inputs for all 4 seasonal coronavirus types, from as little as 10 copies to in excess of a million (Figure 2, NB not all copy numbers were recorded in the diagnostic lab records). 229E and HKU1 samples with less than 90% coverage had a copy number value of c. 100 copies (per 10µl) template or less. By contrast, NL63 and OC43 samples often yielded sequence data with low coverage, even at relatively high copy number, suggesting primer mismatch as the likely cause Figure 2.

A limited subset of predominantly post-pandemic samples were amplified and sequenced both with all 5 primer sets combined in each pool (The Full Vivaldi, containing 70 and 69 primer pairs in pool 1 and 2 respectively) and in some cases additionally with the conventional single primer sets (Targeted Vivaldi, containing 13 to 15 primer pairs in each pool), (Supplementary Figure and table 1). The Full Vivaldi approach did yield some

full genomes but was more prone to produce incomplete genomes compared to the targeted approach. It was, however, still capable of typing seasonal coronavirus positives (Supplementary Figure 1). When tested, all but two of the partial genomes could be converted to near-complete (>96% coverage) by repeating amplification with targeted primers.

A sample positive for both OC43 (3431 copies per 10µl template) and SARS-CoV-2 (959 copies) generated complete coverage for the OC43 genome (UKN23_OC43_5) and a partial SARS-CoV-2 genome (65.95%,) when amplified by the Full Vivaldi method. The SARS-CoV-2 genome coverage could be improved to 83.36% (UKN23_CoV2_1) when amplified with the targeted scheme, whilst the OC43 targeted scheme maintained complete coverage.

A small number of patients infected with OC43 were sampled and sequenced within a typical acute phase timeframe of infection and generated paired genomes of >95% coverage, allowing insight into short-term intra-patient virus evolution and accuracy of base-calling across the 30202 sequenced positions. Three paired samples taken one (n=2) and three days apart exhibited no differences, a further two patients saw single differences over two- and seven-day timepoints, whilst the sixth patient presented two mutations in a six-day interval. Taken together, these limited changes are within the expected mutational rate of seasonal coronavirus genomes (35) and indicative of the high fidelity of the amplicon method.

Genetic epidemiology of seasonal coronaviruses in Nottingham, UK, 2016-2023

The developed amplicon scheme thus generated a considerable number of genomes across the pre- and post-pandemic years at our single regional centre, even when compared to the entire available global dataset. High quality sequences (>95%,) from both the study (64 for 229E, 85 for NL63, 128 for OC43 and 27 for HKU1) and GenBank were aligned and used to construct phylogenetic trees (Figures 3 to 6). Lineages were assigned as per previous studies, summarised in Ye et al (16).

229E

Alphacoronaviral 229E study sequences broadly group into 3 well supported clades (Figure 3, genotypes 6 / 7a / 7b). Three sequences from 2016 and one from January 2017 cluster with contemporary samples from the USA (2016 & 2017) and Germany (2015) and Vietnam (2013) assigned to Genotype 6 in previous studies (11, 16). Sparse sequence number representation, relatively long branch lengths and additional well-supported nodes is suggestive of an unsampled diversity within 229E genotype 6. However, the bulk of study sequences situate in a previously noted emerging lineage 7 (16) and are further segregated into two well supported and populated clusters, nominally 7a and 7b, both of which contain study genomes from 2016.

Interestingly, putative genotype 7a was comprised predominantly from our UK study sequences and none from the most extensive prior global study of 229E genomes, set in China immediately preceding the pandemic (16). Genotype 7a was only sampled elsewhere in the USA between 2015 and 2017, and not after the summer of 2018 in our study having been the significant majority variant in the core 2016/17 coronaviral season. In contrast, Genotype 7b was also seen from the outset of our study period, in samples derived from the USA and the key pre-pandemic Chinese Study (16), as well as post-SARS-CoV-2 emergence in Nottingham as well as China, Russia and the USA.

Increasing phylogenetic granularity, we can observe a single UK study sequence UKN18_229E_4 in June from a patient in their 20s, outside of typical UK coronaviral season) sitting within an extensive sub-cluster of Chinese sequences. Similarly, the earliest study sequence UKN16_229E_1 is more closely related to 2017 and 2018 Chinese sequences than anything else from the UK. Conversely, well supported and populated 7b sub-clusters of predominantly UK sequences with accordingly sparse Chinese

representation can be observed. More recent 2022 and 2023 study sequences are found in these distinct genotype 7b clades.

Taken together, the genetic epidemiology of 229E supports the previously suggested notion of genetic drift (36), potentially with local lineage displacement of 7a by 7b suggesting a more nuanced segregation of geographical transmission at regional levels.

NL63

With the significant NL63 cohort presented here and elsewhere recently (16), previously undetected diversity in circulating NL63 could also be observed. Diverse B and C lineages were sequenced throughout the study timeline (Figure 4), with many subtypes detected within the most heavily sampled 2017/18 core season (Figure 7). Notably in lineage B, the increased sampling supports an expansion of nomenclature toward designation of subtypes, similar to lineage C (37). We have tentatively annotated the well supported lineage B clusters subtypes B1 and B2 to assist with discrimination here and potentially elsewhere.

Like its counterpart Alphacoronavirus 229E, NL63 exhibits geographical segregation within its subtypes. For instance UKN17_NL63_22, situating in a well-supported subgroup of B1 with reference sequences from China (2016-2019) and Japan (2018), whilst the bulk of B1, the 'emerging cluster' reported by Ye et al (16) with many 2016-2019 Chinese samples, contains no other UK study sequences despite our extensive sampling in a similar timeframe. Within Lineage B, our UK sampling presents most commonly as B2, with only single representatives from the USA (2015) and Japan (2017), in contrast to multiple 2017, 2018, 2022 and 2023 UK study sequences.

A similar picture presents with lineage C2, indicating a well-supported sub-cluster containing predominantly study sequences from 2016-18 and 2021/22, alongside contemporary US (2019) Japanese (2019) and Russian (2022) references. Elsewhere more limited numbers of further C2 and C3 sequences were observed pre- and post-pandemic, alongside more numerous Chinese and Japanese reference samples. Intriguingly a single study sample UKN18_NL63_24 stands distinctly from all others, (Figure 4), potentially representing a novel intra-species recombinant of lineages B and C as extensively inferred previously (17, 38).

HKU1

HKU1 is the least sequenced Seasonal Coronavirus globally due to an accordingly lower detection rate, therefore even our relatively limited contribution of 27 high quality (>95% coverage) genomes is significant (Figure 5). Genotypes A (n=13) and B (N=14) were detected in approximately equal numbers across the study period, presenting in sub-clusters with other contemporary sequences; we did not detect the recombinant, and to date rarest, Genotype C (14, 39).

Genotype A study samples from 2016-18 and 2022 segregated with high bootstrap support alongside contemporary Japanese (2014 & 2019), US (2015 & 2021) Chinese (2018) and Russian (2022), hinting at extinction of more historically detected genotype A lineages. Genotype B also offers only limited granularity currently, although the two most recent 2023 study samples segregate with earlier UK study genomes, separate from a cluster of US sequences seen at the onset of the SARS-CoV-2 pandemic (28). Other Genotype B study sequences appear more similar to contemporary Chinese (2015-18), Japanese (2016), Australian (2019), Thai (2017) and US (2017, with a temporally outlying 2005 sequence) isolates.

OC43

OC43 is the most extensively sequenced seasonal coronavirus globally, contributing toward a more complicated genotypic nomenclature. We observed study sequences segregating with genotypes E, G, I, J and K both across the entire study and all within the

most sampled 8-month core season of 2017/18 (Figure 6 and 7). The seasonal lineage data is again suggestive in the better-sampled pre-Pandemic era of a predominance in 2016/17 of genotype I being displaced in the following 2017/2018 season by the emergent genotype K (15). Post-Pandemic, a parity of genotype K with the additional emerging genotype J (15) in the 2021/22 season gives way to a predominance of genotype J in the most recent core season of 2022/2023. Interestingly, sample UKN18_OC43_38 stands alone on a long well supported branch, suggestive of further unsampled diversity in circulating viruses.

Intragenotypic temporal geographical segregation can again be observed, with Genotype E infrequently detected elsewhere globally, but featuring as the second most abundant lineage in our most extensively sampled 2017/18 season (Figure 6 and 7), in stark contrast to an absence of genotype E in 74 contemporary reference genomes from China (16). Genotype J indicated some additional well supported clades, potentially indicative of further emerging lineages set to predominate in future seasons.

Prolonged infection by both Seasonal Alpha- and Betacoronaviruses

Several patients presented multiple samples of the same seasonal coronavirus type, although all but two of the serially samplings were within a 3-week timeframe, and most within a few days (data not shown). However, one patient (in their 20s) was determined to be NL63-positive over 195 days and 7 consecutive timepoints in 2017 (UKN17_NL63_1, 6, 7, 15, 16, 18 & 19), whilst another patient (in their 30s) was OC43-positive over 164 days and 4 consecutive timepoints in 2023 (UKN23_OC43_7, 12, 15 & 17). The sequential samples clustered closely on their respective trees (Figures 4 and 6) Both patients were noted to have been under respiratory surveillance post-bone-marrow transplant and thus highly likely to be immunosuppressed. Two weeks after the final positive timeline, the OC43-positive individual was negative for all targets in the viral respiratory multiplex RT-qPCR. The NL63-positive patient was not screened again after the final positive timepoint.

The proximity of such genetically related positive samples suggested a prolonged infection as previously presented with SARS-CoV-2 (40), so the sequence differences were investigated in isolation for intrapatient evolution (Figure 8). The Alphacoronaviral NL63 patient presented an identical genome 22 days after the initial detection, however as the infection continued, 18 mutations arose in the ORF1ab (n = 3, 2 non-synonymous [NS]), Spike (8, all NS), NS3 (2, 1 NS), Matrix (1 NS), and Nucleocapsid (4, 2 NS and a 9bp). Half of the mutations, involving all genes and including the 3 amino acid nucleocapsid deletion appeared in the last time point.

Similar early and late time course mutation patterns were observed for the Betacoronavirus OC43 with no differences seen in the initial acute phase at 9 days. In contrast, mutations at 9 sites were observed at the 107-day timepoint and 14 further at 164 days, in the ORF1ab (n=10, 8 NS), NS2a (1 NS), Spike (10, 8 NS), Envelope (1 NS) and Membrane (2 NS).

Co- and re-infected patients

In addition to the above mentioned OC43 / SARS-CoV-2 co-infected patient, the diagnostic testing indicated a further individual positive for both Alpha- and Betacoronaviruses (NL63 and HKU1), previously undetected by the non-discriminatory diagnostic assay, with a combined seasonal coronavirus genomic template value of 1771 copies / 10 μ l. A complete HKU1 genome and 47% of the NL63 genome was recovered by the Full Vivaldi method, elevated to 82% by subsequent NL63-exclusive sequencing.

Further to the consecutive timepoint NL63- and OC43-infections, 4 instances of multiple heterologous coronaviral infection in the same season were observed, all with NL63 and OC43 (data not shown). These ranged from as little as 38 to as many as 128 days apart, and all were observed within the most heavily sampled 2017/18 season.

Discussion

The scale and rapidity of the sequencing efforts since the start of the SARS-CoV2 pandemic (41, 42), accompanied by the extensive media coverage and its impact on government health policies, has underscored the importance of viral genomics. While there has been significant global attention on this novel coronavirus spillover, there has been a notable absence of genomic data for the four endemic seasonal coronaviruses. This represents a critical gap in our knowledge that could provide insights into the future course of SARS-CoV-2 in the human population (35, 43, 44). Our study delivers a simple method to improve ongoing surveillance of seasonal coronaviruses and contributes substantially by providing 'backfill' of recent pre-pandemic genomes of all four species.

To reproducibly detect whole genomes from seasonal coronavirus strains (45-47) in unenriched clinical samples, demanded not only an amplicon approach, but also a fine balance between minimising priming sites and maximising the sensitivity of detecting smaller RNA fragments. We thus opted for a 1200bp amplicon scheme as used previously for SARS-CoV-2 (20), rather than the commonly adopted 400bp amplicon scheme (19, 21), larger amplicon schemes (18, 23, 24) or an unbiased metagenomic approach (28, 48). Initial attempts on even a small batch of clinical samples can quickly identify amplicon drop-out, whilst careful continued evaluation is required to guard against evolving drop-out such as seen previously in SARS-CoV-2 (25), and similarly seen in our cohort in post-pandemic 229E infections.

This approach has high sensitivity, with whole genome retrieval from samples with less than 100 genome copies of input, in line with previously reported sensitivity of the 400bp amplicon of 50 copies per reaction (21) and likely orders of magnitude more sensitive than metagenomic approaches to coronaviral genome sequencing (28). The comparable sensitivity of our larger 1200bp to previous 400bp schemes can be explained in part by increasing amplification cycle number from the typically used 35, to 45 (data not shown, (20, 21)). Concerns about the introduction of PCR error and / or detection of contaminating sequences by elevated cycle number can be minimised by not only consistent use of negative controls (49), but also use of proof-reading enzymes (19). The minimal differences observed in our closely but independently sampled OC43 infections – 4 differences across 6 paired c. 30Kb genomes – indicates a robust degree of fidelity in the method but this could be further tested with clonal templates to probe error rates. Similarly, to metagenomic sequencing, strong signals of only fractional parts of a coronaviral genome should be treated with suspicion (50).

To maximise the scope of our method, we ensured our scheme could target an unprecedented 5 near-complete species' genomes from two coronaviral genera in the minimal 2 reaction format (21). To our knowledge, this is the broadest complete genome amplicon method presented to date and showcases the considerable potential of the method in sensitively targeting multiple viruses in whole, or part (51). We validated this by confirming clinically diagnosed coronavirus co-infections and identifying others missed by RT-qPCR multiplex assay panels, which can have poor discriminatory potential.

Combining our novel methodology and extensive diagnostic surplus archive, we have generated the largest single cohort of seasonal coronaviral genomes to date. Thus, when combined with the global dataset, and notably other contemporary in-depth studies from China (16) and the US (28), we were able to gain greater insight into the ebb and flow of seasonal coronaviral genotypes / lineages, such as has been monitored with unprecedented detail with the analogous SARS-CoV-2 variants. In contrast to the currently observed pattern of emergence and apparent total selective sweeps of SARS-CoV-2 lineages (35, 52)(REF), the seasonal coronaviruses were observed to present a more complex co-circulation of genetically distinct lineages rising and falling in predominance, sometimes with contemporary geographical variation and prevalence. For example, 229E saw a progressive sweep in our UK cohort through lineages 6 and 7a pre-pandemic, to

exclusively 7b post-pandemic; however 7b was predominant in a pre-pandemic Chinese cohort (16). Similarly, OC43 lineage I was displaced in predominance by the closely genetically related lineage K (15) between the 2016/17 and 2017/18 seasons. Whilst lineage J, a variant of precursor lineage H (15) was also present in a minority in both these seasons, it presented as the overwhelming OC43 lineage of the 2022/23 UK season. It was generally notable that the season we sampled most extensively (2017/2018) also exhibited the greatest diversity of both alphacoronaviral NL63 and betacoronaviral OC43 lineages, further emphasising the need for greater genomic surveillance to capture the true extent of circulating seasonal coronavirus variability. Underlining this point was the high prevalence of OC43 genotype E in 2017/18, infrequently observed elsewhere globally in recent sampling, but strikingly associated with a fatal encephalitis case previously also in the UK in 2011 (53).

It has been noted for NL63 that genotype switching was not a prerequisite to reinfection of an individual (48), supported by phylogenetic analysis and in contrast to 229E and OC43 (44). The selective force on 229E to antigenically evolve has been further demonstrated *in vivo* with historical sera failing to neutralise more contemporary isolates through significant variation in receptor-binding domains (43) with our data indicating the trend continues. Although an older study suggested re-infection by the same 229E strain was indeed possible in the short term (54), this was in a highly controlled experimental setting and in direct contradiction to a similar earlier study in the 1980s (55). As noted elsewhere (44), HKU1 data is insufficient to draw conclusions about its evolutionary trajectory, however the novel method and relatively large HKU1 cohorts presented here and recently elsewhere (28) could redress this imbalance in future studies. However more broadly, absence of a recombinant HKU1 A / B lineage C extant around the time HKU1 was discovered (6, 14) was in agreement with other key contemporary studies (28, 39).

Overall, this may suggest that dominating SARS-CoV-2 lineages in well-sampled region may not subsequently be swept to extinction but could continue undetected circulation and further evolution in certain populations before re-emergence in others. The unexpected appearance and subsequent domination of novel SARS-CoV-2 variants Delta and Omicron were indeed suggestive of transmission and evolution in unsampled populations (42, 56, 57) or longer term intra-host evolution in immunosuppressed individuals as discussed below. The significant differences between the genomes in this single centre study and the other largest contemporary seasonal coronavirus cohort to date, from 36 hospitals in Beijing with a catchment of 25 million people (16), underscores the need for greater and broader surveillances of respiratory viral infections, not only for seasonal coronaviruses, but other pathogens also (58).

The effect of non-pharmaceutical interventions to reduce pandemic transmission of SARS-CoV-2 has had a significant effect on the well-monitored circulation of Influenza, notably with the apparent disappearance of Influenza B Yamagata (59). These measures may also have exerted pressure on some seasonal coronavirus lineages, with a significant delay in the typical US seasonal onset observed in 2020/21 attributed to the SARS-CoV-2 pandemic (10). Pre-pandemic patterns of alternating seasonal betacoronaviral dominance of OC43 and HKU1, punctuated by parity (10), may be further perturbed by significantly enhanced population immunity (by infection and / or vaccination) of SARS-CoV-2 as a third beta-coronavirus (60). But ultimately immunity against re-infection, cross-protective or otherwise, is short-lived (61), whilst protection against severe disease remains (60, 62).

Paradoxically 'out of season' seasonal coronaviral infections were observed. In some instances, phylogenetic analysis was suggestive of importation to the UK by international travel, as we and others reported previously for SARS-CoV-2 (41, 42). Conversely, there may be continued low-level circulation within the UK and elsewhere. In other instances, out of season positives were the result of likely persistent infections acquired by the immunocompromised within the typical epidemic months. Long-term

SARS-CoV2 infection of individuals (especially immune-suppressed), leading to intra-patient virus evolution has been described (40, 63), and offered as a potential mechanism for the emergence of some highly evolved variants of concern (63, 64). We similarly demonstrate that increased genomic sequencing of seasonal coronaviruses in even our single centre cohort can reveal prolonged infection and evolution in both beta- and alphacoronaviruses. Perhaps most strikingly we observed a 9bp nucleotide deletion immediately prior to the start of the C-terminal domain of the nucleocapsid in the disordered linker region responsible for RNA binding and oligomerisation (65, 66) in the last timepoint of our chronically infected NL63 patient. This had parallels in SARS-CoV-2, with a 3 AA nucleocapsid deletion relative to the ancestral Wuhan strain a feature of the Omicron lineage, albeit in the N terminal domain (67). The recent spillover of a novel canine coronavirus in Malaysia also reported a larger 12 AA deletion in the middle of the N protein, hypothesised to be associated with adaptation to a new host (68), similar to previously reported nucleocapsid deletions (and insertions) found to determine nuclear localization in MERS and SARS-CoV-1 infections (69). However, caution must be taken in extrapolating any isolated changes in immunocompromised patients to effects on transmissibility in the wider population.

In conclusion, given the recent spillover of SARS-CoV-2, its uncertain evolutionary path, the concurrent circulation and potential co-infection with endemic coronaviruses, and the propensity of coronaviruses for genetic recombination, there is a pressing need for enhanced surveillance of seasonal coronaviruses. The methodology outlined in this study closely resembles widely accepted amplicon schemes used for SARS-CoV-2, but with a notable advancement: it offers unprecedented coverage across two genera and five species. Furthermore, it has demonstrated high sensitivity in both pre- and post-pandemic cohorts of all four seasonal coronaviruses. Therefore, it is well-suited for widespread adoption, serving as a valuable tool to complement the increasing genomic sequencing efforts of other respiratory viruses. This will provide crucial insights for informed epidemiological analyses and public health decision-making.

Conflict of Interest

All authors declare no conflict of interest.

Funding Source

The study was funded by COG-UK (COVID-19 Genomics UK Consortium).

Ethical Approval statement

The COG-UK consortium work was approved by the PHE Research Ethics and Governance Group (R&D NR0195). Use of pre-pandemic residual diagnostic nucleic acids and associated anonymized patient information was covered by ethical approval granted to the Nottingham Health Science Biobank Research Tissue Bank, by the North West - Greater Manchester Central Research Ethics Committee, UK, reference 15/NW/0685. Nottingham University Hospitals National Health Service Trust (Nottingham, UK) further approved investigation of diagnosed coronavirus positive samples under clinical audit number 23-078C.

References

1. Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, et al. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol.* 2016 Jun;24(6):490-502.
2. Woo PCY, de Groot RJ, Haagmans B, Lau SKP, Neuman BW, Perlman S, et al. ICTV Virus Taxonomy Profile: Coronaviridae 2023. *J Gen Virol.* 2023 Apr;104(4).

3. Hamre D, Procknow JJ. A new virus isolated from the human respiratory tract. *Proc Soc Exp Biol Med*. 1966 Jan;121(1):190-3.
4. McIntosh K, Dees JH, Becker WB, Kapikian AZ, Chanock RM. Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. *Proc Natl Acad Sci U S A*. 1967 Apr;57(4):933-40.
5. van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, Wolthers KC, et al. Identification of a new human coronavirus. *Nat Med*. 2004 Apr;10(4):368-73.
6. Woo PC, Lau SK, Chu CM, Chan KH, Tsoi HW, Huang Y, et al. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol*. 2005 Jan;79(2):884-95.
7. Nickbakhsh S, Ho A, Marques DFP, McMenamin J, Gunson RN, Murcia PR. Epidemiology of Seasonal Coronaviruses: Establishing the Context for the Emergence of Coronavirus Disease 2019. *J Infect Dis*. 2020 Jun 16;222(1):17-25.
8. Veiga A, Martins LG, Riediger I, Mazetto A, Debur MDC, Gregianini TS. More than just a common cold: Endemic coronaviruses OC43, HKU1, NL63, and 229E associated with severe acute respiratory infection and fatality cases among healthy adults. *J Med Virol*. 2021 Feb;93(2):1002-7.
9. Torjesen I. Covid-19 will become endemic but with decreased potency over time, scientists believe. *BMJ*. 2021 Feb 18;372:n494.
10. Shah MM, Winn A, Dahl RM, Kniss KL, Silk BJ, Killerby ME. Seasonality of Common Human Coronaviruses, United States, 2014-2021. *Emerg Infect Dis*. 2022 Oct;28(10):1970-6.
11. Al-Khannaq MN, Ng KT, Oong XY, Pang YK, Takebe Y, Chook JB, et al. Diversity and Evolutionary Histories of Human Coronaviruses NL63 and 229E Associated with Acute Upper Respiratory Tract Symptoms in Kuala Lumpur, Malaysia. *Am J Trop Med Hyg*. 2016 May 4;94(5):1058-64.
12. Shao N, Zhang C, Dong J, Sun L, Chen X, Xie Z, et al. Molecular evolution of human coronavirus-NL63, -229E, -HKU1 and -OC43 in hospitalized children in China. *Front Microbiol*. 2022;13:1023847.
13. Zhu Y, Li C, Chen L, Xu B, Zhou Y, Cao L, et al. A novel human coronavirus OC43 genotype detected in mainland China. *Emerg Microbes Infect*. 2018 Oct 30;7(1):173.
14. Woo PC, Lau SK, Yip CC, Huang Y, Tsoi HW, Chan KH, et al. Comparative analysis of 22 coronavirus HKU1 genomes reveals a novel genotype and evidence of natural recombination in coronavirus HKU1. *J Virol*. 2006 Jul;80(14):7136-45.
15. Zhang Z, Liu W, Zhang S, Wei P, Zhang L, Chen D, et al. Two novel human coronavirus OC43 genotypes circulating in hospitalized children with pneumonia in China. *Emerg Microbes Infect*. 2022 Dec;11(1):168-71.
16. Ye R-Z, Gong C, Cui X-M, Liu J-Y, Fan H, Xie H, et al. Continuous evolution and emerging lineage of seasonal human coronaviruses: A multicenter surveillance study. *Journal of Medical Virology*. 2023;95(6):e28861.
17. Muller NF, Kistler KE, Bedford T. A Bayesian approach to infer recombination patterns in coronaviruses. *Nat Commun*. 2022 Jul 20;13(1):4186.
18. Dong X, Deng YM, Aziz A, Whitney P, Clark J, Harris P, et al. A simplified, amplicon-based method for whole genome sequencing of human respiratory syncytial viruses. *J Clin Virol*. 2023 Apr;161:105423.
19. Francis RV, Billam H, Clarke M, Yates C, Tsoleridis T, Berry L, et al. The Impact of Real-Time Whole-Genome Sequencing in Controlling Healthcare-Associated SARS-CoV-2 Outbreaks. *J Infect Dis*. 2022 Jan 5;225(1):10-8.
20. Freed NE, Vikova M, Faisal MB, Silander OK. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. *Biol Methods Protoc*. 2020;5(1):bpaa014.
21. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc*. 2017 Jun;12(6):1261-76.
22. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature*. 2016 Feb 11;530(7589):228-32.

23. Maes M, Khokhar F, Wilkinson SAJ, Smith AD, Kovalenko G, Dougan G, et al. Multiplex MinION sequencing suggests enteric adenovirus F41 genetic diversity comparable to pre-COVID-19 era. *Microb Genom*. 2023 Jan;9(1).
24. Joffret ML, Polston PM, Razafindratsimandresy R, Bessaud M, Heraud JM, Delpeyroux F. Whole Genome Sequencing of Enteroviruses Species A to D by High-Throughput Sequencing: Application for Viral Mixtures. *Front Microbiol*. 2018;9:2339.
25. Borcard L, Gempeler S, Terrazos Miani MA, Baumann C, Grädel C, Dijkman R, et al. Investigating the Extent of Primer Dropout in SARS-CoV-2 Genome Sequences During the Early Circulation of Delta Variants. *Frontiers in Virology*. 2022 2022-April-13;2.
26. Kovacs D, Mambule I, Read JM, Kiran A, Chilombe M, Bvumbwe T, et al. Epidemiology of Human Seasonal Coronaviruses Among People With Mild and Severe Acute Respiratory Illness in Blantyre, Malawi, 2011-2017. *J Infect Dis*. 2024 Feb 14.
27. Zhao MC, Wen C, Sun L, Duan SX, Zang KX, Wang L, et al. Epidemiology and Clinical Characteristics of Seasonal Human Coronaviruses in Children Hospitalized in Hebei Province, China Before and During the COVID-19 Pandemic. *Risk Manag Healthc Policy*. 2023;16:1801-7.
28. Chow EJ, Casto AM, Rogers JH, Roychoudhury P, Han PD, Xie H, et al. The clinical and genomic epidemiology of seasonal human coronaviruses in congregate homeless shelter settings: A repeated cross-sectional study. *Lancet Reg Health Am*. 2022 Nov;15:100348.
29. Chellapuri A, Smitheman M, Chappell JG, Clark G, Howson-Wells HC, Berry L, et al. Human parainfluenza 2 & 4: Clinical and genetic epidemiology in the UK, 2013-2017, reveals distinct disease features and co-circulating genomic subtypes. *Influenza Other Respir Viruses*. 2022 Nov;16(6):1122-32.
30. Howson-Wells HC, Tsoleridis T, Zainuddin I, Tarr AW, Irving WL, Ball JK, et al. Enterovirus D68 epidemic, UK, 2018, was caused by subclades B3 and D1, predominantly in children and adults, respectively, with both subclades exhibiting extensive genetic diversity. *Microb Genom*. 2022 May;8(5).
31. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018 Sep 15;34(18):3094-100.
32. Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods*. 2017 Apr;14(4):407-10.
33. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004 Aug 19;5:113.
34. O'Toole Á, Hill V, Jackson B, Dewar R, Sahadeo N, Colquhoun R, et al. Genomics-informed outbreak investigations of SARS-CoV-2 using civet. *PLOS Global Public Health*. 2022;2(12):e0000704.
35. Markov PV, Ghafari M, Beer M, Lythgoe K, Simmonds P, Stilianakis NI, et al. The evolution of SARS-CoV-2. *Nat Rev Microbiol*. 2023 Jun;21(6):361-79.
36. Chibo D, Birch C. Analysis of human coronavirus 229E spike and nucleoprotein genes demonstrates genetic drift between chronologically distinct strains. *J Gen Virol*. 2006 May;87(Pt 5):1203-8.
37. Wang Y, Li X, Liu W, Gan M, Zhang L, Wang J, et al. Discovery of a subgenotype of human coronavirus NL63 associated with severe lower respiratory tract infection in China, 2018. *Emerg Microbes Infect*. 2020;9(1):246-55.
38. Pollett S, Conte MA, Sanborn M, Jarman RG, Lidl GM, Modjarrad K, et al. A comparative recombination analysis of human coronaviruses and implications for the SARS-CoV-2 pandemic. *Sci Rep*. 2021 Aug 30;11(1):17365.
39. Chen X, Zhu Y, Li Q, Lu G, Li C, Jin R, et al. Genetic characteristics of human coronavirus HKU1 in mainland China during 2018. *Arch Virol*. 2022 Nov;167(11):2173-80.
40. Kemp SA, Collier DA, Datir RP, Ferreira I, Gayed S, Jahun A, et al. SARS-CoV-2 evolution during treatment of chronic infection. *Nature*. 2021 Apr;592(7853):277-82.
41. Chappell JG, Tsoleridis T, Clark G, Berry L, Holmes N, Moore C, et al. Retrospective screening of routine respiratory samples revealed undetected community transmission and missed intervention opportunities for SARS-CoV-2 in the United Kingdom. *J Gen Virol*. 2021 Jun;102(6).

42. du Plessis L, McCrone JT, Zarebski AE, Hill V, Ruis C, Gutierrez B, et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science*. 2021 Feb 12;371(6530):708-12.
43. Eguia RT, Crawford KHD, Stevens-Ayers T, Kelnhofer-Millevolte L, Greninger AL, Englund JA, et al. A human coronavirus evolves antigenically to escape antibody immunity. *PLoS Pathog*. 2021 Apr;17(4):e1009453.
44. Kistler KE, Bedford T. Evidence for adaptive evolution in the receptor-binding domain of seasonal coronaviruses OC43 and 229e. *Elife*. 2021 Jan 19;10.
45. Corman VM, Muth D, Niemeyer D, Drosten C. Hosts and Sources of Endemic Human Coronaviruses. *Adv Virus Res*. 2018;100:163-88.
46. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol*. 2019 Mar;17(3):181-92.
47. Forni D, Cagliani R, Arrigoni F, Benvenuti M, Mozzi A, Pozzoli U, et al. Adaptation of the endemic coronaviruses HCoV-OC43 and HCoV-229E to the human host. *Virus Evol*. 2021;7(2):veab061.
48. Kiyuka PK, Agoti CN, Munywoki PK, Njeru R, Bett A, Otieno JR, et al. Human Coronavirus NL63 Molecular Epidemiology and Evolutionary Patterns in Rural Coastal Kenya. *J Infect Dis*. 2018 May 5;217(11):1728-39.
49. Lopez-Labrador FX, Brown JR, Fischer N, Harvala H, Van Boheemen S, Cinek O, et al. Recommendations for the introduction of metagenomic high-throughput sequencing in clinical virology, part I: Wet lab procedure. *J Clin Virol*. 2021 Jan;134:104691.
50. de Vries JJC, Brown JR, Couto N, Beer M, Le Mercier P, Sidorov I, et al. Recommendations for the introduction of metagenomic next-generation sequencing in clinical virology, part II: bioinformatic analysis and reporting. *J Clin Virol*. 2021 May;138:104812.
51. Brinkmann A, Uddin S, Ulm SL, Pape K, Forster S, Enan K, et al. RespiCoV: Simultaneous identification of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and 46 respiratory tract viruses and bacteria by amplicon-based Oxford-Nanopore MinION sequencing. *PLoS One*. 2022;17(3):e0264855.
52. Viana R, Moyo S, Amoako DG, Tegally H, Scheepers C, Althaus CL, et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature*. 2022 Mar;603(7902):679-86.
53. Morfopoulou S, Brown JR, Davies EG, Anderson G, Virasami A, Qasim W, et al. Human Coronavirus OC43 Associated with Fatal Encephalitis. *N Engl J Med*. 2016 Aug 4;375(5):497-8.
54. Callow KA, Parry HF, Sergeant M, Tyrrell DA. The time course of the immune response to experimental coronavirus infection of man. *Epidemiol Infect*. 1990 Oct;105(2):435-46.
55. Reed SE. The behaviour of recent isolates of human respiratory coronavirus in vitro and in volunteers: evidence of heterogeneity among 229E-related strains. *J Med Virol*. 1984;13(2):179-92.
56. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*. 2021 May;593(7858):266-9.
57. Tegally H, Moir M, Everatt J, Giovanetti M, Scheepers C, Wilkinson E, et al. Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa. *Nat Med*. 2022 Sep;28(9):1785-90.
58. Tang JW, Lam TT, Zaraket H, Lipkin WI, Drews SJ, Hatchette TF, et al. Global epidemiology of non-influenza RNA respiratory viruses: data gaps and a growing need for surveillance. *Lancet Infect Dis*. 2017 Oct;17(10):e320-e6.
59. Paget J, Caini S, Del Riccio M, van Waarden W, Meijer A. Has influenza B/Yamagata become extinct and what implications might this have for quadrivalent influenza vaccines? *Euro Surveill*. 2022 Sep;27(39).
60. Kundu R, Narean JS, Wang L, Fenn J, Pillay T, Fernandez ND, et al. Cross-reactive memory T cells associate with protection against SARS-CoV-2 infection in COVID-19 contacts. *Nat Commun*. 2022 Jan 10;13(1):80.

61. Edridge AWD, Kaczorowska J, Hoste ACR, Bakker M, Klein M, Loens K, et al. Seasonal coronavirus protective immunity is short-lasting. *Nat Med*. 2020 Nov;26(11):1691-3.
62. Lineburg KE, Grant EJ, Swaminathan S, Chatzileontiadou DSM, Szeto C, Sloane H, et al. CD8(+) T cells specific for an immunodominant SARS-CoV-2 nucleocapsid epitope cross-react with selective seasonal coronaviruses. *Immunity*. 2021 May 11;54(5):1055-65 e5.
63. Chaguza C, Hahn AM, Petrone ME, Zhou S, Ferguson D, Breban MI, et al. Accelerated SARS-CoV-2 intrahost evolution leading to distinct genotypes during chronic infection. *Cell Rep Med*. 2023 Feb 21;4(2):100943.
64. Gonzalez-Reiche AS, Alshammary H, Schaefer S, Patel G, Polanco J, Carreno JM, et al. Sequential intrahost evolution and onward transmission of SARS-CoV-2 variants. *Nat Commun*. 2023 Jun 3;14(1):3235.
65. Zhang B, Tian J, Zhang Q, Xie Y, Wang K, Qiu S, et al. Comparing the Nucleocapsid Proteins of Human Coronaviruses: Structure, Immunoregulation, Vaccine, and Targeted Drug. *Front Mol Biosci*. 2022;9:761173.
66. Szelazek B, Kabala W, Kus K, Zdzalik M, Twarda-Clapa A, Golik P, et al. Structural Characterization of Human Coronavirus NL63 N Protein. *J Virol*. 2017 Jun 1;91(11).
67. Suharsono H, Mahardika BK, Sudipa PH, Sari TK, Suardana IBK, Mahardika GN. Consensus insertion/deletions and amino acid variations of all coding and noncoding regions of the SARS-CoV-2 Omicron clades, including the XBB and BQ.1 lineages. *Arch Virol*. 2023 May 8;168(6):156.
68. Vlasova AN, Diaz A, Damtie D, Xiu L, Toh TH, Lee JS, et al. Novel Canine Coronavirus Isolated from a Hospitalized Patient With Pneumonia in East Malaysia. *Clin Infect Dis*. 2022 Feb 11;74(3):446-54.
69. Gussow AB, Auslander N, Faure G, Wolf YI, Zhang F, Koonin EV. Genomic determinants of pathogenicity in SARS-CoV-2 and other human coronaviruses. *Proc Natl Acad Sci U S A*. 2020 Jun 30;117(26):15193-9.



Figure 1: Diagnosed Seasonal Coronavirus infections, Nottingham University Hospitals NHS Trust UK, February 2016 to July 2023. Seasonal coronavirus testing was undertaken with the AusDiagnostics Respiratory 16-plex assay beginning in February 2016 and able to discriminate seasonal coronavirus type, but this was not always recorded, in which case ‘sCoV’ was indicated. In February 2020 Seasonal Coronavirus detection was switched off to avoid clinical confusion at the onset of the SARS-CoV-2 pandemic. Seasonal Coronavirus testing was resumed in May 2021 in a more selective capacity, where after no diagnostic distinction was made between 229E, NL63, OC43 & HKU1 on a newly reconfigured AusDiagnostics panel. All available samples collected post-SARS-CoV-2 emergence were attempted for WGS and where successful the type is presented.

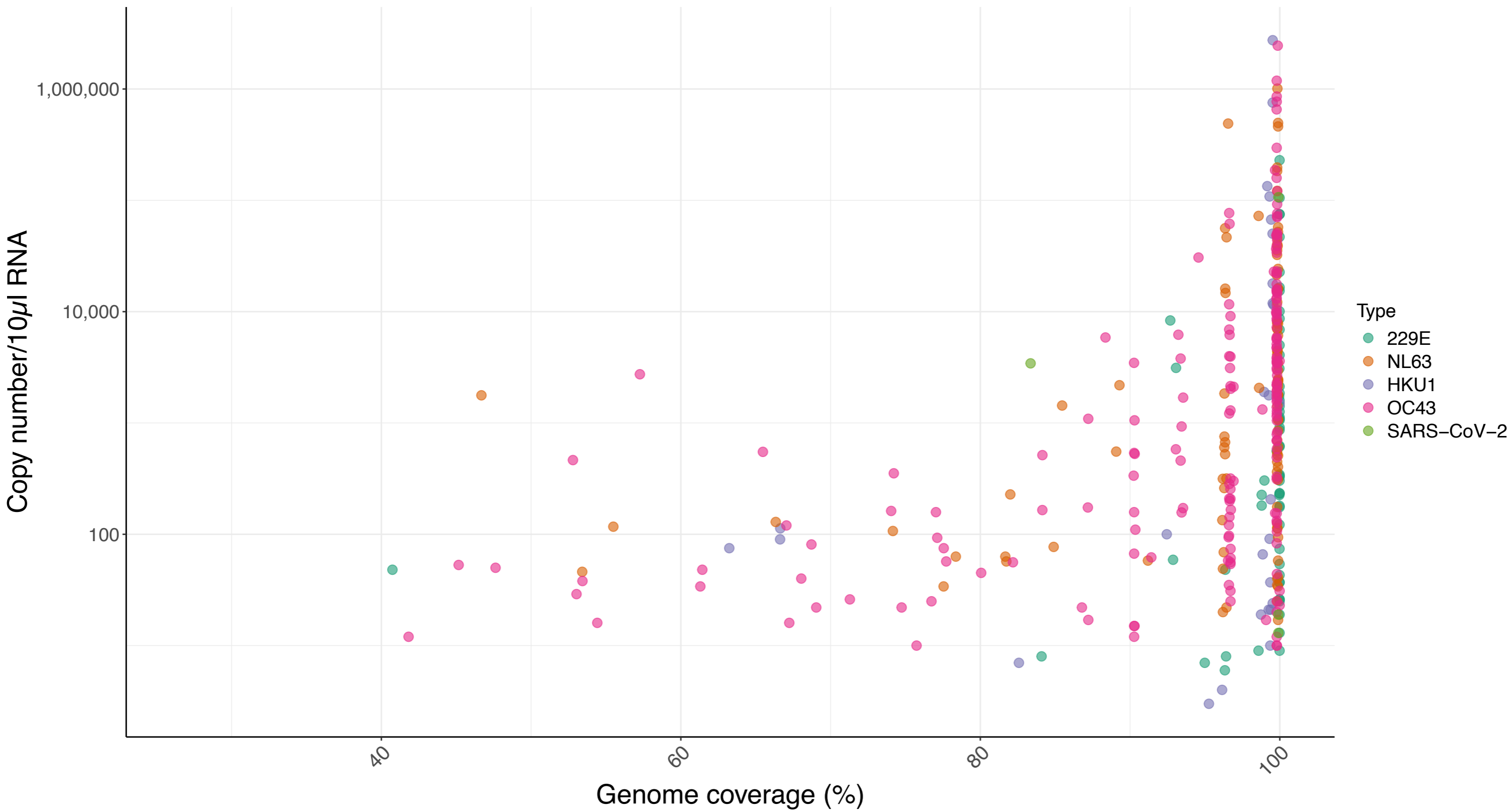


Figure 2: Genome coverage of Seasonal Coronaviruses by the ‘Vivaldi’ 1200bp amplicon scheme.

The genome coverage of each Seasonal Coronavirus 229E, NL63, HKU1 & OC43 (x axis) and additionally SARS-CoV-2, expressed as a percentage of the total region targeted by the amplicon scheme is shown compared to the template copy number calculated by the diagnostic assay (Copies per 10 µl RNA, AusDiagnostics, y-axis).

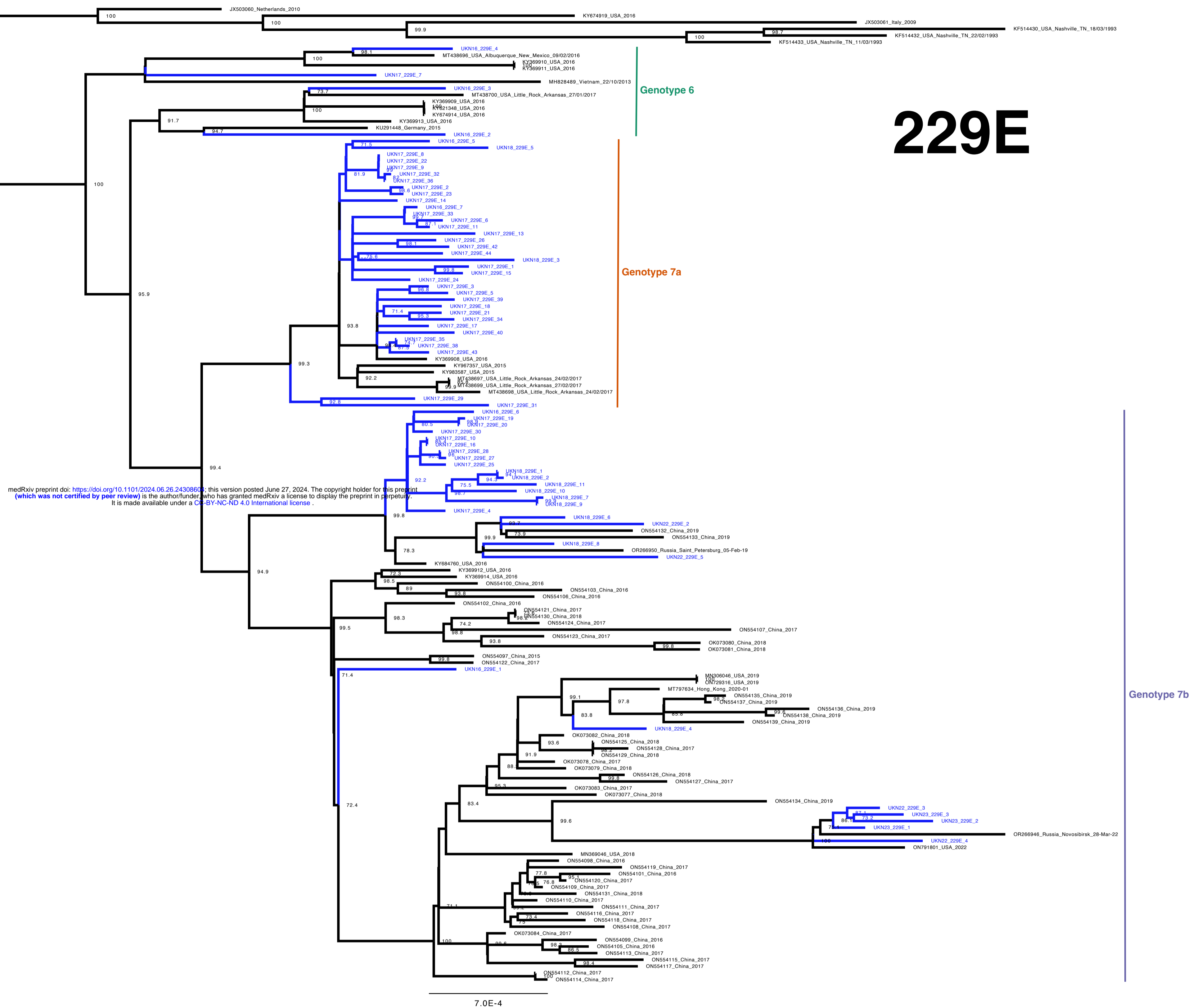


Figure 3: Phylogeny of 229E Seasonal Coronavirus. Phylogenetic relationship by maximum-likelihood method of all Nottingham, UK, study sequences with >95% coverage (blue text) and publicly available genomes (black text) retrieved from GenBank in December 2023 (identified by accession number/country [with region where noted]/year). Numbers to the right of tree nodes indicate SH-aLRT (Shimodaira–Hasegawa approximate likelihood ratio test) branch support, with values <70 not shown. Branch lengths are drawn to a scale of nucleotide substitutions per site, with scale indicated. Tentative genotypes based on well supported clustering (>90 branch support) with lineage exemplar sequences as presented by Ye et al (16) are indicated by vertical bars.

HKU1

DQ339101_China_Hong_Kong_2000s
DQ415897_China_Hong_Kong_2000s
DQ415912_China_Hong_Kong_2000s
DQ415899_China_Hong_Kong_2000s
DQ415913_China_Hong_Kong_2000s
DQ415898_China_Hong_Kong_2000s

Genotype C

medRxiv preprint doi: <https://doi.org/10.1101/2024.06.26.24308604>; this version posted June 27, 2024. 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](https://creativecommons.org/licenses/by-nc-nd/4.0/).

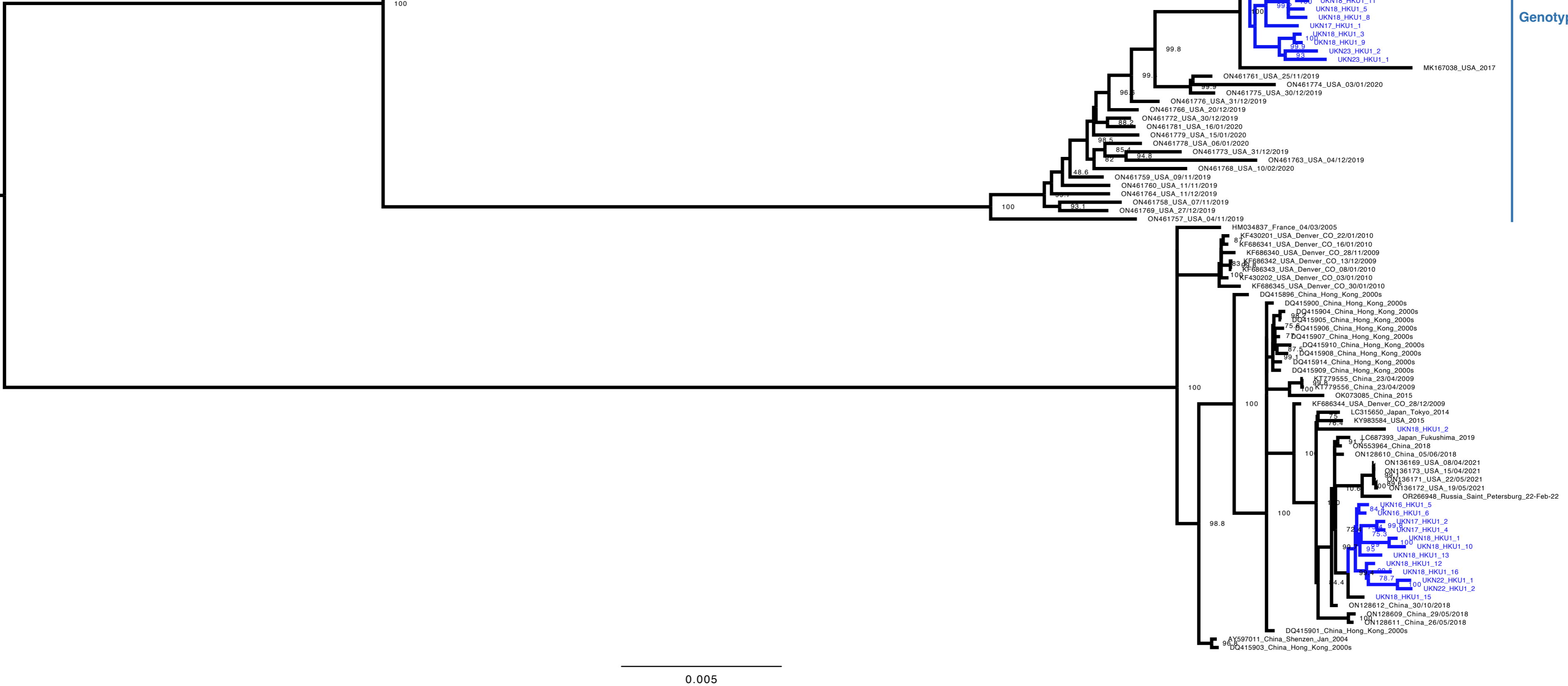
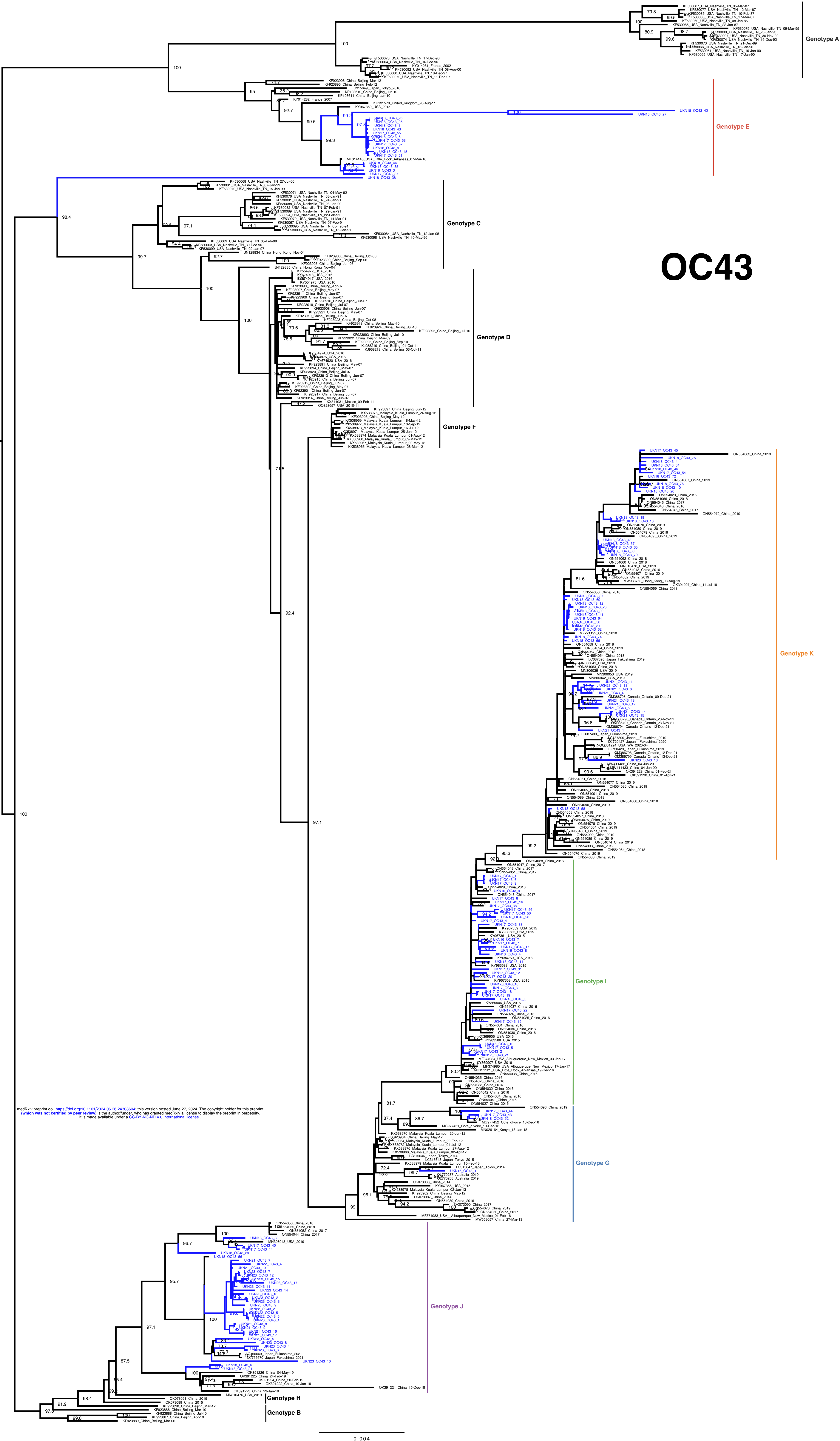


Figure 5: Phylogeny of HKU1 Seasonal Coronavirus. Phylogenetic relationship by maximum-likelihood method of all Nottingham, UK, study sequences with >95% coverage (blue text) and publicly available genomes (black text) retrieved from GenBank in December 2023 (identified by accession number/country [with region where noted]/year). Numbers to the right of tree nodes indicate SH-aLRT (Shimodaira–Hasegawa approximate likelihood ratio test) branch support, with values <70 not shown. Branch lengths are drawn to a scale of nucleotide substitutions per site, with scale indicated. Tentative genotypes based on well supported clustering (>90 branch support) with lineage exemplar sequences as presented by Ye et al (16) are indicated by vertical bars.



medRxiv preprint doi: <https://doi.org/10.1101/2024.06.26.24308604>; this version posted June 27, 2024. 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.

Figure 6: Phylogeny of OC43 Seasonal Coronavirus. Phylogenetic relationship by maximum-likelihood method of all Nottingham, UK, study sequences with >95% coverage (blue text) and publicly available genomes (black text) retrieved from GenBank in December 2023 (identified by accession number/country [with region where not-]/year). Numbers to the right of tree nodes indicate SH-aLRT (Shimodaira-Hasegawa approximate likelihood ratio test) branch support, with values <70 not shown. Branch lengths are drawn to a scale of nucleotide substitutions per site, with scale indicated. Tentative genotypes based on well supported clustering (>90 branch support) with lineage exemplar sequences as presented by Ye et al (16) are indicated by vertical bars.

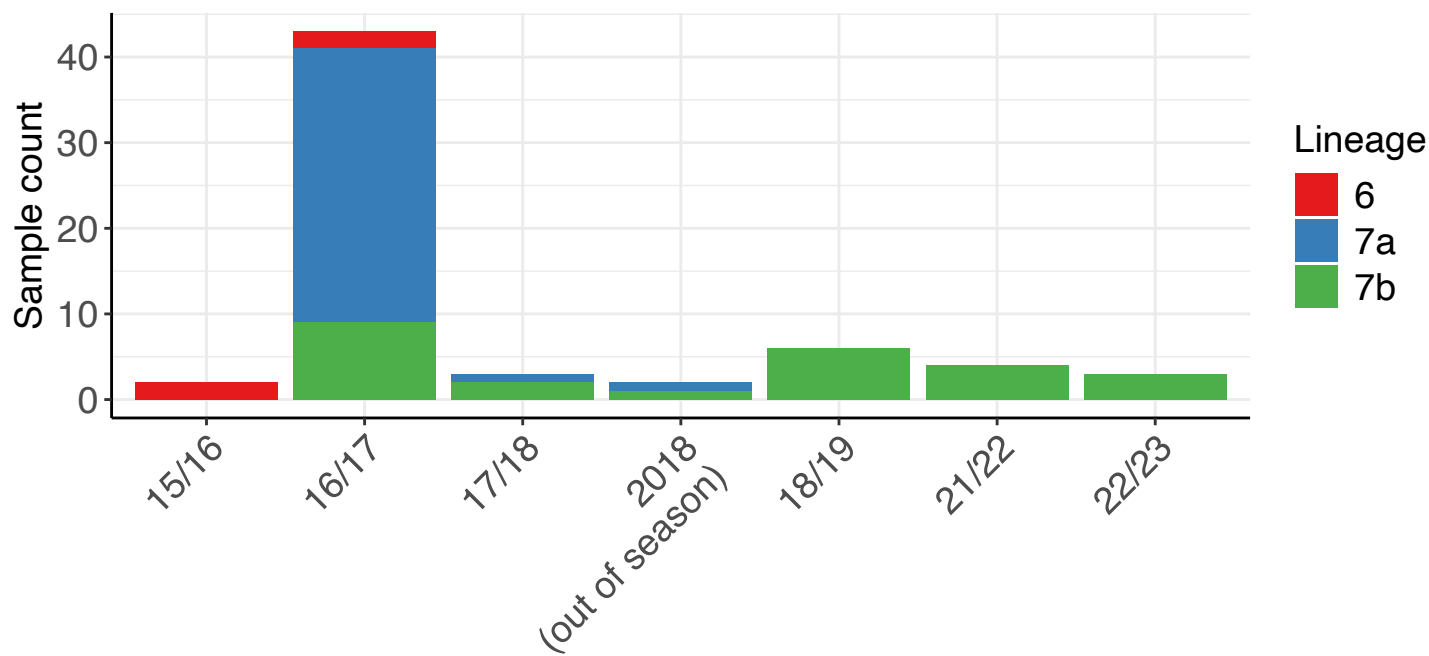
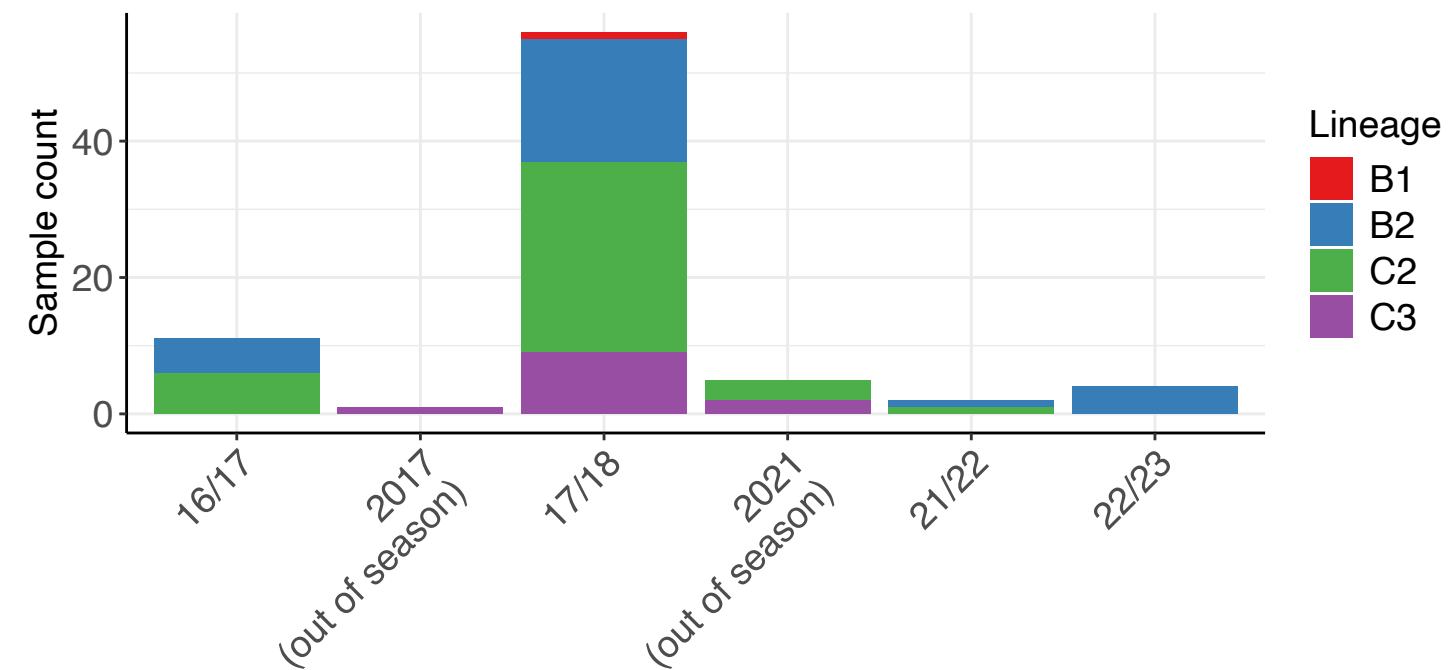
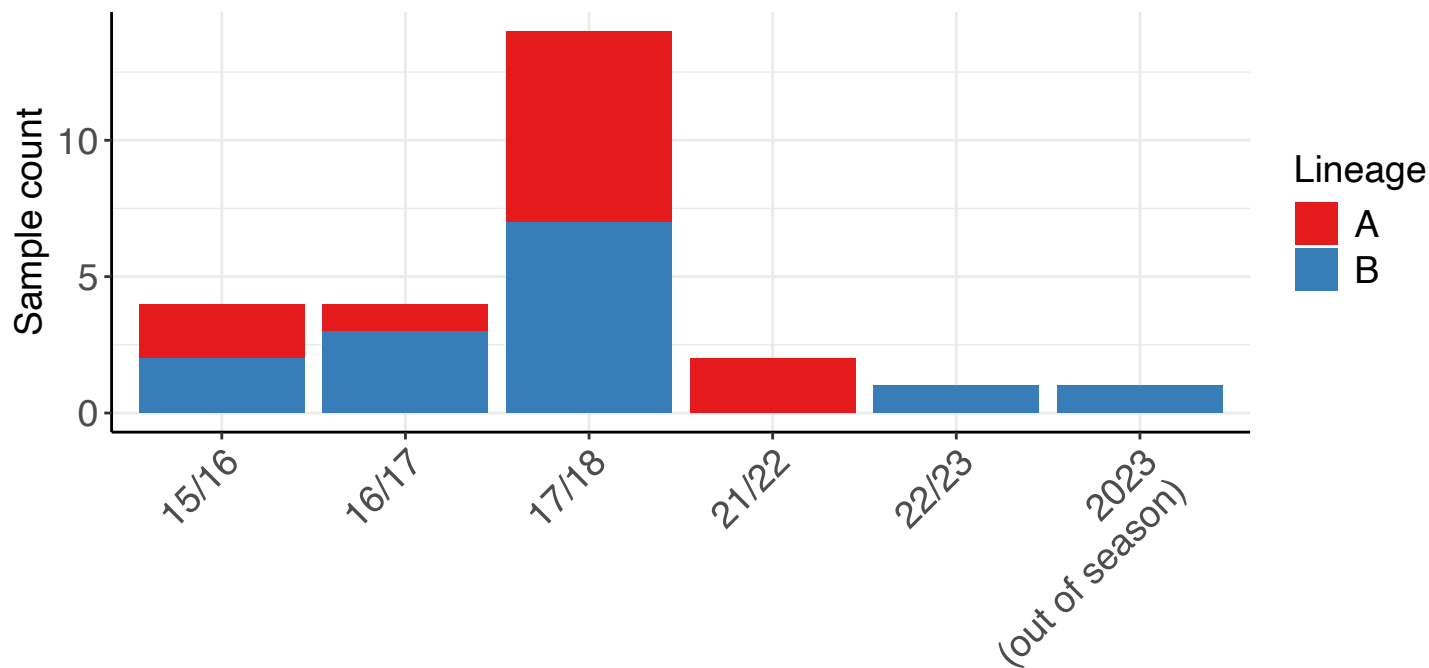
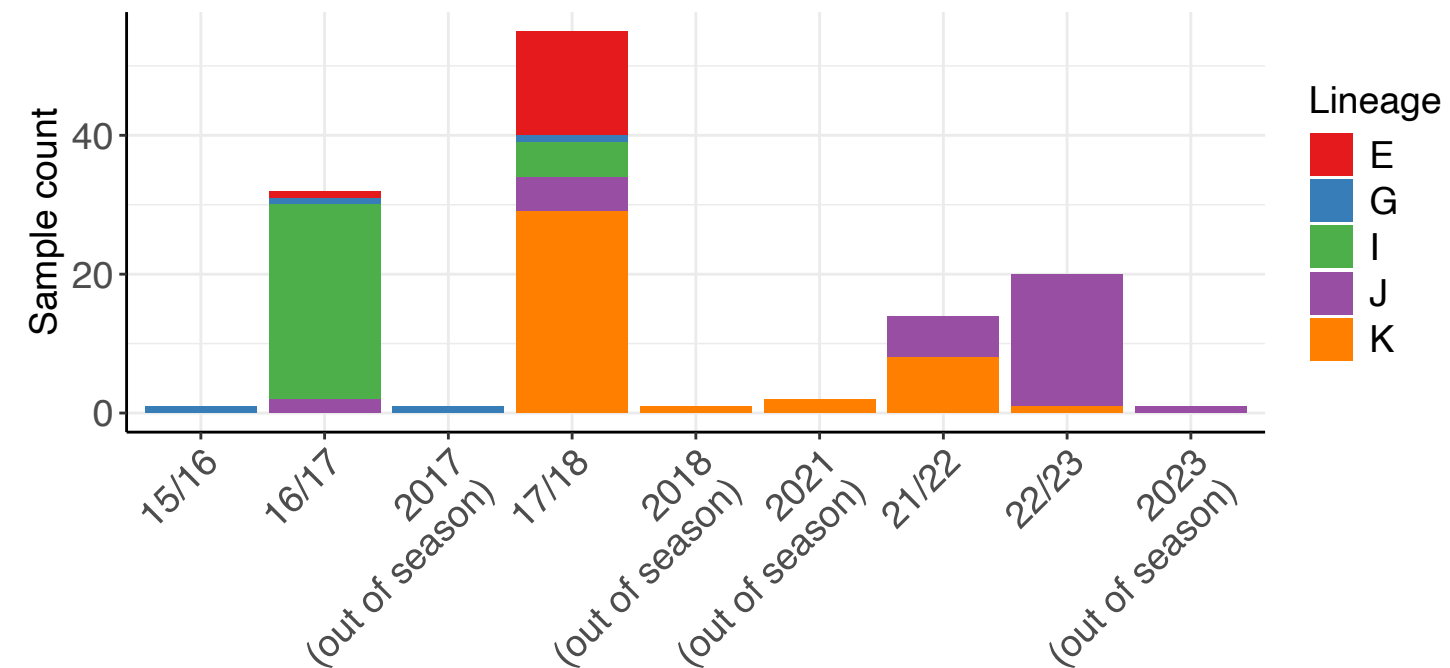
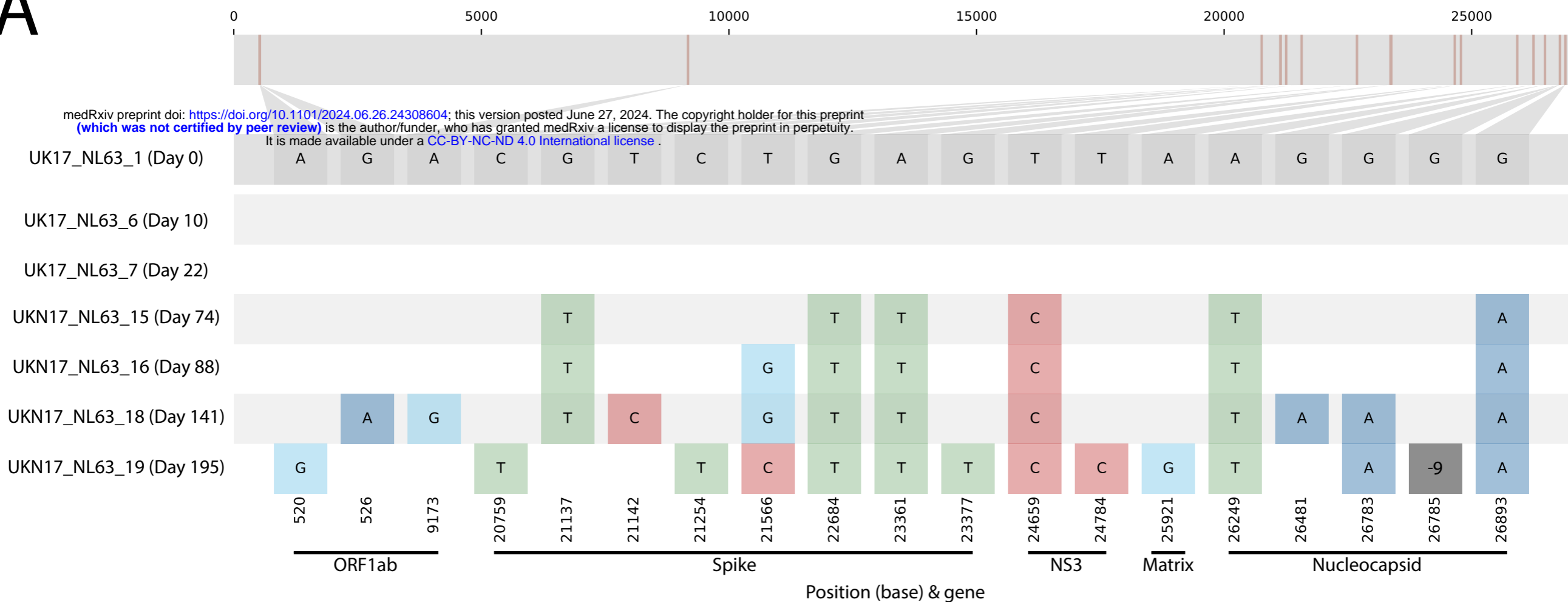
A: 229E**B: NL63****C: HKU1****D: OC43**

Figure 7: Lineages counts of Seasonal Coronaviruses by season. The lineages of 229E, NL63, HKU1 & OC43 Seasonal Coronaviruses (panels A to D respectively) were determined for each genome sequence of greater than 95% coverage, based on well supported clustering (>90 branch support) with lineage exemplar sequences as presented by Ye et al (16). Core infection season was defined as October to May (e.g. 16/17 for samples collected between October 2016 and May 2017), whilst June to September was considered out of typical season (e.g. 2017 for a sequence generated in June 2017).

A



B

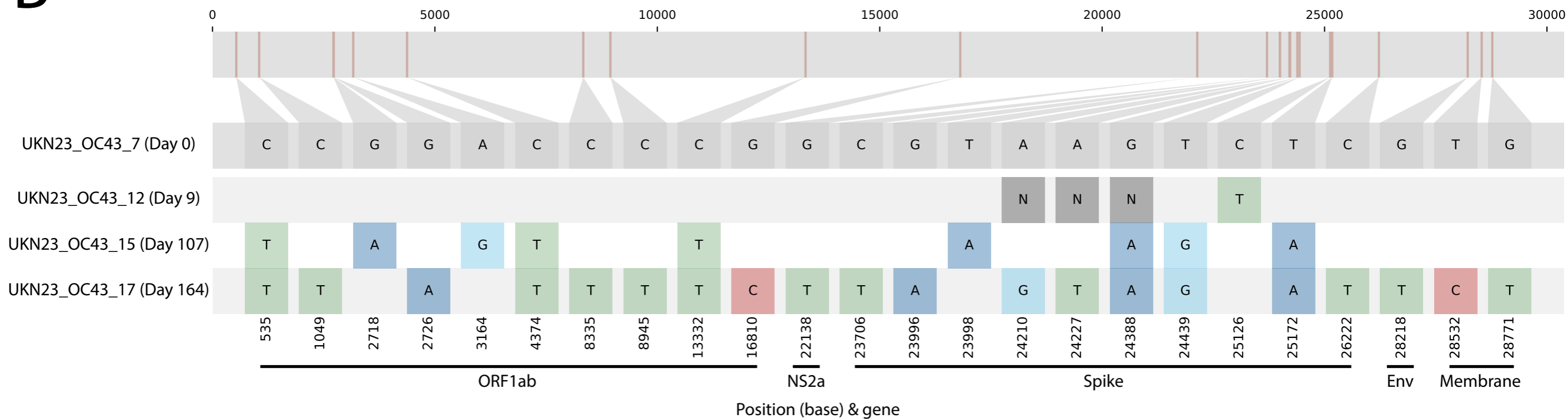
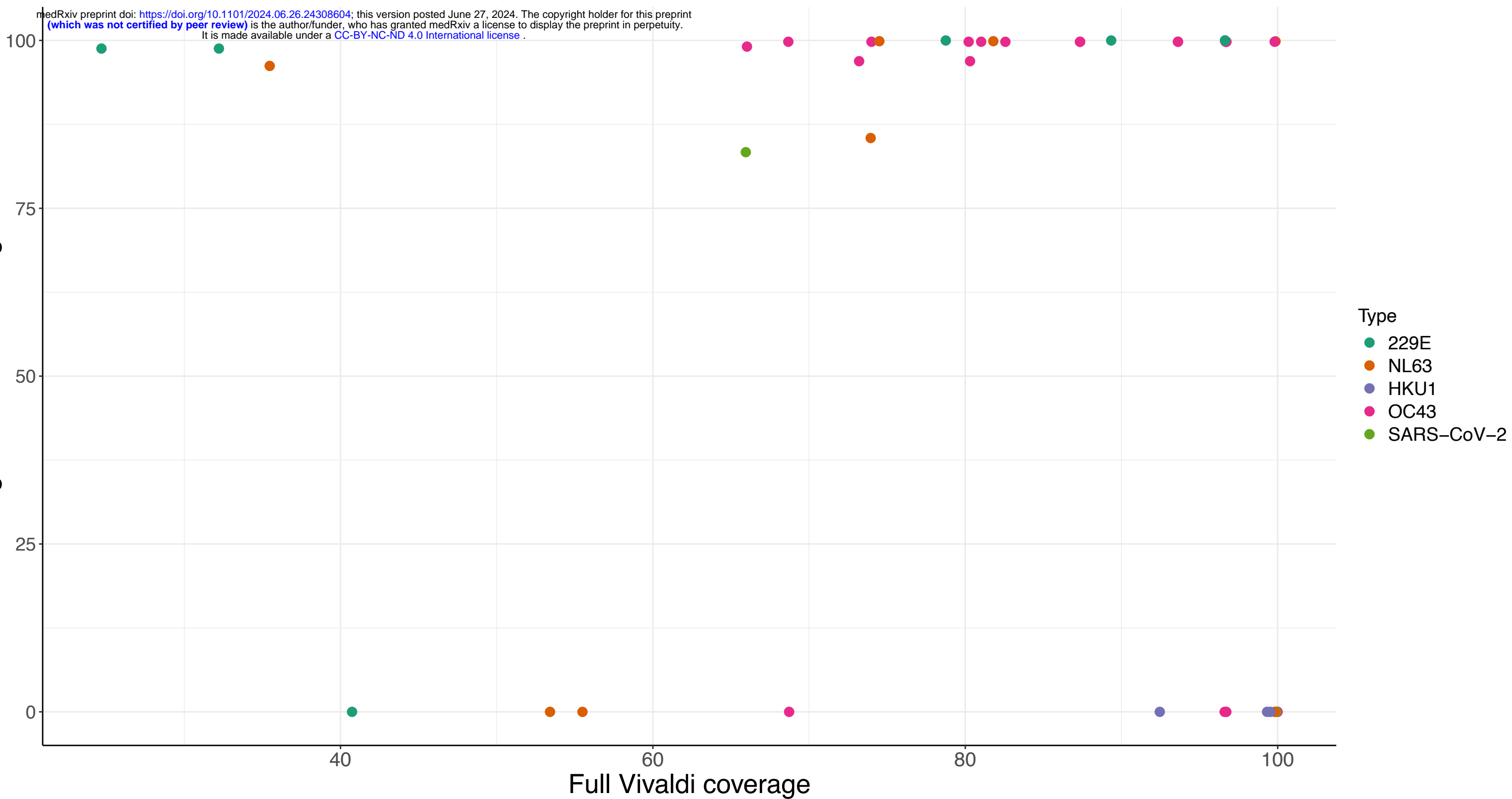


Figure 8: Putative prolonged Seasonal Coronaviral infection. Two sets of genome sequences were derived from the same individuals on dates suggestive of prolonged single infections, for both Alphacoronavirus NL63 (panel A) and Betacoronavirus OC43 (panel B). Sequence differences were visualised by Snip-it plots from the CIVET tool (<https://github.com/artic-network/civet>), using the first sample as the ‘Day 0’ timepoint and reference sequence. Genome nucleotide positions and genes are indicated relative to reference sequences LC488390 and LC756670 for NL63 and OC43 respectively. ‘N’ indicates sequence read dropout and ‘-9’ a 9bp deletion in the final NL63 nucleocapsid sequence.

Targeted Vivaldi coverage



Suppl. Figure 1: Genome coverage comparing Full Vivaldi four seasonal coronaviruses and SARS-CoV-2 combined multiplex to type-matched, targeted sequencing. The genome coverage of selected Seasonal Coronavirus 229E, NL63, HKU1 & OC43 and additionally SARS-CoV-2 samples when amplified by PCR multiplex containing all primer sets (Full Vivaldi, x axis), is shown compared to the genome coverage achieved with a single set of coronavirus-targeting primers (Targeted Vivaldi, y-axis). Genome coverage is expressed as a percentage of the total region targeted by the amplicon scheme. Samples with '0' coverage by the targeted method were sequenced only by the Full Vivaldi method.