SARS-CoV-2 Research Summary
Recombination in SARS-CoV-2

Prepared by:
Anna Maria Niewiadomska
& Richard H. Scheuermann

(Updated April 2022)

Summary
With the recent expansion of the SARS-CoV-2 Omicron variant lineages (BA.1 and BA.2), as well as their concurrent circulation with multiple Delta variant sublineages, several recombinant lineages have recently been discovered. Below, we summarize what is currently known about these lineages as well as provide some information on recombination in the *Coronaviridae* family in general and methods for recombination detection.

Recombination in Coronaviridae
Recombination has long been recognized as a method that many viruses use to increase their genomic diversity. This in turn can lead to selective biological traits that affect viral evolutionary trajectories, affecting diverse viral attributes from immune escape to more significant effects such as host species zoonotic events. These recombination events can lead to the emergence of novel virus variants in new host populations. The *Coronaviridae* family is well known to undergo both intra-species and, less frequently, interspecies recombination (1). Several recent reviews have focused on recombination in SARS-CoV-2 and the *Coronaviridae* family (2,3,4,5)

Co-infections and Sequencing Artefacts

Several factors complicate the identification of bona fide viral recombinants, including: 1) simultaneous detection of multiple viruses due to co-infections, 2) laboratory contamination, 3) variant related sequencing failures, 4) bioinformatic pipeline issues, and 5) convergent mutations.

1) Co-infection of hosts with 2 or more variants can occur during periods of high endemicity, when two (or more) variants are co-circulating. This in turn can lead to amplification of both variants from the same sample, resulting in a chimeric viral consensus genome that contains base calls from both separate variants.
2) Laboratory procedures are at risk of contamination when processing several patient samples, and can similarly lead to apparent chimeric viral consensus genomes.
3) As variations in new lineages accumulate, these can interfere with PCR primer binding, resulting in either amplification inefficiency or, if the difference is significant enough, complete amplicon dropouts, which can be misinterpreted as reversion to ancestral sequence if not evaluated carefully.
4) Sequence assembly pipelines and consensus calling parameter settings can also result in base calling errors, particularly in cases of mixed genomes, or in the misinterpretation of ambiguous nucleotide calls as ancestral.
5) Finally, if a recombinant sequence is suspected, the occurrence of convergent mutations in variants with the same single nucleotide polymorphisms, can make identification of a specific breakpoint difficult or impossible, particularly if breakpoints and/or convergent mutations occur at the genome extremities.

SARS-CoV-2 Research Summary
Recombination in SARS-CoV-2

Prepared by:
Anna Maria Niewiadomska
& Richard Scheuermann

Software/Methods used to detect recombination

Several recombination detection programs have been historically used to detect recombination in multiple virus families. However traditional tools can often miss recombination events when the parent sequences have a high degree of similarity, as is the case with SARS-CoV-2 variants. Therefore several groups have developed novel algorithms specifically tailored for the detection of recombination in SARS-CoV-2. Below we provide a list of commonly used methods for recombination identification in SARS-CoV-2, as well as for other viruses.

Developed specifically for SARS-CoV-2:

- Bolotie (detecting recombinations in viruses using large data sets with high sequence similarity). Command line program.
- RIPPLES (Recombination Inference using Phylogenetic PLacEmentS): Available as a command line program that detects long branches in a mutation annotated tree, then analyzes individual segments to detect up to two breakpoints. Can accept large datasets.
- Sc2rf - SARS-Cov-2 Recombinant Finder: a command line program designed to search for potential SARS-CoV-2 recombinants.

Broad application recombination detection:

- 3SEQ: command-line program. Input is a nucleotide sequence file (phylip or aligned fasta format); must include at least 3 sequences (2 parents, 1 child); program checks all triplet combinations for a mosaic recombination signal.
- GARD: available with a graphical user interface (GUI) or downloadable as a CL program; must include at least 3 sequences (2 parents, 1 child); uses likelihood-based model selection to search multiple sequence alignments for recombination breakpoints and identify putative recombinant sequences.
- PHI (pairwise homoplasy index): available as software packages on several platforms; this statistical method is able to detect recombination and compute significance by permutation tests, but is unable to identify breakpoints.
- RDP5 (Recombination Detection Program 5): A downloadable program with a GUI (Windows or Mac with windows emulator); combines multiple recombination detection methods in one package.
- Simplot: Similarity Plotting, bootscanning, quick tree generation, and informative-sites analysis.

Other programs can also be found here.

SARS-CoV-2 Research Summary
Recombination in SARS-CoV-2

Prepared by:
Anna Maria Niewiadomska
& Richard Scheuermann

Recombinant Lineage Nomenclature

In order to better define emerging SARS-CoV-2 epidemiological lineages, the "Phylogenetic Assignment of Named Global Outbreak Lineages" or "pangolin" tool was developed to assign alphanumeric lineage names (6). Later, provisional rules for naming recombinant lineages were defined by the team (7). Below, we provide a summary table of all recombinant SARS-CoV-2 lineages that have been assigned a pango lineage name (verified by the Pango Network Team), or have been referenced in a publication to date. Table includes lineage name, pango designation issue github number, proposed parental lineages, breakpoints, number of breakpoints, location first detected and references available (if any).
Note, genuine recombinants will often show evidence of epidemiological clustering, such as multiple sequences with similar spatiotemporal characteristics, originating from different laboratories.

Name	Github#	Lineage 1	Lineage 2	Proposed Breakpoint(s)	# brkpt	Location first detected	References
XA	NA	B.1.1.7	B.1.177	21,255 ≥ ≤ 21,764 (S)	1	UK	(8)
XB	#189	B.1.634	B.1.631	22,775 ≥ ≤ 22,778 (S)	1	N. America	(9)
XC	#263	Delta (AY.29)	B.1.1.7	Between ORF6 and ORF7a	1	Japan	(10)
XD	#444	Delta (AY.4)	BA.1	~21,643; ~25,581	2	France	(12)
XE	#454	BA.1	BA.2	≥ 11,537 (NSP6)	1	UK	-
XF	#445	Delta	BA.1	≥ 5,386 (NSP3)	1	UK	-
XG	#447	BA.1	BA.2	≥ 6,513 (NSP3)	1	Denmark	-
XH	#448	BA.1	BA.2	≥ 10,448 (NSP5 or NSP6)	1	Denmark	-
XJ	#449	BA.1	BA.2	≥ 13,448 (NSP12)	1	Finland	-
XK	#460	BA.1	BA.2	≥ 19,995 (NSP15)	1	Belgium	-
XL	#464	BA.1	BA.2	≥ 8393 (NSP3)	1	UK	-
XM	#472	BA.1.1	BA.2	≥ 17,410 (NSP13)	1	Multiple EU	-
XN	#480	BA.1	BA.2	2,834 ≥ ≤ 4,183 (NSP3)	1	UK	-
XP	#481	BA.1.1	BA.2	27,385 ≥ ≤ 29,509 (ORF6-N)	1	UK	-
XQ	#468	BA.1.1	BA.2	4,322 ≥ ≤ 5,385 (NSP3)	1	UK	-
XR	#469	BA.1.1	BA.2	4,322 ≥ ≤ 4,891 (NSP3)	1	UK	-
XS	#471	Delta	BA.1.1	9,055 ≥ ≤ 10,447 (NSP4/NSP5)	1	USA	-
NA	#439	Delta (AY.119)	BA.1.1	≤ 22,202	1	USA	(11)
NA	NA	Delta	BA.1	≤ 22,034	1	USA	(11)

SARS-CoV-2 Research Summary

Recombination in SARS-CoV-2

Prepared by:
Anna Maria Niewiadomska
& Richard Scheuermann

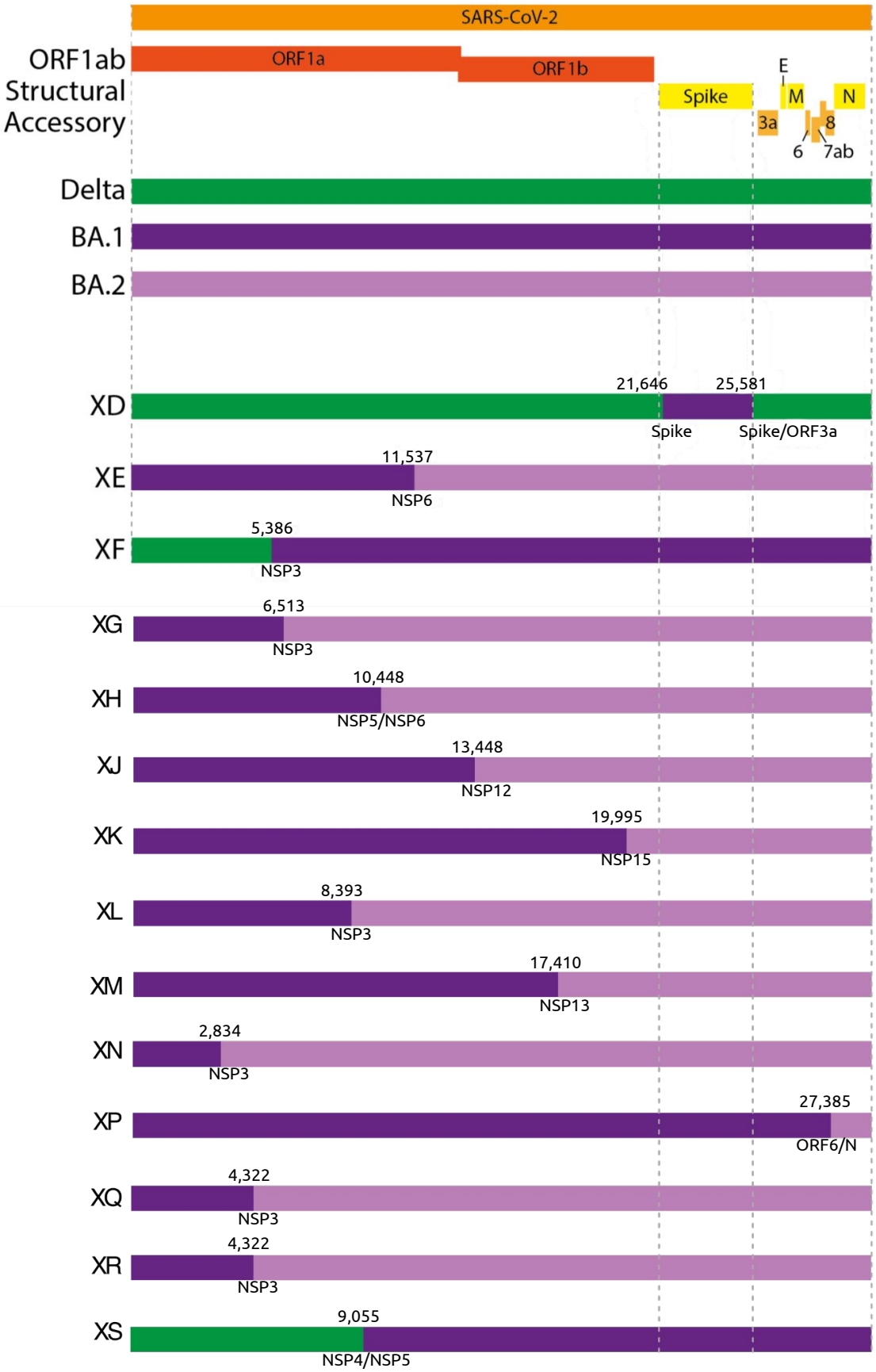

Figure 1: Proposed genomic structures of Omicron recombinant lineages (Adapted from UKHSA technical briefing 39): This figure shows the SARS-CoV-2 genome and genes (orange, red, yellow). Reference genomes colors: Delta (Green), BA.1 (purple), BA.2 (lilac). Recombinant lineages are colored by the associated references, along with inferred break point coordinates and genes*.

* Exact breakpoint coordinates and the genes in which they occur can be difficult to precisely define, and thus coordinates are merely minimum possible coordinates with respect to the 5' end of the genome.

Prepared by:
Anna Maria Niewiadomska
& Richard Scheuermann

References

1) Graham, R. L., & Baric, R. S. (2010). Recombination, Reservoirs, and the Modular Spike: Mechanisms of Coronavirus Cross-Species Transmission. *Journal of Virology*, *84*(7), 3134. https://doi.org/10.1128/JVI.01394-09

2) Pollett, S., Conte, M. A., Sanborn, M., Jarman, R. G., Lidl, G. M., Modjarrad, K., & Maljkovic Berry, I. (2021). A comparative recombination analysis of human coronaviruses and implications for the SARS-CoV-2 pandemic. *Scientific Reports*, *11*(1), 17365. https://doi.org/10.1038/S41598-021-96626-8

3) Fischer, W., Giorgi, E. E., Chakraborty, S., Nguyen, K., Bhattacharya, T., Theiler, J., Goloboff, P. A., Yoon, H., Abfalterer, W., Foley, B. T., Tegally, H., San, J. E., de Oliveira, T., Wilkinson, E., Msomi, N., Iranzadeh, A., Fonseca, V., Doolabh, D., Mlisana, K., … Korber, B. (2021). HIV-1 and SARS-CoV-2: Patterns in the evolution of two pandemic pathogens. *Cell Host & Microbe*, *29*(7), 1093. https://doi.org/10.1016/J.CHOM.2021.05.012

4) Giorgi, E. E., Bhattacharya, T., Fischer, W., Yoon, H., Abfalterer, W., & Korber, B. (2021). Recombination and low-diversity confound homoplasy-based methods to detect the effect of SARS-CoV-2 mutations on viral transmissibility. *BioRxiv*, 2021.01.29.428535. https://doi.org/10.1101/2021.01.29.428535

5) Garushyants, S. K., Rogozin, I. B., & Koonin, E. V. (2021). Template switching and duplications in SARS-CoV-2 genomes give rise to insertion variants that merit monitoring. *Communications Biology 2021 4:1*, *4*(1), 1–9. https://doi.org/10.1038/s42003-021-02858-9

6) O'Toole, Á., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J. T., Colquhoun, R., Ruis, C., Abu-Dahab, K., Taylor, B., Yeats, C., du Plessis, L., Maloney, D., Medd, N., Attwood, S. W., Aanensen, D. M., Holmes, E. C., Pybus, O. G., & Rambaut, A. (2021). Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evolution*, *7*(2). https://doi.org/10.1093/VE/VEAB064

7) Pybus, O. (2021). *Pango Lineage Nomenclature: provisional rules for naming recombinant lineages - SARS-CoV-2 coronavirus* - *Virological*. https://virological.org/t/pango-lineage-nomenclature-provisional-rules-for-naming-recombinant-lineages/657

8) Ben Jackson, Rambaut, A., Pybus, O. G., Robertson, D. L., Connor, T., Loman, N. J., & Consortium, T. C.-19 G. U. (COG-U. (2021). *Recombinant SARS-CoV-2 genomes involving lineage B.1.1.7 in the UK - SARS-CoV-2 coronavirus / SARS-CoV-2 Molecular Evolution* - *Virological*. https://virological.org/t/recombinant-sars-cov-2-genomes-involving-lineage-b-1-1-7-in-the-uk/658

9) Gutierrez, B., Castelán Sánchez, H. G., Da, D., Candido, S., Jackson, B., Fleishon, S., Ruis, C., Delaye, L., Rambaut, A., Pybus, O. G., & Escalera-Zamudio, M. (2021). Emergence and widespread circulation of a recombinant SARS-CoV-2 lineage in North America. *MedRxiv*, 2021.11.19.21266601. https://doi.org/10.1101/2021.11.19.21266601

10) Sekizuka, T., Itokawa, K., Saito, M., Shimatani, M., Matsuyama, S., Hasegawa, H., Saito, T., Kuroda, M., & (COG-JP), C.-19 G. S. N. in J. (2021). Genome Recombination between Delta and Alpha Variants of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). *MedRxiv*, 2021.10.11.21264606. https://doi.org/10.1101/2021.10.11.21264606

11) Bolze, A., White, S., Basler, T., Rossi, A. D., Roychoudhury, P., Greninger, A. L., Hayashibara, K., Wyman, D., Kil, E., Dai, H., Cassens, T., Tsan, K., Nguyen, J., Ramirez, J., Carter, S., Cirulli, E. T., Barrett, K. S., Belda-Ferre, P., Washington, N. L., … Luo, S. (2022). Evidence for SARS-CoV-2 Delta and Omicron co-infections and recombination. *MedRxiv*, 2022.03.09.22272113.

12) Simon-Loriere, E., Montagutelli, X., Lemoine, F., Donati, F., Touret, F., Bourret, J., Prot, M., Munier, S., Attia, M., Conquet, L., Nguyen, S., Amara, F., Maisa, A., Fournier, L., Brisbarre, A., Dehan, O., Levillayer, L., Gunalan, V., (DCGC), T. D. C.-19 G. C., … Paquin, A. (2022). *Rapid characterization of a Delta-Omicron SARS-CoV-2 recombinant detected in Europe.*