1. Recombination in Coronaviruses Other Than SARS-CoV-2

Recombination represents a major contributor to RNA virus evolution [1] together with re-assortment (which exclusively operates in RNA viruses with segmented genomes). Recombination can occur both in segmented [2,3] and non-segmented RNA viruses [4–7] and avoids an accumulation of irreversible deleterious mutations typical of asexual reproduction (so called "Muller's ratchet" [8]). "Donor" and "acceptor" are conventional terms used to refer to the strain represented in a greater and lesser amount, respectively. Recombination within different sublineages of the same virus invariably requires co-circulation and co-infection of the same host.

Recombination can be difficult to detect whenever the sublineages have minimal differences and requires whole-genome sequencing (WGS). Unlabeled private mutations can help track the spread of the recombinant lineage more easily: they are defined as private mutations that are neither reversions nor labeled (i.e., they are not mutations to a genotype that is known to be common in a clade) [9]. Deletions are generally considered useful landmarks for recombination because they are unlikely to be reverted (except through recombination), but they can spontaneously occur across different sublineages (convergent evolution) independently from recombination (as seen, e.g., by the NSP6 SGF- reported in Alpha, Beta, and Gamma variants of concern of SARS-CoV-2 [10]).

We review here former evidence for recombination in betacoronaviruses and then focus on SARS-CoV-2 recombinants.

Coronaviridae undergo both homologous recombination (HR) and nonhomologous recombination (NHR). Only a minority of recombinants are likely detected in surveys since most of them are unlikely to be fitter than the currently dominant strain.

2. Recombination in SARS-CoV-2

2.1. Recombinant Origin of SARS-CoV-2

Li et al. initially showed in March 2020 that SARS-CoV-2's entire receptor-binding motif (RBM) was introduced through recombination with coronaviruses from pangolins, possibly a critical step in the evolution of SARS-CoV-2's ability to infect humans [40]. This was later confirmed by Zhu et al. in December 2020 [41]. However, more recently, using sliding window bootstrap (SWB) to highlight the regions supporting phylogenetic relationships, SARS-CoV-2 was defined as a mosaic genome composed of regions sharing recent ancestry with three bat SCoV2rCs recently discovered in the Yunnan region of China (RmYN02, RpYN06, and RaTG13) or related to more ancient ancestors in bats from Yunnan and Southeast Asia [42], with no evidence of direct recombination with pangolin viruses.

2.2. Super-Infection or Co-Infection with Different SARS-CoV-2 Lineages

SARS-CoV-2 can be named according to different phylogenetic systems, which can often but not always be reconciled. The Global Initiative on Sharing All Influenza Data (GISAID) phylogeny classifies clades with progressive letters (https://www.gisaid.org/index.php?id=208, accessed on 29 April 2022). The Phylogenetic Assignment of Named Global Outbreak LINeages (PANGOLIN) nomenclature uses an alphabetical prefix and a numerical suffix to identify descendants (https://www.pango.network/the-pango-nomenclature-system/statement-of-nomenclature-rules/, accessed on 29 April 2022). NextStrain uses a year-letter system (https://docs.nextstrain.org/projects/ncov/en/latest/reference/naming_clades.html, accessed on 29 April 2022). Finally, the WHO uses progressive Greek letters to dynamically identify variants of interests (VOI) or concern (VOC) (https://www.who.int/activities/tracking-SARS-CoV-2-variants, accessed on 29 April 2022).

A few months after the initiation of the COVID-19 pandemic, co-infections were documented without any evidence of recombination. The first detailed case was described in February 2021 as co-infection from NextStrain 20A and 20B lineages, which was followed up for kinetics of relative abundance: a Portuguese patient had a prolonged viral shedding case (97 days long), first with a severe disease manifestation followed by a short second hospitalization episode, in an otherwise healthy young female [43].

2.3. Evidence for Recombination in SARS-CoV-2

There is both *in silico* [51] and in vivo [52] evidence for recombination of different SARS-CoV-2 strains. Studies relying on linkage disequilibrium identified that SARS-CoV-2 recombination occurs at very low levels [52–54] or does not occur at all [55–60]. Several alternative methods are available for reconstructing genealogies explicitly in the presence of recombination, both with [61] and without [62–64] making the parsimony assumption, but none is tailored to the particular problem of detecting recombination in the presence of recurrent mutation. In fact, many tests of recombination assume that all mutations can only occur once at each site, and hence, recurrent mutation from convergent evolution (as it occurs in SARS-CoV-2) and systematic errors can confound signatures of recombination [7,27,36,65].

Hence, novel methodological approaches have been developed to detect recombinant genomes in SARS-CoV-2 lineages. Ignatieva et al. proposed a parsimony-based greedy heuristic algorithm for reconstructing plausible ancestral recombination graphs (KwARG) [66]: it does not scale well to large datasets but was powerful enough for disentangling the effects of recurrent mutation from recombination in the history of a sample [67]. Turakhia et al. developed Recombination Inference using Phylogenetic PLacEmentS (RIPPLES) to break the sequence into distinct segments that are differentiated by mutations on the recombinant sequence and separated by up to two breakpoints: for each set of breakpoints, RIPPLES places each of its corresponding segments using maximum parsimony to find the two parental nodes—hereafter termed donor and acceptor. RIPPLES is very fast with a large dataset but is biased against identifying recombination events near the edges of the viral genome. They identified 606 recombination events by investigating a 1.6M sample tree, showing that approximately 2.7% of sequenced SARS-CoV-2 genomes have recombinant ancestry, that recombination breakpoints occur disproportionately in the Spike protein region, and that cases were coinfected with 2–3 SARS-CoV-2 variants on average [68].