https://twitter.com/Daoyu15/status/1630378021667602433

Daoyu
@Daoyu15
.
5:23 PM · Feb 27, 2023
The current environment and evolutionary landscape for SARS-CoV-2 is much different from the pre-pandemic era in term of allowed evolutionary changes.
Why?
The rampant use of reverse transcriptase in SARS-CoV-2 testing and genome sequencing is on a scale that is unlike anything

Before the pandemic, and the SARs-CoV-2 pandemic is the first time (if not counting prior research lab activities) where A Coronavirus sequences encounters reverse transcriptase on a daily basis—this is present in both qRT/PCR (“nucleic acid test”) tests and genome sequencing, as

The common step for the synthesis of cDNA before both is to use a reverse transcriptase and random hexanucleotide primers to convert RNA into DNA. Extracted and amplified SARS-CoV-2 nucleic acid materials are not considered infectious, so in many mass testing facilities these

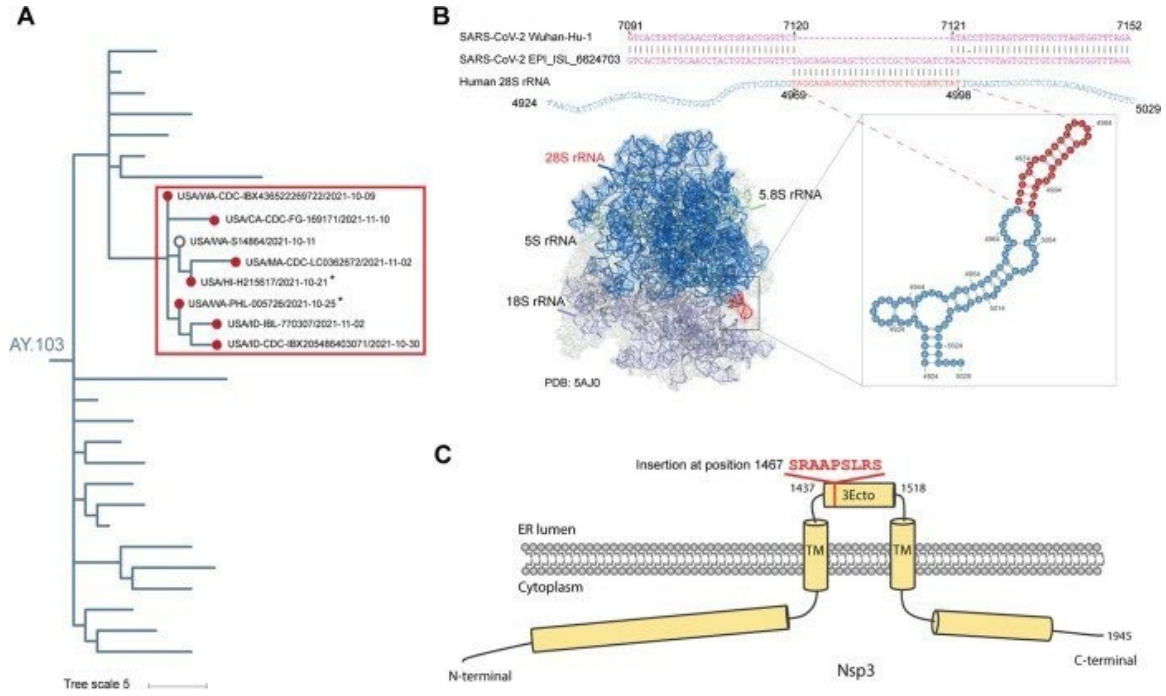
Amplicons were disposed without complete inactivation of the DNA.
Bacterial cells especially those found in nature are often competent—meaning that they will sample DNA, not just plasmid DNA but any kinds of single-stranded or double-stranded DNA—from the environment.

After sampling, ssDNA is allowed to interact with DNA within the bacterial genome, where not only HDR(Homology-directed recombination) but also NHEJ (non-homologous end joining) and MHEJ (microhomology-directed end joining) takes place. Yes. DNA recombination can often not

Require surrounding homology to take place because a DNA based organism have multiple repair pathways and many different machineries for DNA recombination that are absent in RNA viruses.
Every one of the >10 million SARs-CoV-2 genomes on GISAID and at least 10 times more

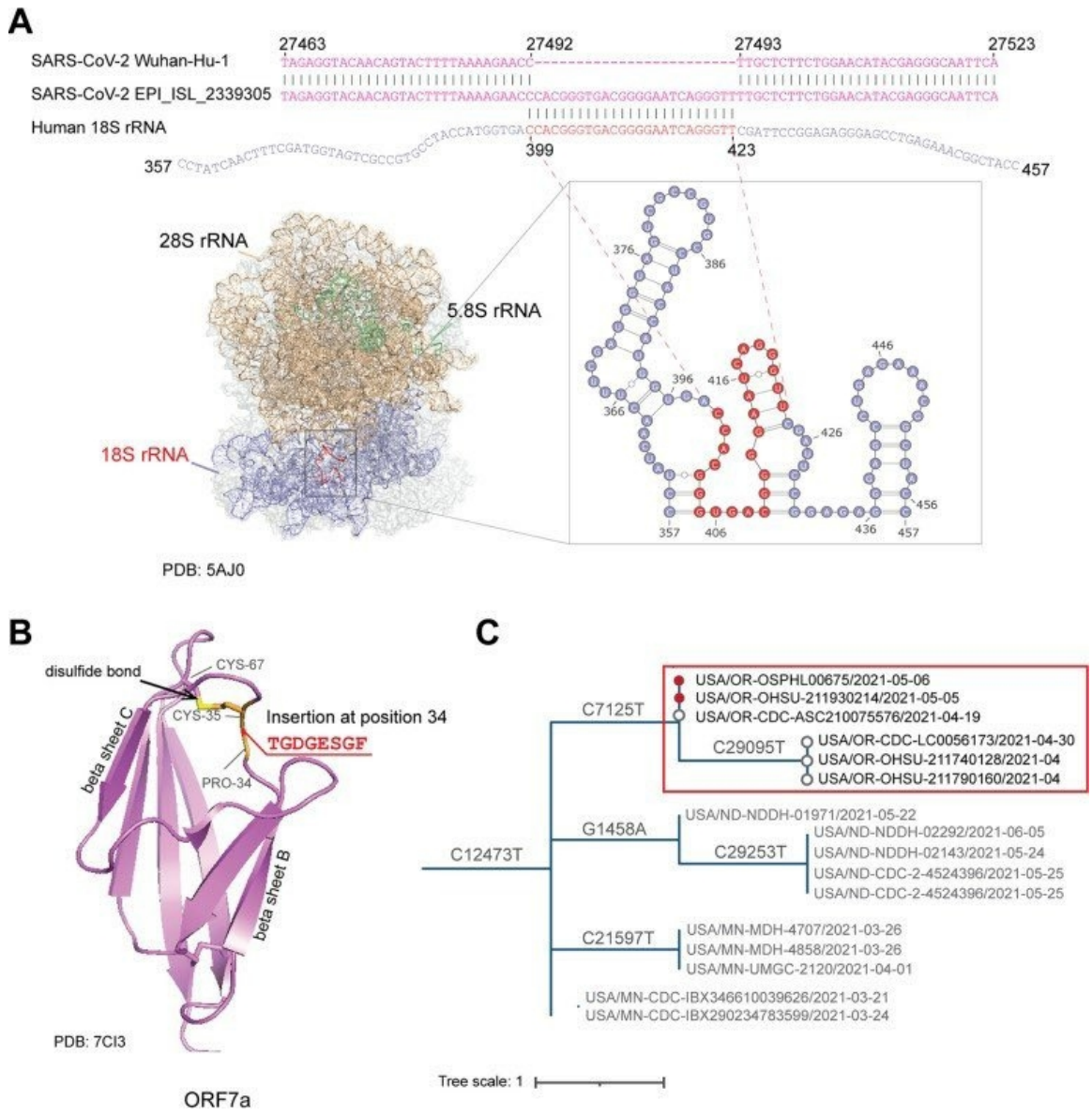
Than this number that weren’t on GISAID due to low quality in the sequencing process, originated from the unnatural process of reverse transcription and amplification. That bacteria can then cache for later use.

What this means for the evolution of SARs-CoV-2?
It enables the genome to sample fragments of nucleic acids that were not possible before the pandemic in nature.
Both bacterial recombination, DNA recombination within a human host and sometimes, genome sequencing-related



Recombination can take place. One of The “18S rRNA inserts” that were found in some SARS-CoV-2 genomes in 2021 is a prime example of TA ligation cloning being the cause of the insertion, as the insert is bordered by two “T” nucleotides—TA ligation is a common process used during

ILLUMINA library preparation, where residual libraries are often not autoclaved as in year 2021. The Reverse transcriptase itself can also lead to insertions, as unlike the RNA-dependent RNA polymerase, Reverse transcriptases have terminal deoxynucleotidyl transferase activity



Meaning that the small (one or two bp) homology become extended into larger homology by random nucleotide insertions onto the end of the cDNA. This causes a similar 18S insertion in the same lab complex where the homology is a single T or a single C. As these cDNA strands become

Taken up by bacteria, inserted into the genome, and then amplified,

https://zenodo.org/record/3786451#.Y_1bsyW8klQ
(as seen in one of the early pandemic Bacterial assemblies, protein EEU8328811.1, where the SRA contained only bacteria and SARS-CoV-2) they eventually get into the environment,

https://zenodo.org/record/3786451#.ZDmnSLzMKJB

Enter the human intestinal tract, where the DNA from the environment is taken up during the digestion process and interaction with human microflora, gets transcribed and then recombine frequently with the SARS-CoV-2 that often also replicate there—the addition of homology

Segments already done during cDNA production (most common) or within bacteria (less common). If this then confer an immunological advantage, it can then replicate, form a small cluster before in most cases, disappear as these chimeric nucleotide sequences are originally adapted

For bacterial growth advantage and pose little benefit once they reenter the SARs-CoV-2 genome, mostly only immunological and temporary in nature. Never in history were so many genomes of one virus being sequenced this many times in humans, and never in history were the reverse

Transcriptase being used on a global scale at such a high frequency. The human SARS-CoV-2 in the pandemic, after the beginning of worldwide mass testing, therefore gained the quality in term of evolution not only of RNA viruses but also retroviruses, thanks to the

Never-before-seen focus on its genome by man-made reverse transcriptases.
One of the reasons why “insertions” in the SARS-CoV-2 genome only begun to pop up after the end of 2020.
Once you begun doing reverse transcription on an industrial scale, the world, the environment and

The environment within cell culture labs can no longer be said to be free of “splints” of DNA/RNA where SARS-CoV-2 could use to circumvent the homology barrier against recombination.

4 billion human infections vs at most 1000 hypothetical animal infections, 10 million sequenced pandemic SARS-CoV-2 genomes vs less than 1000 positive prepandemic samples for Sarbecoviruses. Rampant reverse transcription in 2021-2023 vs no reverse transcription on the SARS-CoV-2

Progenitor genome at the “right” locations for “spillover” before the end of 2019. (Other than inside research labs)

Daoyu
@Daoyu15
.
Mar 2
This is one of the things that can be caused by reverse transcriptases. Especially certain bacterial plasmids that proliferated with sequences that were duplicated from reverse-transcribed SARs-CoV-2 S1-S2.

---

https://mobile.twitter.com/Daoyu15/status/1630041787233611776

Daoyu
@Daoyu15
It should be noticed that once CGG-CGG is present within the SARS-CoV-2 S1-S2 junction, then it can obviously be duplicated into other locations of the genome, with some mutations happening as it needs to then suit their role at that location.
Also important is that these are

Only discovered three years into the pandemic and with over 4 billion human infections. This is a higher number than the entire supply of wild or farmed animals in China and 6 orders of magnitudes higher than any of the potential supply route to Huanan specifically.

The “lineage” of course have now went extinct, due to the immune burden once it is in the general population.

And found after more than 10000000 genomes have been sequenced on GISAID. This is 4 orders of magnitude higher than the entire supply of genomes previously sequenced for Sarbecoviruses and incidentally also 4 orders of magnitude higher than the maximal number of animals that

Can become involved in a HSM spillover scenario.

This have similar mechanism as the R203K/G204R in the B.1.1 lineage. However, strong immunogenicity mean that it is now extinct.