Twitter Thread by Francisco de Asis

Francisco de Asis
@franciscodeasis

[Thread] In-silico molecular overclocking of RaTG13.
TLDR: 191-nt RdRp segments of SARS-CoV-2, RaTG13 and Ra7896 show an unexpected molecular clock behavior. In-silico synonym mutations is one probable explanation.

	W-H-1	15710	15714	15720	15729	15738	15747	15750	15753	15763	15789	15801	15804	15810	15813	15837	15891	15895	15897	15898	15900
	RaTG13	15692	15696	15702	15711	15720	15729	15732	15735	15745	15771	15783	15786	15792	15795	15819	15873	15877	15879	15880	15882
	7896	401	405	411	420	429	438	441	444	454	480	492	495	501	504	528	582	586	588	589	591
Wuhan-Hu-1 - MN908947.3		T	C	C	T	C	T	T	A	C	G	T	T	C	T	T	A	C	A	G	T
RaTG13 - MN996532.1		T	T	T	T	C	T	T	A	C	A	C	T	C	C	T	A	C	A	G	T
7896 - MN312671		T	C	C	C	T	C	C	G	T	G	T	C	T	T	C	T	A	G	A	T

Assuming SARS-CoV-2 is an ancestor of Ra7896 (used in RaTG13) does not fully explain it. Implied evolutionary rate would be in the order of 10^-2 substitutions/site/year in a well-conserved part of the genome, far from a normal ~10^-3 for a complete genome
https://t.co/5blQ4IxsxY

> I think I know what happened here in THIS 191-nt segment. All 3 statements are true and:
> - published RaTG13 is real Ra7896
> - SARS-CoV-2 backbone is parent of real Ra7896
> - published Ra7896 is real Ra79XX (any of the other 7 of the clade)https://t.co/E1YrOZanxq
>
> — Francisco de Asis (@franciscodeasis) March 16, 2021

So, it is not only SARS-CoV-2 having its molecular clock frozen, but also RaTG13 molecular clock running more than expected.
@Nerdhaspower and @Quay_Dr have already noted strange patterns of synonym mutations along the genome of RaTG13

I always make this assumption: WIV ability of faking sequences is not unrestricted. They would never fake aa seqs, due to the protein folding problem. In a few months or years, they would be discovered
https://t.co/13athRfPIa

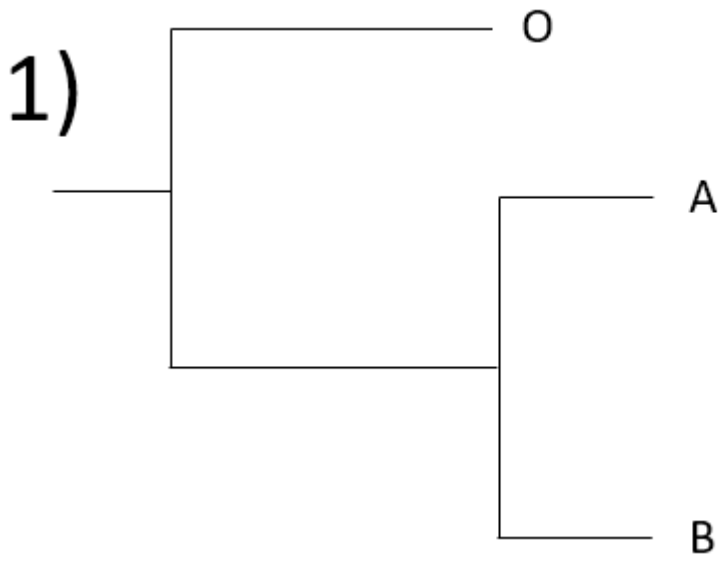
So, in silico, WIV would never make:
- non-synonymous mutations
- splicing within genes
But they could make:
- synonymous mutations
- swap genes or the complete genome from other real viruses

Why fabricating RaTG13?: To make it appear more distant than it really is. You probably never heard of that 98.65% identity between SARS-CoV-2 RdRp and Ra4991 RdRp. It was very dangerous!
https://t.co/WurzUVwW0W

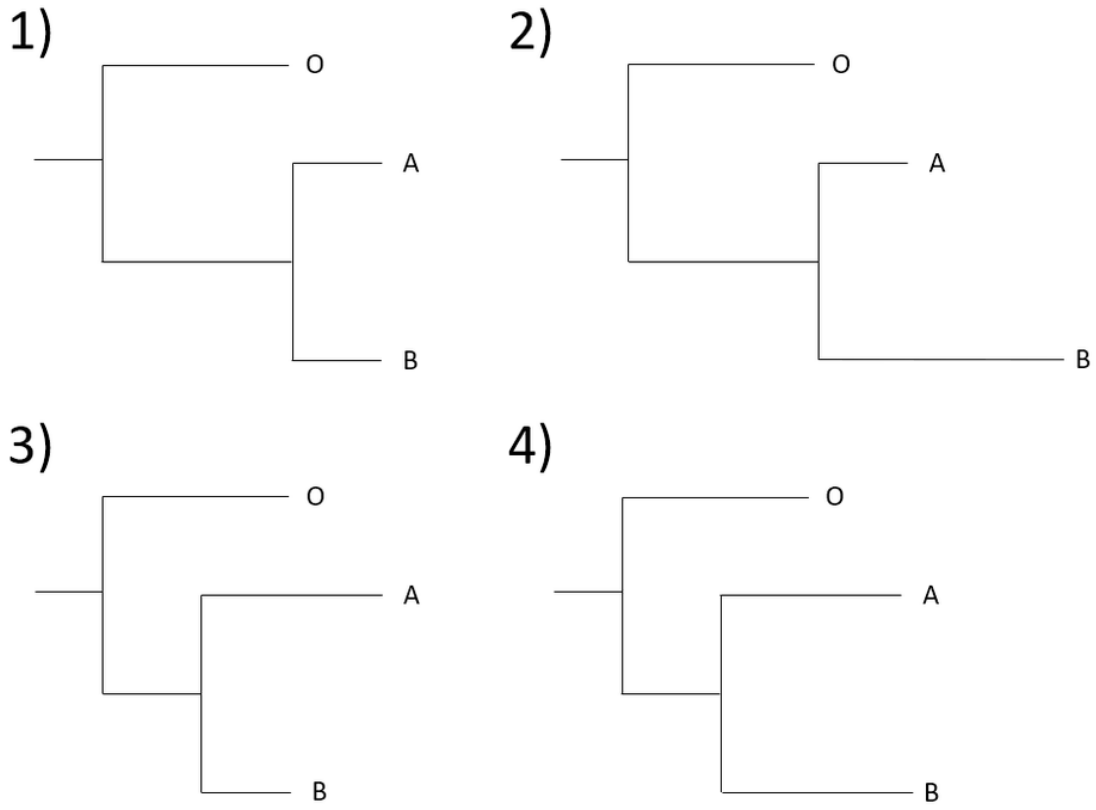
> Exciting time machine journey to WIV the day the pandemic is declared
> [BLASTing SARS-CoV-2 RdRp]
> "Oh, let's say \u201chigh sequence identity\u201d and move to a lower complete genome identity with a sequence we now disclose as new despite having it since 2018"https://t.co/OUpPNeJ7KF pic.twitter.com/0Q4wnNkfj3
>
> — Francisco de Asis (@franciscodeasis) March 13, 2021

But, if you want to make one virus appear more distant to another one without being noticed, you cannot just make orthogonal synonymous mutations, because you can get caught with the phylogenetic trees if you do not do it wisely

WIV forgot that they would eventually publish 7896 RdRp that could serve as a close outer group for SARS-CoV-2 and RaTG13.
Imagine Fig. 1 is real/base situation, A is fixed and you want B more distant.
Note: A is SARS-CoV-2, B is RaTG13 and O is the outer group (clade 7896)



If you just add orthogonal synonymous mutations to B, the clock is distorted (Fig 2, A & B not contemporaries). Correct way of faking would have been making a few backward synonymous mutations towards the outgroup (Fig 3), and then a few orthogonal synonym mutations (Fig 4)



It seems that WIV forgot the backward mutations!
I was thinking if it was better to keep this secret until WIV publish their next paper of the clade 7896 to let them commit the error again. But it clearly shows up in any tree. Their problem was not checking it this short segment