Investigating two possible Origins of SARS-CoV-2 - an RNA Analysis on Tree Spaces

Roland Moore, Vic Patrangenaru and Adam Dixon

January 5, 2021

Abstract

One regards spaces of rooted trees as stratified spaces. Spaces of rooted phylogenetic trees are used to define the genetic distance from the root species to its leafs. In particular, tress with three have a well defined genetic distance between older species and recent species. One applies such ideas to analyze RNA sequences of SARS-CoV-2 from multiple sources, by building samples of trees with three leafs in MEGA, and by running nonparametric statistics for intrinsic means on such rooted trees with three leafs. One computes the evolutionary distance to the root information, to discern if SARS-CoV-2 viruses that were sequenced at the early stage of the COVID 19 pandemic have more likely the bat-virus RaTG13, or a Wuhan-virus as ancestor, and our data points to the second as being evolutionary closer.

Keywords: phylogenetic tree, tree space, stratified space, RNA aligned sequences, SARS-CoV-2, genetic distance

1 Introduction

Each historical era has been coined by scientific and technological progress, or regress, and the twenty first century is no exception; it is the century of Biology (see eg Glover (2012)[6]). At the beginning of the century of Biology, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) techniques have intensified, as a basis of the acclaimed CRISPR-Cas9 and CRISPR-Cas13 genome editing methods, that can be programmed to target specific stretches of genetic code to edit DNA or RNA at given locations, for various purposes, to begin with, presumably as diagnostic tools.

With an ever increasing human population, one should nevertheless recognize that some biologists with access to CRISPR-Cas, might be drawn by neo-Malthusianists to use their skills into developing a biological weapon, including but not limited to an airborne pathogen targeting innocent civilians.

Organisms with **functions**, so that can interbreed, form a **specie**, and phylogeny is the ancestral history of a specie or a class of species. Note that great minds got interested in the development an external fast adaptable immune system using small organisms, may help immunocompromised individuals to fight disease using programable bacteria, making it to mutate faster, with aid of plasmids and viruses; with a CRISPR-Cas9 fast amplification device, one could for example create new biological functions in a controllable way (see Gromov(2018)[8]). The idea is to modify the RNA or DNA of simple organisms, for **gain of function** to improve the ability of such species to multiply at an exponential rate (see Gromov(2018)[8]), in a hope of serving medicine, potentially better than classical biochemical cures. There are companies that aim at using messenger RNA (mRNA) techniques to help create new treatments of hard to fight diseases, such as cancer, or Covid-19 (see e.g. Moderna (2021)[3], Copin et al.(2021)[5]). Indeed after the failure of earlier HIV

vaccine trials (see Adepoju (2020)[1]), Oxford University began Phase I HIV vaccine trials, with results expected to be available in April 2022, and Moderna is set to follow suit, while Regeneron's monoclonal antibody cocktail against Covid 19, is already FDA approved.
With a frenzy of mRNA vaccination research during the century of Biology, including successes and failures, it is hard to fathom, why the current pandemic could not source from a biotechnology experiment that went wrong, especially since the presumption of an intermediate host could not be validated, as no animal host was found, close to two years after the start of the epidemic. Given advances in CRISPR biotechnology, editing relatively short genome sequences, such as viral sequences should be a fairly easy task.
While one should not discard the hypothesis that SARSCov2 could well be a lab creation (see [15]), based on the half century or so time period, that would take RaTG13 virus to evolve into SARSCov2, and on the public opinion (see Campbell(2021)[2]), absent sequences of lab edited bat virus sequences, testing such a hypothesis is a "needle in a haystack" problem, therefore in this paper we limit ourselves to a weaker hypothesis testing problem, namely based on available SARSCov2 RNA sequences, we test if SARSCov2 specimens are on average closer or farther apart from a RaTG13 bat-SARS virus than from a Wuhan-SARS virus. Our analysis is run on phylogenetic tree spaces. To lower the dimensionality, while at the same time, increase the sample size, we focus on rooted trees with three leafs.
A phylogenetic tree (PT) is an equivalence class 1-connected directed graphs, having a root (the common ancestor of all other vertices), and leafs, that are vertices with no descendents, that are representing (the DNA/RNA of) a given species, assumed to be known, while intermediate species (vertices between leafs and the root) are unknown.

Our paper is organized as follows: in Section 2 one introduces a basic Biology dictionary, to describe the data that is used in our analysis. The third section is dedicated to stratified spaces, with a focus on spaces of phylogenetic trees with a fixed number of leafs. For rooted trees, one defines the evolutionary distance between a root specie and a leaf specie. This distance plays a key role in the hypothesis testing analysis. Section four is computational, being dedicated to RNA alignment and RNA based phylogenetic tree building. Section five is dedicated to testing the hypothesis testing on which of the two viruses is most likely to be at the origin of the SARS-Cov2, based on their average genetic distance to the toot of a tree: the bat-SARS virus than from a Wuhan-SARS virus. The paper ends with a discussion on implications of the findings in section four, and cautionary recommendations on CRISPR-Cas advances.

2 DNA structure and RNA data

Gregor Mendel (1865,[9]) was the first one to suspect that *genes* are heritable factors causing differences in organisms' physical traits such as shape, size or color. Genes are composed of a sequence of macromolecules, DNA (Deoxyribonucleic Acid)-adenine, thymine, guanine, and cytosine (ATGC) sequence units, and RNA (Ribonucleic Acid)- adenine, uracil, guanine, and cytosine (AUGC) sequence units. The four bases consist of two groups of base pairs: purines (AG) and pyrimidines (UC in RNA, TC in DNA). For example, about 3.3 billion of these DNA base pairs exist in the human genome, all genetic information of a human. The precise alignment of the base pairs in the double helix structure of the DNA, determines the role of any particular gene. The double helix structure of the DNA (due to the two parental heritages of the offspring) has strings of AT alternating with another string of GC like steps on staircase. The hall of fame DNA history began with Rosalind Franklin and Raymond Gosling historical X-ray diffraction Photo 51 (see [13]), correcting an earlier model by Pauling and Corey(1953)[11].

Most viruses causing serious diseases in humans, including SARS-Cov2 have their genome made entirely of RNAs. SARS-Cov2 is such a virus, a coronavirus. Coronaviruses have the largest known RNA genomes, between 27K and 32K basepairs (bp) in length (see Clinton Smith and Denison (2013)[7]). An similar RNA based virus, SARS-CoV, which caused a brief pandemic in 2003 and had similar characteristics with SARS-CoV-2. Knowledge and experience from the SARS-CoV pandemic helped expedited the studies and vaccine making on SARS-CoV-2 since they are similar in many ways [12]. However, their genomes suggest they came from different sources [14], and are therefore, is not the same virus. SARS-CoV is known to be relatively more fatal than SARS-CoV-2 but spreads less in infections. For a history of the SARS-Cov Pandemic in 2003 see Cristianini and Hahn (2006) [17]. Note that while the World Health Organization (WHO), was first notified in 2006 of the severe acute respiratory syndrome (SARS), from a relatively small private hospital in Hanoi, capital of Vietnam, and issued a global alert on 15 Mar 2003, labelling this SARS as a "worldwide health threat" when similar cases were found in Canada, Indonesia, Philippines, Singapore and Thailand, it was later learnt that the first case of SARS appeared in a Chinese province in 2002. By late April, there was 4,300 SARS cases and 250 deaths. It has been documented that, only in April 2003, China apologized for its slow outbreak response. According to CDC, this 2003 SARS-CoV outbreak was a virus that had never been found in humans before. It infected 8,096 people of which 774 died. Only eight people from US got infected but none died. Within a span of six months of this SARS-CoV outbreak, it cost the world about $40 billion to contain it (see CDC-fact sheet-SARS(2004)[16]). On 5 Oct 2012, the National Select Agent Registry Program declared SARS coronavirus a **select agent**. A select agent is a toxin, virus, or bacterium which has the potential to become a severe threat to public health and safety (From https://www.cdc.gov/sars/index.html)

The SARS-CoV Genome consist of close to 30K bp of single stranded RNA sequence, significantly different from any previously known human coronaviruses. It turned out that SARS-Cov was a zoonotic (animal source infection) type of disease. Phylogenetic analysis revealed SARS-CoV originated from a *palm civet*, a small mammal, that is part of the diet in the Guangdong province in China. When it comes to SARS-Cov2, their RNA genome is sequence is tens of thousands letters long, where each letter represents one of the bases, and for computational convenience the uracil, is represented by the letter T. The RNA data used in our analysis, can be found at ani.stat.fsu.edu/∼vic/Data. Here is an example of a the initial portion of a datum sequence : hCoV-19/Belgium/GHB-03021/2020—EPI_ISL_407976
NNNNNNNNNNNNNNNNNNNNNNNNNAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTC
TCTAAACGAACTTTAAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTAT
ATTAATAACTAATTACTGTCGTTGACAGGACACGAGTAACTCGTCTATCTTCTGCAGGCTG..etc.

3 Stratification of Tree Spaces with m Leafs

A stratified space (space with a manifold stratification) is a metric space $\mathcal{M}$ that admits a filtration by closed subspaces, $\emptyset = F_{-1} \subseteq F_0 \subseteq F_1 \cdots \subseteq F_n \subset \cdots = M = \cup_i F_i$ such that for each $i = 1, \ldots, n, F_i \setminus F_{i-1}$ is empty or is an i-dimensional manifold, called the i-th stratum. For examples of stratified spaces encountered in Statistics, see Shen et al.(2021)[10]). For more details of spaces of trees with a fixed number of leafs, regarded at stratified spaces, we are also referring to Shen et al.(2021)[10].

Definition 1. *A tree with m leafs is a connected, simply connected graph, with a distinguished vertex, labeled o, called the root, and m vertices of degree 1, called leafs, that are labeled from 1 to p. In addition, we assume that with all interior edges have positive lengths. An edge of a p-tree is called interior if it is not*

connected to a leaf.

Tree spaces T_m, of trees with m leafs, were introduced in Billera et.al. (2001)[20], and, spaces of phylogenetic trees with a m leafs are our examples of tree spaces with m leafs. Note that phylogenetic trees are useful in the epidemiological reconstruction of the path of a viral infection. They trace histories of infectious bacteria and viruses to find the mutations they have undergone. This may give insight into how they appeared over time, and consequently, lead to effective drug development for curing such infections. Other applications of phylogenetic trees are in assessment of DNA evidence in court cases, in tracing individual ancestry, as well as in conservation of biology of rare species, by measuring and/or displaying the biodiversity in any ecosystem. Therefore phylogenetic trees can also help shaping conservation policy leading to decisions necessary to prevent specie's extinction. Even the Human species, currently under unprecedented Covid-19 attacks, is one of the species that may become endangered, in absence of tough policies for stopping the current pandemic, including by enforcing policies that would limit dangerous developments of Biotechnology.
The key procedure in obtaining phylogenetic tree (PT) data is the alignment of RNA sequences, as each RNA sequence is regarded a leaf of a rooted tree. Alignment of a sample of size nm of RNA sequences is obtained by segmenting them into identical subsequences (residues), and representing them as nm rows within a huge thin matrix. Gaps are inserted between the residues, so that identical or similar characters are aligned in successive columns. Clustal Omega is a multiple sequence alignment program that uses seeded guide trees and hidden Markov models profile-profile techniques to generate alignments between three sequences or more, up to 4000 sequences, assuming the total data size is up to 4MB (see [**?**]). Once the data was aligned, one may subtract a sample of size n of m sequences that are used to build build trees with $m-1$ leafs, and a known root, an individual RNA that temporarily precedes the leafs RNA's. Note that given a sample of RNA sequences, the smaller m, the larger the data set on the space of rooted trees. Branches are meant to suggest genetic information transmission path from one ancestor to the next, and longer branch lengths implies more genetic changes occurred; see Figure1

In general, there are two homologous interpretations that can be inferred when sequence data is compared for phylogenetic analyses:

- A smaller number of changes (shorter branches) reflects more relatedness.
- A bigger number of changes (longer branches) reflects less relatedness.

In Shen et al 92021)[10] one considers a sample of size 72 of SARS-Cov2 RNAs (also known as hCoV-19), that were present in Covid infected patients in various areas of the World. They built a rooted tree based on the aligned sequences, that is represented in Figure 2 below.

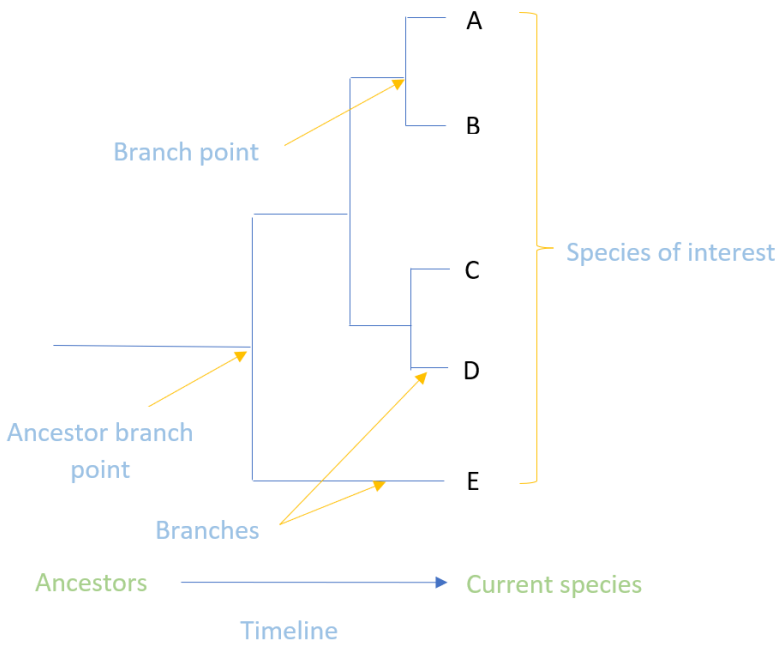

Figure 1: How species are displayed as leafs on a phylogenetic tree

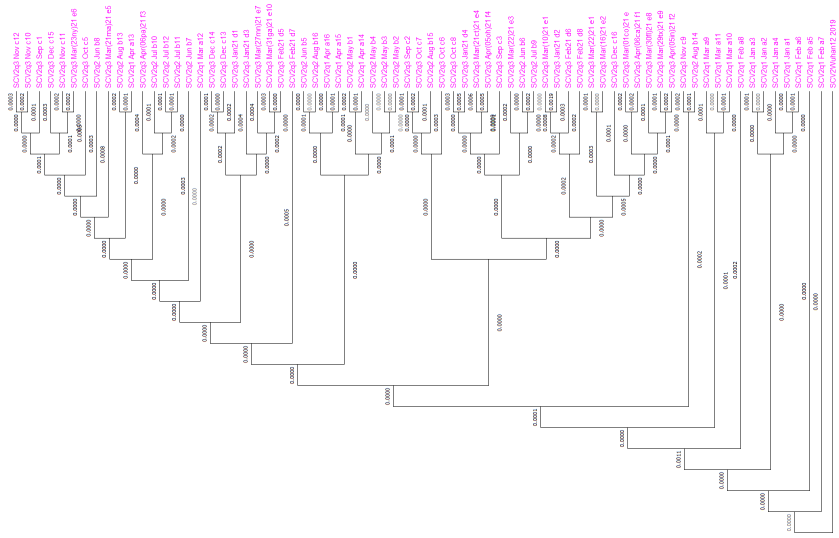

Figure 2: PT built from 72 RNA SARS-Cov2 sequences

Note that one tree with a large number of leafs, is however not helping the statistical tree analysis , since the sample size is one. Therefore given that this data set is fairly small, it is preferable to use rooted PTs with three leafs. The space T_3 is a spider with three legs (see Figure 3 from Billera et al(op.cit.))

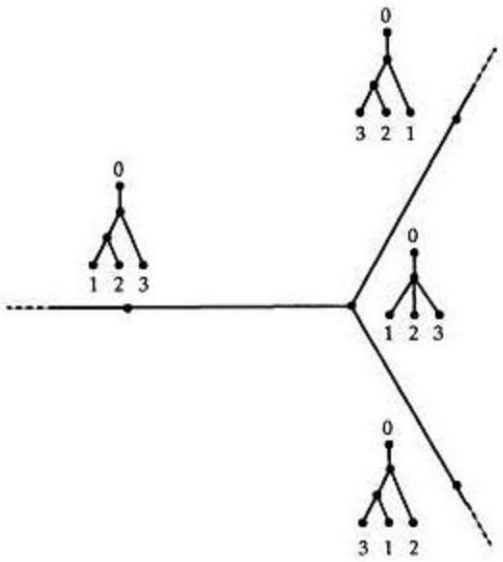

Figure 3: Tree space T_3

Thus our leafs data consists in copies of the virus the hCoV-19 based 3-leaf trees. Note that the 3-spider representation of T_3 is for rooted PTs with three leafs and an unknown root. However for an understanding of the origin of SARS-Cov2, a more reasonable analysis consists in considering trees with a **known candidate RNA**, therefore the we define the *evolutionary distance* from the ancestor (known tree root), to a current species, as sum of the inner edge distance and the distance from the known root to the first branching point of the PT. Therefore, the data considered in this paper, sits on the object space $M_3 = \mathbb{R} \times T_3$, which can be identified with an open book with three leafs, the first coordinate, being the distance from the root to the first branching point.

4 From RNA Sequence Data to Numerical Data via MEGA

In 2021 a new software, Molecular Evolutionary Genetic Analysis (MEGA) was made publicly available for PT analysis (see [18]). The 72 RNA sequences mentioned in the previous section, were aligned in MEGA. A portion of the aligned sequences is displayed in Figure 4. One mutation can be seen in line 16 of that figure. In naming the RNA sequences, in that figure, the if a sequence was collected in the year 2020, that year is not specified. However if the the year of RNA collection was 2021, that year is specified, to avoid confusion.

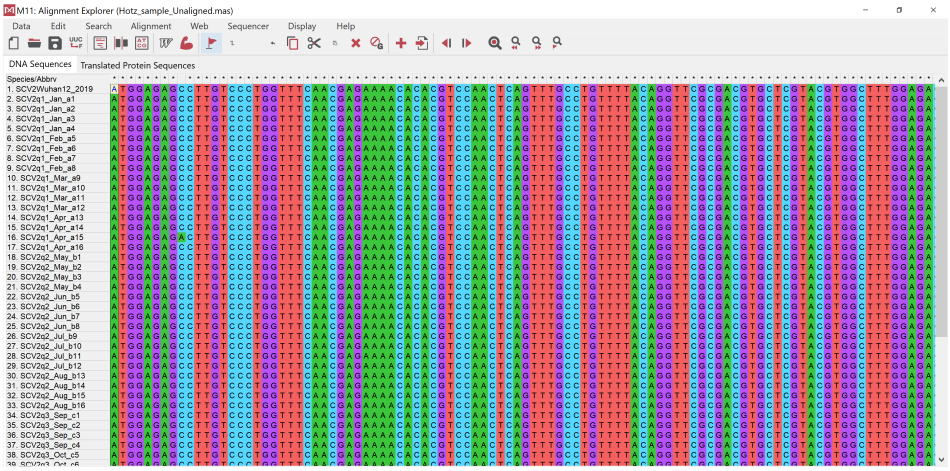

Figure 4: Sample of all RNA sequences (29K bp each)

The next step consisted in building 3-Leafs PT's in MEGA, with a known root, that could be either a Wuhan-SARS-Cov2 RNA, from the beginning of the pandemic in the Hubei region of China, and, for the same leafs, after alignment building trees with a root at a Bat-SARS RNA sequence. An example of tree building using alignment steps taken in MEGA is presented in Figure 5 and in Figure 6.

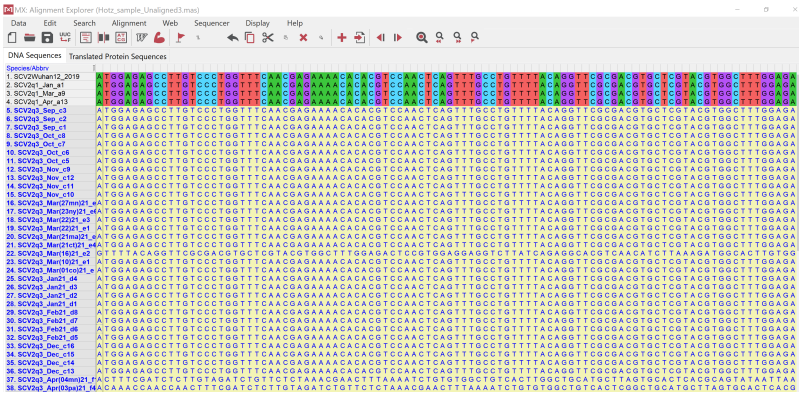

Figure 5: Selected 4 sequences from period 1

Figure 6: Isolated 4 sequences ready for alignment

Next in Figure 7, we display the corresponding rooted tree.

The next step consisted in grouping the three leafs of the rooted PT's data into two random samples of equal sizes. To insure independentce, the subset of leafs of the Wuhan-SARS-Cov2 rooted PT's was disjoint from the subset of leafs of the Bat-SARS rooted PT's. Also, the sample sizes were equal, thus $n_1 = n_2 = 12$. Finally the evolutionary distances were computed based on the branches and trunks distances for each of the resulting trees. These are are recoded in the table below. **Table.** Distances for the BAT-SARS rooted trees Wuhan-SARS-Cov2 rooted trees.

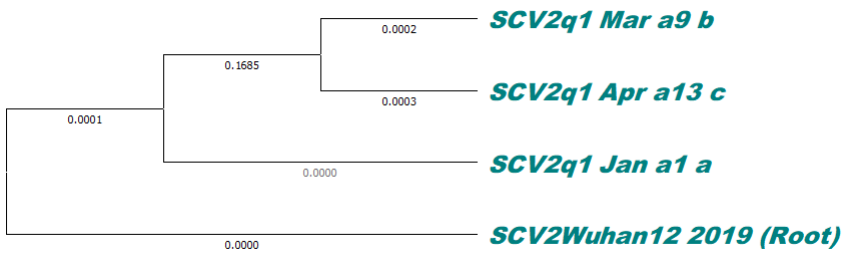

Figure 7: 3-leafs phylogenetic tree example to generate T_3 space data

Bat-SARS root	Inner edge Bat Sars	Wuhan-SARS root	Inner edge Wuhan-SARS
0.4857	0.0002	0.0000	0.0000
0.3225	0.0028	0.0000	0.0000
0.2923	0.0027	0.0000	0.1693
0.2898	0.0001	.1687	0.0001
0.4856	0.0002	.1688	0.0000
0.4856	0.0001	.1688	0.0000
0.4855	0.0003	0.1692	0.0002
0.4853	0.0009	0.1699	0.0014
0.4860	0.0005	0.1679	0.0003
0.4844	0.0023	0.1684	0.0005
0.2917	0.0047	0.0000	0.0000
0.2882	0.0011	0.0000	0.0000

5 Investigating the Origin of SARS-CoV-2 RNA Sequences Using Evolutionary Distances on Tree Spaces

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causes the infectious coronavirus disease 2019 (Covid-19), which is a vascular disease with pulmonary pneumonia. The main goal is to use stratified space analysis with three-leafs phylogenetic trees to find out if the RNA sequence of this virus has undergone some significant mutations since the onset of the pandemic in US. This investigation is done comparing distributions of evolutionary distances on T_3, or on its proxy, the 3-Spider, distances from leafs representing individual subsequent SARS-Cov-2 RNA sequences to the root of the tree. Given that the two hypotheses are **1. Covid jumped from bats to humans, via an intermediate animal, vs 2. Covid19 was leaked from the Wuhan Institute of Virology (WIV)**, and due to lack of cooperation of WIV authorities, the alternative hypothesis was that Covid19 appeared in Wuhan, tus, trees are being built from 3 leaf trees having as root either a Bat-SARS-Cov, or a Wuhan-SARS-Cov-2 sequence. Note that in an earlier paper, Moore (2021)[4] collected a representative sample, and used phylogenetic trees with only three leafs to compare SARS-CoV-2 RNA sequences samples collected within three different time periods; the first and second halves of 2020 (periods 1 and 2), and then the first four months of 2021 (period 3). He then applied the stickiness theorem in Hotz et al(2014)[19], results that will be published in a subsequent paper.

Fir the evolutionary distance based SARS-Cov2 origin hypothesis testing, we note that the PT SARS-Cov2 on T_3, is in fact a univariate testing problem, since in the 3-spider representation, S_3 is a 1D stratified space. The SARS-Cov2 hypothesis testing $H_0 : Batvs.H1 : Wuhan$, can be therefore formulated in terms of the mean evolutionary distances to the origin, as follows: Let X_B be the distance from a point representing a tree rooted at a BAT-SARS sequence to the center of the spider S_3 and X_W be the distance from a point representing a tree rooted at a BAT-SARS sequence to the center of the spider S_3.

- Accept the bat origin if $H_0 : \mu_{X_B} \leq \mu_{X_W}$ is accepted with the alternative $H_0 : \mu_{X_B} > \mu_{X_W}$.
- Accept the Wuhan origin if $H_0 : \mu_{X_B} \geq \mu_{X_W}$ is accepted with the alternative $H_0 : \mu_{X_B} < \mu_{X_W}$.

Based on the data on hand, in the above table (add data in columns 1 and 2, respectively columns 3 and 4). Based on the data in table 4, the Bat-SARS hypothesis is strongly rejected, using nonparametric bootstrap. Indeed, the studentized sample mean difference 95% nonparametric bootstrap confidence interval, using 97.5% and 2.5% percentiles cutoffs is $(6.767418, 11.372894)$. We used $N = 10000$ resamples with repetition.

6 Concluding remarks

It is well documented that bats are not a common fixture in Wuhan, China. On the other hand, it is also well known that the report at the World Health Organization, was done by a team that included Peter Daszak (EcoHealth Alliance, USA) who had a conflict of interest, being coauthor with the WIV virologist Zheng-Li Shi[21]. Therefore, the report conclusion is flawed. In this short data driven computational paper, we concluded that SARS-Cov2 is most likely originated in Wuhan, China, and there is a high suspicion, given the two year plus long history of the Covid 19 pandemic, that, unfortunately, SARS-Cov2, has a too long evolutionary distance from the Bat-SARS RNA rooted trees, compared with the Wuhan-SARS-Cov2 RNA rooted trees, thus, in the absence of bats in Wuhan, other than those in WIV, and more importantly, absent an intermediate animal career of SARS-Cov2, we accept the WIV Lab made origin of SARS-Cov2.
Our analysis, based on the evolutionary distance, between older and recent similar coronaviruses, also validates the fact that it should take about half century for the BAT-SARS RNA to evolve into the SARS-Cov2. Gain of function for viruses should be banned.

References

[1] Paul Adepoju (2020). Moving on from the failed HIV vaccine clinical trial. *The Lancet*, **Vol 7,** Nr 3, e149-e214.

[2] Campbell, J.(2021). https://www.youtube.com/c/Campbellteaching/community

[3] Moderna (2021) https://www.modernatx.com/modernas-mrna-technology.

[4] Moore, R.(2021). *Investigating Significant Mutations of US SARS-CoV-2 RNA Sequences Using Stratified Spaces, And Genetic Connection to Drug-Induced Liver Injury (DILI) Through Statistical Learning Methods*. PhD Dissertation, Florida State University.

[5] R. Copin,A. Baum,E. Wloga,K.E. Pascal,S. Giordano, B. O. Fulton, A. Zhou, N. Negron, K. Lanza, N. Chan, A. Coppola, J. Chiu, M. Ni, Y. Wei, G. S. Atwal, A. Romero Hernandez, K. Saotome, Y. Zhou, M. C. Franklin, A. T. Hooper, S. McCarthy, S. Hamon, J. D. Hamilton, H. M. Staples, K. Alfson, R. Carrion Jr.,2 Shazia Ali,1 Thomas Norton,1 Selin Somersan-Karakaya,1 Sumathi Sivapalasingam,1 Gary A. Herman, D. M. Weinreich, L. Lipsich, N. Stahl, A. J. Murphy, G. D. Yancopoulos and Ch. A. Kyratsous (2021). The monoclonal antibody combination REGEN-COV protects against SARS-CoV-2 mutational escape in preclinical and human studies. *Cell*, **184**, 3949—3961.

[6] Anne Glover (2012). *The 21st Century: The Age of Biology*. OECD Forum on Global Biotechnology, Paris 12 November 2012. https://www.oecd.org/sti/emerging-tech/A Glover.pdf

[7] Everett Clinton Smith, Mark R Denison(2013). Coronaviruses as DNA wannabes: a new model for the regulation of RNA virus replication fidelity. *PLoS PLOS Pathogens* **9**(12): e1003760. https://doi.org/10.1371/journal.ppat.1003760

[8] Gromov, M.(2018). *Mikhail Gromov - Mathematics behind massive artificial evolution/selection processes* Talk at IHES, Paris, France. https://www.youtube.com/watch?v=g4Wl3Ggho6k

[9] Mendel, G.(1865). Versuche $\ddot{u}$ber Pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr, 1865, Abhandlungen*: 3—47

[10] C. Shen, V. Patrangenaru and R. Moore(2021). A Phylogenetic Trees Analysis of SARS-CoV-2. *arXiv:2106.06918*.

[11] Linus Pauling and Robert B. Corey(1953). A Proposed Structure For The Nucleic Acids. *Proc Natl Acad Sci U S A*. 1953 Feb; 39(2): 84—97

[12] Padron-Regalado, E. (2020, April 23). *Vaccines for SARS-CoV-2: Lessons from Other Coronavirus Strains*. Infectious diseases and therapy. https://pubmed.ncbi.nlm.nih.gov/32328406/.

[13] https://en.wikipedia.org/wiki/Rosalind_Franklin

[14] WHO (2020, March 26). *Origin of SARS-CoV-2*. https://www.who.int/health-topics/coronavirus/origins-of-the-virus

[15] https://en.wikipedia.org/wiki/Investigations$_into_the_origin_of$_COVID-19

[16] CDC Fact Sheet: Basic Information about SARS(2004). https://www.cdc.gov/sars/about/fs-SARS.pdf

[17] Cristianini, N., & Hahn, M. W. (2006). Introduction to computational genomics: a case studies approach. *Cambridge University Press*.

[18] Sudhir Kumar, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* 35:1547-1549.

[19] Thomas Hotz, Stephan Huckemann, Huiling Le, James S. Marron, Jonathan C. Mattingly, Ezra Miller, James Nolen, Megan Owen, Vic Patrangenaru and Sean Skwerer (2013). Sticky Central Limit Theorems on Open Books. *Annals of Applied Probability,* **23**, 2238–2258.

[20] Billera, Louis J.; Holmes, Susan P.; Vogtmann, Karen(2001). Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27**, no. 4, 733–767.

[21] P. Zhou, H. Fan, T. Lan, X.-L. Yang, W.-F. Shi, W. Zhang, Y. Zhu, Y.-W. Zhang, Q.-M. Xie, S. Mani, X.-S. Zheng, B. Li, J.-M. Li, H. Guo, G.-Q. Pei, X.-P. An, J.-W. Chen, L. Zhou, K.-J. Mai, Z.-X. Wu, D. Li, D. E. Anderson, L.-B. Zhang, S.-Y. Li, Z.-Q. Mi, T.-T. He, F. Cong, P.-J. Guo, R. Huang, Y. Luo, X.-L. Liu, J. Chen, Y. Huang, Q. Sun, X.-L.-L. Zhang, Y.-Y. Wang, S.-Z. Xing, Y.-S. Chen, Y. Sun, J. Li, P. Daszak, L.-F. Wang, Z.-L. Shi, Y.-G. Tong and J.-Y. Ma (2018). Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. *Nature.* **556**, 255—258 .