

Response to “On the origin and continuing evolution of SARS-CoV-2”

n0rr

Mar 6

In our phylogenetic analyses, "L type" likely formed a monophyletic derived clade.
"L type" in Tang et al. is same to "Group CTC" in our phylogenetic tree.

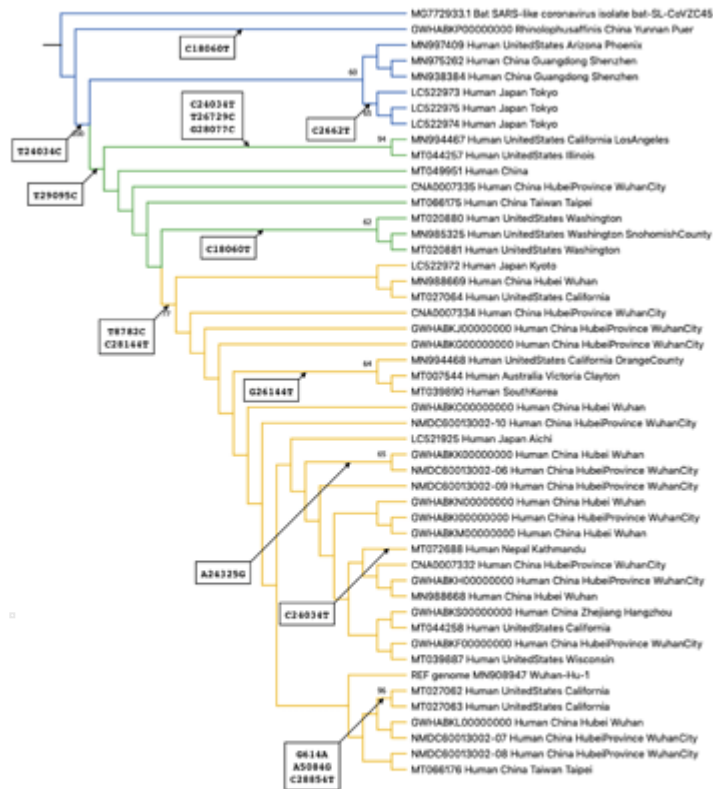


Figure 2. Topological phylogeny of SARS-CoV-2 estimated with maximum-likelihood

A maximum-likelihood phylogenetic analysis based on the multiple sequence alignment of ORF were performed under the condition of separating each codon position. Blue, green, and yellow branches indicate group TCT, TCC, and CTC respectively. Substitutions on the parsimony-informative sites were mapped to the corresponding branches. SARS-CoV-2 isolated from human formed a monophyletic clade with 100% bootstrap value. Group CTC formed a monophyletic clade with 74% bootstrap value. Group TCC was paraphyletic respect to the group CTC. Group TCT was paraphyletic respect to the clade consists of group CTC and TCT. Final log-likelihood was -55794.362675. AIC score was 112649.332493. Branch length was meaningless.

Phylogenetic analyses of the severe acute respiratory syndrome coronavirus 2...

Worldwide Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection is disrupting in the economy and anxiety of people. The public anxiety has increased the psychological burden on government and healthcare professionals, resulting in a...

luj

Mar 7

Response to MacLean et al.'s 'Response to "On the origin and continuing evolution of SARS-CoV-2"'

Jian Lu

State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing, 100871, China

To whom correspondence should be addressed: LUJ@pku.edu.cn

The criticisms by MacLean et al. (Response to "On the origin and continuing evolution of SARS-CoV-2") of the Tang et al.'s recent publication ("On the origin and continuing evolution of SARS-CoV-2")

on National Science Review (NSR) will be briefly answered below. Both parties have agreed that full-length exchanges should appear in NSR, which has generously offered to host this very public and open debate.

There are two main criticisms by MacLean et al. First, they argued that S and L types have no functional significance since we did not measure within-host infections. To evolutionary geneticists like ourselves, the "functional" test is the frequency of each mutation in the population. If a mutation has a fitness (or transmission) advantage, it will be much more common than the neutral prediction. Here, we use the synonymous mutations that do not alter protein sequences as the neutral reference.

As shown in Fig. 2 of Tang et al., there is only one nonsynonymous mutation that appears in > 50% of the samples and that is the L mutation. In contrast, there are 7 synonymous mutations. In the complete absence of selection, one expects to see a nonsynonymous : synonymous ratio of $\sim 3.5 : 1$. In short, while we expect to see $7 \times 3.5 = 24.5$ nonsynonymous mutations in the population, we see only one, which is the L mutation. It is not clear why MacLean et al. failed to see this simple and central point.

We may then flip the coin and observe the ancestral S variant. It occurs in only one of the 27 samples (3.7%) collected before Jan. 7 (mainly from Wuhan). In contrast, it occurs in 28 of the 73 samples (38.4%) after Jan. 7 (mainly from outside of Wuhan). Again, the difference is startling and highly significant by any statistical test. MacLean et al. claimed that the two patterns can be explained by pure stochastic forces, such as population bottleneck. These finer points will be debated in NSR. We only wish to add one point. Stochastic changes are pronounced only when the population size is small. Even if each infected patient has only ONE viral particle, the effective population size of the virus is not small.

The second criticism of MacLean et al. is that there is no purifying (i.e., negative) selection against non-synonymous mutations in the circulating viral population. We shall again refer to the ratio of mutations at > 50% in frequency. Recall that the expected nonsynonymous: synonymous mutations should be $24.5 : 7$ but we observed only $1 : 7$. If this 24.5 fold difference is not due to negative selection, what is it then?

This 24.5 fold difference stands in sharp contrast with the test that MacLean et al. used to argue against purifying (their Table 1). The technical errors will be discussed in detail in the formal debate. In particular, because deleterious mutations occur in low frequencies, often as singletons (i.e., one occurrence in all samples), the study of purifying selection has to filter out deleterious mutations that have not been purged yet. Apparently, these authors fail to take into account the fundamental rule of population genetics in their analyses.

As the editor of NSR states in public, the proper place to have a thoughtful debate is on NSR, which will invite experts to review the submissions. Our public statement on this matter prior to the account published by NSR will be limited to the messages here.

While the debate above is important, its fairly academic. What is missing is discussion about the use of "aggressive", which is not a standard epidemiological term. Any work on the subject of COVID-19 is going to draw attention, so we all need to be very mindful about the public health messaging. The way that I assume that you are using "aggressive" is to imply transmission rate or fitness, as you are only looking at the frequencies of the lineages. Many in the public, however, see "aggressive" and think severity or virulence. It's this part that is drawing the most attention to your paper and insinuating unnecessary fear. The press loves this as articles about fear generate more clicks. Now there are common threads all over twitter that suggest if you are infected with the "L" strain, you will be more likely to have severe disease and die. We now have to spend considerable energy undoing this misinformation. Much of the damage, however, is already done. Please think about the messaging around future papers to avoid confusion and unnecessary panic.

Please also address this in your response. It will be very important to have clarification from the authors.

Sorry for the tone, I do not intend any direct offense to you or your co-authors. This is a learning experience for all of us.

Best,

Nate

luj

Mar 8

@ [grubaughlab](#) Thanks. The advice is well taken.

fuyutao

Mar 10

The genomic sequence hCoV-19/USA/IL1/2020 was interpreted in the NSR paper as a patient carrying a mixture of L and S types. The patient later infected a family member. The virus isolated from the family member is hCoV-19/USA/IL2/2020, which is the S type. This single case doesn't support S type being "less aggressive".

luj

Mar 10

@ [fuyutao](#)

Thank you for providing the information about hCoV-19/USA/IL2/2020. We did not talk about his patient in our paper. Would you please tell me in which part of our paper we showed this as evidence to support "S type being less aggressive" ?

Thanks,
Jian

oscar.maclean

Mar 10

Response to Lu's response to MacLean et al.'s Response to "On the origin and continuing evolution of SARS-CoV-2"

Oscar A. MacLean*, Richard Orton, Joshua B. Singer, David L. Robertson.

MRC-University of Glasgow Centre for Virus Research (CVR).

*To whom correspondence should be addressed: Oscar.MacLean@glasgow.ac.uk.

As Jian Lu on behalf of Tang et al. has chosen to post their response here, we see no reason to replicate our critique in the NSR journal. Concerning their response, respectfully to these authors, they are missing our key issues with their manuscript:

As we pointed out in our initial response, there are many epidemiological factors which may drive changes in frequency of a mutation in a global viral outbreak which involve no natural selection. The culmination of these processes drives mutation frequency changes, these changes are fully expected. It is thus meaningless to state as the authors have that "the "functional" test is the frequency of each mutation in the population".

The claim that a difference in S and L frequencies across time would be "significant by any measure" is the key issue here. For example, a simple Fisher's exact test (which was used in Tang et al's original paper) measures the probability of seeing an observed difference in frequencies across countries or time, if the underlying frequencies were identical. This test does not assess if those mutation frequency changes observed across the time period are driven by selection or a null model under genetic drift. The question that is being addressed is not "is there a true difference in frequencies" (stochastic processes will cause changes in the frequency of mutations, which will likely drive significant Fisher's exact tests), rather it's whether this change in frequency over time is distinct from that under a neutral null evolutionary model. The former and the latter questions are very different, with greatly differing wider significance. We would very much like to see an analysis answering the latter question be performed, but would be immensely surprised if it had the power to detect any signatures of selection given the current data.

Any analysis of allele frequencies also needs to consider another important factor that we neglected to mention in our initial reply: biased sampling. Each sample that is sequenced is not independent from another. For example, as contact tracing is a significant driver of case detection, there will be a correlation between samples detected, driving oversampling of particular genotypes and mutations. Additionally, the sampling of infections for sequencing is greatly biased by the country they occur in, for example, 80% of COVID-19 cases to date (9/3/2020) come from China, but a far smaller proportion (~40% as of 2/3/2020) of sequenced genomes do. This biased sampling will further exaggerate these epidemiological variations driving variations in observed mutation frequencies.

Lu has also restated their claim that "there is only one nonsynonymous mutation that appears in > 50% of the samples and that is the L mutation. In contrast, there are 7 synonymous mutations" without any further qualification, completely ignoring our explanation in our first posting of why their methodology is flawed and why there are not seven such mutations.

Some additional points:

We did not make the claim "...that there is no purifying (i.e., negative) selection against non-synonymous mutations in the circulating viral population." We simply highlighted a lack of power to detect it with the limited data currently available. The authors are over-interpreting our results here.

Whilst the paper by Charlesworth and Eyre-Walker (2008), which suggests removing variants under 5% frequency in count based analysis of polymorphisms, is fairly well cited (158 as of 9/3/2020), we think describing this data processing step as "the fundamental rule of population genetics" is unwarranted. However, if this frequency threshold were to be used, only three nonsynonymous and two synonymous mutations would remain for analysis (using data from 2/3/2020). This further highlights the lack of power in the current SARS-CoV-2 polymorphism data to make inferences about the pattern of purifying selection.

Lu provides a 3.5:1 nonsynonymous to synonymous ratio with no explanation of the methodology that has been used to generate it, or why it differs from our 2.43:1 estimate. Counting the relative numbers of sites for a given gene/genome is a non-trivial process, and the choice of model will be a significant driver in the estimated ratios.

To reiterate our first posting, Tang et al. have not provided any evidence that there are two major types of SARS-CoV-2, and certainly no evidence that part of the outbreak has been "more aggressive". We agree with Nathan Grubaugh's post that such a claim is misleading and has led to the spread of misinformation in the press. Rather than focus on our response, the authors should urgently correct the confusion they are responsible for.

Reference

Charlesworth, Jane, and Adam Eyre-Walker. "The McDonald–Kreitman test and slightly deleterious mutations." *Molecular biology and evolution* 25.6 (2008): 1007-1015.

[richard.orton](#)

Mar 11

It was shown experimentally during the 2013-2016 outbreak of Ebola virus in West Africa that some of the non-synonymous mutations in the glycoprotein increased infectivity in human cells ([virology.ws](#), [Diehl et al 2016](#), [Urbanowicz et al., 2016](#)). In the case of SARS-CoV-2 here, it seems a little premature to be labelling mutations "aggressive" just based on frequencies and without any experimental validation yet, as many epidemiological factors can have a major influence, as well as sampling issues. The S and L "types" relate to a C/T mutation at 28,144 resulting in the amino acid serine (S) or leucine (L) at this site. Both these mutations are still circulating in the epidemic, and as the virus is naturally evolving (and passing through various bottlenecks), this leads to further distinct clusters of viral samples as new mutations are acquired and passed on. For example, within the L cluster, a subcluster containing non-synonymous mutations at genome positions 14,408 and 23,403 and a synonymous mutation at 3,037 is now apparent, which consistently appear together on the viral genome; a further subcluster additionally containing mutations at 3 neighbouring positions is also apparent within this. Whilst the S cluster has a subcluster containing non-synonymous mutations at genome positions 17,747 and 17,858 along with a nonsynonymous mutation at 18,060 (many of the recent sequences from the USA are within this cluster). All these clusters (or "types") are readily apparent from the phylogenetic trees, but it is not clear if any of the mutations give the virus a fitness advantage. Only experiments can truly determine the role of any of these mutations

and whether they affect transmissibility/infectivity etc, or whether their accumulation is just due to random stochastic/bottleneck effects. Both the S and L viruses are still circulating, and there are a lot of other mutations accumulating across the viral genome.



Figure. Heatmap visualisation of all synonymous (yellow) and non-synonymous (red) nucleotide mutations from all complete human SARS-CoV-2 genome sequences available from [GISAID](#) on 11th March 2020. Created in R using d3heatmap (the hclust algorithm for grouping related samples is a little crude). Columns represent individual samples, rows represent genome positions. The S and L clusters are labelled with blue and purple boxes respectively, and genome positions mentioned in the text are labelled, numbering relates to the genome positions in the GenBank sequence [MN908947](#) . We would like to thank all the authors who have kindly deposited and shared genome data on [GISAID](#) ; a table with genome sequence acknowledgments can be found on the [CoV-GLUE website](#) .

luj

Mar 13

[@oscar.maclean](#) oscar.maclean](<http://virological.org/u/oscar.maclean>)

An update of Response to MacLean et al.'s 'Response to "On the origin and continuing evolution of SARS-CoV-2"'

Jian Lu

State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing, 100871, China

To whom correspondence should be addressed: LUJ@pku.edu.cn

Since MacLean et al. saw "no reason to replicate their critique in the NSR journal", on behalf of all authors, I update our response here.

The first debated issue is about the two types (S and L) we defined to describe the circulating virus population. Our definition was based on two loci that are ~20 kb apart in the ~30 kb viral genome, each has a minor allele frequency of ~30%, and show nearly complete linkage ($r^2 = 0.95$). Hence, it is a simple FACT that there are two major clones or types circulating in the population. As for the functional interpretation, the frequency of S type changed from 3.8% to 37.8% between samples collected before and after Jan. 7 2020. We believed the nearly 10-fold difference, supported by $P = 0.0008$, was worth reporting to the community. Nevertheless, as we stated in Tang et al., we were fully aware that the finding might be biased by the limited samples available at that moment. We note that it is important for scientists to find a balance between being mute and reporting the observations during such a pandemic. We encourage people to keep an eye on the potential connection between virus types and clinical symptoms. Meanwhile, we recognize that within the context of Tang et al. the term "aggressive" is misleading and should be replaced by a more neutral term "a higher frequency". We have made an addendum to <https://doi.org/10.1093/nsr/nwaa036> to clarify this.

The second debated issue is about the detection of purifying selection. MacLean et al. argued that they did not detect purifying selection acting on the nonsynonymous mutations ($P = 0.113$, Table 1). To do that, MacLean et al. estimated the relative number of sites using the Goldman and Yang (1994) codon model. They concluded that "It estimates there are 2.43 times more nonsynonymous than synonymous sites in the SARS-CoV2 genome." Based on their calculation, the ratio of nonsynonymous (N) to synonymous (S) sites should be 3.43 : 1. We obtained very similar ratios (roughly 22,789 N to 6,443 S sites, which is roughly 3.5: 1) whenever we used the parameter settings MacLean et al. did or various other parameters.

MacLean et al. used an expected N:S site ratio of 2.43:1 to test if the observed 169 mutations (105 N + 64 S) have been shaped by purifying selection. They failed to detect purifying selection because of a non-significant P value (= 0.113) they obtained by using a chi-square test. This is simply WRONG.

First, MacLean et al. used a ratio of 2.43: 1 to calculate the expected number of N and S mutations, which gave them a wrong conclusion ($P = 0.113$, as shown in Table 1 of MacLean et al.). When the ratio of 3.43: 1 is used in the calculation, the P value should be 0.003, if we used the chi-square test as MacLean et al. did.

Second, the chi-squared test MacLean et al. used is not appropriate. They should contrast the numbers of mutated N and S sites (105 and 64, respectively) to the numbers of non-mutated N sites and S sites in the virus genome, which are 22,684 and 6,379, respectively, according to our calculation. These numbers are consistent with MacLean et al.'s claim that "there are 2.43 times more nonsynonymous than synonymous sites in the SARS-CoV2 genome." Then, the P-value should be 1.04×10^{-6} .

Hence, both ways would overturn MacLean et al.'s conclusion. MacLean et al. failed to detect the signature of purifying selection based on erroneous calculations.

Moreover, consistent with our results, purifying selection on SARS-Cov-2 was also detected in a recent independent study (<http://doi.org/10.1093/cid/ciaa203>).

We are pleased to see the addendum added to the Tang et al. manuscript in National Science Review, which we have copied below for reference. However we note that the online abstract, which will be by far the most read part of the paper, is unchanged in the strength of its claims "On the other hand, the S type, which is evolutionarily older and less aggressive".

On the specific methodological comment on our purifying selection analysis:

To investigate the origin of the estimates of nonsynonymous to synonymous site ratios given by Tang et al. above, we ran the PAML software they cited (Yang 2007) using two different models. These two models differ in their estimation of codon frequencies, the 1x4 model uses average base frequencies across sites, and the codon frequency model uses the observed counts of each codon in the alignment. The estimates of ratios of nonsynonymous to synonymous sites these models produced ranged from 2.76 to 3.75. All models for these count estimates are fundamentally wrong, as they are approximations using limited data. Even the gold standard for generating this data, mutation accumulation experiments, are limited by the difficulty in observing lethal mutations. However, given that PAML uses a more powerful maximum likelihood framework and has been cited over 5,000 times (as of 16/3/20), we will happily use these estimated ratios, rather than our own (2.43). Both of these ratios from PAML would produce a significant chi squared test on our count data in Table 1 of the original post $P < 0.036$.

We thank the authors for bringing our attention to this. We therefore agree it is a fair conclusion that significant evidence of purifying selection can be observed, which is filtering out nonsynonymous mutations before they can be observed in the outbreak. We would note that this is a subtly different result from that in the original Tang et al. paper, which suggested evidence of purifying selection suppressing the frequency of the observed mutations in the outbreak. Our criticism of that analysis remains unchanged.

Tang et al's addendum for reference:

"In our recent publication, we showed that among circulating SARS-CoV-2 (with 103 genomes analyzed) two different viral genomes co-exist. We identified them as lineages L and S. The concerned amino acid we used to define the L and S lineages is located in ORF8 (open reading frame 8), which plays a yet undefined role in the viral life cycle. Based on the finding that "L" lineage has a higher frequency than lineage S, we described the L lineage as aggressive. We now recognize that within the context of our study the term "aggressive" is misleading and should be replaced by a more precise term "a higher frequency". In short, while we have shown that the two lineages naturally co-exist, we provided no evidence supporting any epidemiological conclusion regarding the virulence or pathogenicity of SARS-CoV-2. By saying so, corrections will be made in the print version of this paper to avoid being misleading."

Reference

Yang, Ziheng. "PAML 4: phylogenetic analysis by maximum likelihood." *Molecular biology and evolution* 24.8 (2007): 1586-1591.

@oscar.maclean

A quick comment is that MacLean et al. still did not perform the statistical test properly. When they tested whether the observed N/S number (105/ 64) departures from the expected N/S ratio (2.76 to 3.75, according to their updated analysis), they should use a binomial test, rather than the Chi-squared test as presented in their Table 1. Using the binomial test, we get $P = 0.0009$ under the expected N/S ratio of 2.76, and $P = 4.41e-07$ under the expected N/S ratio of 3.75.

Alternatively, if MacLean et al. insisted on using the chi-squared test, the most appropriate approach is to compare the number of mutated N and S sites to the number of non-mutated N and S sites, as I showed previously.