nature research

Peer Review Information

Journal: Nature Microbiology
Manuscript Title: A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology
Corresponding author name(s): Dr Andrew Rambaut, Dr Edward (C) Holmes, Prof Oliver (George) Pybus

Reviewer Comments & Decisions:

Decision Letter, initial version:

Dear Eddie,

Thank you for your patience while your manuscript "A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology" was under peer-review at Nature Microbiology. It has now been seen by 3 referees, whose expertise and comments you will find at the of this email. You will see from their comments below that while they find your work of interest, some important points are raised. We are very interested in the possibility of publishing your study in Nature Microbiology, but would like to consider your response to these concerns in the form of a revised manuscript before we make a final decision on publication.

All three of our expert reviewers confirm that this is a timely proposal and provide guidance for improvements in revision. Reviewer #1 was unconvinced this was a good fit for Nature Microbiology, but we will overrule that concerns, which are not technical. Reviewer #1 queries the dynamic nature of the naming of lineages which s/he argues could cause confusion. Please clarify how this would work in your revised manuscript. In a similar vein, reviewer #2 asks how sequencing archival samples may alter the rooting or branches in the tree. Reviewer #3 also brings up a similar issue but this time mentions recombinant viruses.

1. Please ensure that the dynamic nature of the nomenclature is made crystal clear insofar as it can be by addressing concerns about the 'what ifs' for added samples and renamed lineages.

2. Reviewer #2 asks if you might use 'fixed date' releases. This seems an interesting idea as it might provide a 'history of the viral lineages' over time in an easy to access format. Your thoughts on this idea would be welcome. Related to this, and raised by the same reviewer, who will maintain and update the nomenclature over time? Where will it be stored?

3. Reviewer #2 suggests seeking involvement of the ICTV Coronaviridae Study Group. In principle this is a good suggestion going forwards for the field, but at the same time we are keen to progress your thoughts to the community at the earliest possible time as this proposal is timely and there may be other proposals in the works from other teams. Please let us know your thoughts on this as soon as practical. One idea we had was to contact them so that you can (perhaps) say the proposal is under

consideration (indeed you may already have done so) but not wait for full consideration should that be likely to take longer than seems ideal.

Please can you let me know when to expect a revised version, so that we can prioritize your article when it is resubmitted?

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions at http://www.nature.com/nmicrobiol/info/final-submission/

The usual length limit for a Nature Microbiology Article is six display items (figures or tables) and 3,000 words. We have some flexibility, and can allow a revised manuscript at 3,500 words, but please consider this a firm upper limit. There is a trade-off of ~250 words per display item, so if you need more space, you could move a Figure or Table to Supplementary Information.

Some reduction could be achieved by focusing any introductory material and moving it to the start of your opening 'bold' paragraph, whose function is to outline the background to your work, describe in a sentence your new observations, and explain your main conclusions. The discussion should also be limited. Methods should be described in a separate section following the discussion, we do not place a word limit on Methods.

Nature Microbiology titles should give a sense of the main new findings of a manuscript, and should not contain punctuation. Please keep in mind that we strongly discourage active verbs in titles, and that they should ideally fit within 90 characters each (including spaces).

We strongly support public availability of data. Please place the data used in your paper into a public data repository, if one exists, or alternatively, present the data as Source Data or Supplementary Information. If data can only be shared on request, please explain why in your Data Availability Statement, and also in the correspondence with your editor. For some data types, deposition in a public repository is mandatory - more information on our data deposition policies and available repositories can be found at https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data.

Please include a data availability statement as a separate section after Methods but before references, under the heading "Data Availability". This section should inform readers about the availability of the data used to support the conclusions of your study. This information includes accession codes to public repositories (data banks for protein, DNA or RNA sequences, microarray, proteomics data etc…), references to source data published alongside the paper, unique identifiers such as URLs to data repository entries, or data set DOIs, and any other statement about data availability. At a minimum, you should include the following statement: "The data that support the findings of this study are available from the corresponding author upon request", mentioning any restrictions on availability. If DOIs are provided, we also strongly encourage including these in the Reference list (authors, title, publisher (repository name), identifier, year). For more guidance on how to write this section please see:
http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf

To improve the accessibility of your paper to readers from other research areas, please pay particular attention to the wording of the paper's opening bold paragraph, which serves both as an introduction

and as a brief, non-technical summary in about 150 words. If, however, you require one or two extra sentences to explain your work clearly, please include them even if the paragraph is over-length as a result. The opening paragraph should not contain references. Because scientists from other sub-disciplines will be interested in your results and their implications, it is important to explain essential but specialised terms concisely. We suggest you show your summary paragraph to colleagues in other fields to uncover any problematic concepts.

If your paper is accepted for publication, we will edit your display items electronically so they conform to our house style and will reproduce clearly in print. If necessary, we will re-size figures to fit single or double column width. If your figures contain several parts, the parts should form a neat rectangle when assembled. Choosing the right electronic format at this stage will speed up the processing of your paper and give the best possible results in print. We would like the figures to be supplied as vector files - EPS, PDF, AI or postscript (PS) file formats (not raster or bitmap files), preferably generated with vector-graphics software (Adobe Illustrator for example). Please try to ensure that all figures are non-flattened and fully editable. All images should be at least 300 dpi resolution (when figures are scaled to approximately the size that they are to be printed at) and in RGB colour format. Please do not submit Jpeg or flattened TIFF files. Please see also 'Guidelines for Electronic Submission of Figures' at the end of this letter for further detail.

Figure legends must provide a brief description of the figure and the symbols used, within 350 words, including definitions of any error bars employed in the figures.

Please include a statement before the acknowledgements naming the author to whom correspondence and requests for materials should be addressed.

Finally, we require authors to include a statement of their individual contributions to the paper -- such as experimental work, project planning, data analysis, etc. -- immediately after the acknowledgements. The statement should be short, and refer to authors by their initials. For details please see the Authorship section of our joint Editorial policies at http://www.nature.com/authors/editorial_policies/authorship.html

When revising your paper:

* include a point-by-point response to any editorial suggestions and to our referees. Please include your response to the editorial suggestions in your cover letter, and please upload your response to the referees as a separate document.

* ensure it complies with our format requirements for Letters as set out in our guide to authors at www.nature.com/nmicrobiol/info/gta/

* state in a cover note the length of the text, methods and legends; the number of references; number and estimated final size of figures and tables

* resubmit electronically if possible using the link below to access your home page:

[REDACTED]

*This url links to your confidential homepage and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this e-mail to co-authors, please delete this link to your homepage first.

Please ensure that all correspondence is marked with your Nature Microbiology reference number in the subject line.

Nature Microbiology is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. This applies to primary research papers only. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

We hope to receive your revised paper within three weeks. If you cannot send it within this time, please let us know.

We look forward to hearing from you soon.

Yours sincerely,
Paula

****

Reviewer Expertise:

Referee #1: Epidemiology
Referee #2: Taxonomy
Referee #3: Viral phylodinamics

Reviewers Comments:

Reviewer #1 (Remarks to the Author):

As the SARS-CoV-2 pandemic has spread, thousands of virus genome sequences have been generated. However, there is no coherent or accepted system for nomenclature. Such a system should be flexible and dynamic to reflect the birth and death of viral lineages. The authors have proposed a nomenclature system that focuses on virus lineages that contribute to global transmission and on genetic diversity by declassifying virus lineages that become inactive. The authors obtained 2685 SARS-CoV-2 genomes from the GISAID database, and identified 13 lineages, as well as some additional descendant lineages, by using this system.

There may be other, possibly better systems for SARS-CoV-2 nomenclature, but this is a reasonably good proposal. However, whether this manuscript is suitable for Nature Microbiology is questionable because this is just a proposal and no research was involved. This manuscript is more suitable for the Bulletin of the World Health Organization.

Other points:

1. Page 5. How did the authors classify lineages as 'active', 'unobserved', or 'inactive'?

2. Page 6. The date of the genome downloading from GISAID should be April 9, 2020, if the information shown in the method section is correct.

3. Page 7. The authors suggest that the status of the currently circulating lineages should be assessed at regular intervals and unobserved or inactive lineages should be no longer labeled. Does this mean that the number of names will be changed regularly to reflect the current epidemiological situation? I think changing the number of names would cause confusion.

Reviewer #2 (Remarks to the Author):

Rambaut et al., proposed a nomenclature of SARS-CoV-2 lineages to assist research on epidemiology and decision making during the COVID-19 pandemic. It is the non-ordinary proposal at the extraordinary times, as stated the clearly by the authors, who include leaders in the field. They listed major challenges, which include a global and fast expanding scale of the pandemic, an unprecedented pace of genome sequencing, and the lack of a template for the lineage nomenclature to emulate. To address these challenges, the authors offered the nomenclature that is based on the virus phylogeny and designated to provide a real-time bird-view perspective on the diversity of hundreds thousands genome sequences collected worldwide. To do so, the authors developed a set of rules to produce a hierarchical 4-level nomenclature of labels that is flexible and dynamic. It focuses on viruses of the most abundant and currently circulating lineages that facilitates a timely overview of the pandemic at global and local scales. The authors further proposed this nomenclature be compatible with other nomenclatures.

Major concerns.

1. The proposed nomenclature is the root-dependent that is evolutionary sensible. However, this makes stability of the entire nomenclature system dependent on stability of internal nodes of the virus tree. Although major sequencing effort may be directed toward analyzing most recent cases of the SARS-CoV-2 pandemic (terminal nodes), we may expect researchers go also after archive specimens. Their analysis might lead to revision of the already delineated lineages and their nomenclatures, including those for the basal A and B lineages. How could this complication be addressed within the proposed system?

2. One way to deal with the above complication is producing fixed dated releases of the lineage classification and nomenclature, fully independent from prior releases. Is this what the authors want to do? If not, the authors should also explain how comparison between results obtained and published at different time points will be ensured.

3. According to my reading Table 1 and Figure 1, lineage B may be most sensitive to future updates. Its monophyly is poorly supported, and it has a long internal branch leading to a major subset including B1.1-B1.20 lineages. The very delineation of this lineage seems to be an example of an expert decision which was made using a special provision in the rules for exceptions, whose rationale and consequences were only briefly commented on (p. 7, a sentence in the first paragraph). It is plausible that the future expanded sampling of the virus may lead to the split of the B lineage, close to its root. This will affect the nomenclature of all descending lineages. If the authors agree with this concern, they may want to use this example to illustrate their reasoning regarding exceptions to the rules and its applicability to other possible cases.

4. Who will oversee application of the nomenclature to newly sequenced genomes, especially if the

label designation will involve a rule exception? Could authors of those new sequences designate lineage labels on their own or they must consult the authors of this manuscript for assistance?

5. Further to the above and especially if the authors' ambition is to make their nomenclature a global standard, I would encourage the authors to contact John Ziebuhr, Chair of the Coronaviridae Study Group, regarding this proposed nomenclature before it is published. Although formally there is no authority overseeing classification and nomenclature matters below species rank of the coronavirus taxonomy, the CSG has effectively served this function over many years, which is in line with a prevailing practice in virology.

Specific suggestions:

1. Title: consider including "lineages" after SARS-CoV-2.
2. Text: literature sources cited must be expanded considerably in the manuscript, especially regarding classification issues.
3. Text: subsection headings to separate background, different parts of the proposal and its discussion would improve reading experience.
4. Text: indicate a virus evolution time-frame or a divergence range which is going to be covered by this nomenclature.
5. p. 2 and elsewhere: could lineage be 'reactivated', even retroactively, due to expanded genome sequencing effort?
6. p. 2: please specify the taxonomy of SARS-CoV-2.
7. p. 8. Methods. Due to practical considerations, the authors have accepted incomplete genome sequences to the analysis. It would be highly informative to demonstrate, if it is feasible, a relationship between a classification that is based on the incomplete genome sequences and another one that is based on a subset that includes the complete genome sequences. Basically, the question is about assessing an impact of the variation in the trimmed terminal sequences and the non-resolved positions which were excluded from the analysis.
8. Table 1 and Figure 1: there is a mismatch between the lineages detailed and labelled in these two illustrations; this should be addressed.
9. Table 1, and Figure 1: please indicate 'active' and 'inactive' lineages. Are 'unobserved' and 'inactive' synonyms?
10. Figure 1: scale bar is missing. Not clear what "triangles" stand for.

Reviewer #3 (Remarks to the Author):

This is is a sensible and timely proposal. I have just a few points to consider:

1. Introductory paragraph: The term "global pandemic" is redundant in my books. Seems to have gained wide use and there are more important things to worry about. But maybe consider changing to "pandemic"?

2. Agree a nomenclature system is needed and this one seems the best thought through. It is practical and epidemiologically relevant, it should be flexible and useful for a long time to come.

3. Regarding including recombinants in the system: "if recombinant lineages arise, exhibit onward spread, and satisfy the requirements for lineage designation outlined above, then they will be assigned the next available alphabetical prefix irrespective of what their parent lineages are": I can imagine that satisfying the 70% bootstrap value criterion might not be reliable or meaningful for recombinant

lineages. Any thoughts on that and whether that particular criterion might need to be relaxed or modified for recombinants?

4. I couldn't see the scale bar in the figure.

Nice system!

Author Rebuttal to Initial comments

Reviewer #1 (Remarks to the Author):

As the SARS-CoV-2 pandemic has spread, thousands of virus genome sequences have been generated. However, there is no coherent or accepted system for nomenclature. Such a system should be flexible and dynamic to reflect the birth and death of viral lineages. The authors have proposed a nomenclature system that focuses on virus lineages that contribute to global transmission and on genetic diversity by declassifying virus lineages that become inactive. The authors obtained 2685 SARS-CoV-2 genomes from the GISAID database, and identified 13 lineages, as well as some additional descendant lineages, by using this system.
Response: Since submitting this paper we have continually updated the lineage designation and are now working with more than 35,000 SARS-CoV-2 genomes from which we have identified 81 lineages (as well as 17 putative lineages). We have also created software for assigning new genomes that we briefly describe here.
Other points:
1. Page 5. How did the authors classify lineages as 'active', 'unobserved', or 'inactive'?
Response: We have now clarified how each lineage is classified into each of the categories as follows: "Accordingly, lineages of SARS-CoV-2 documented within the last month are defined here as 'active', those last seen >1 month but <3 months ago are classified as 'unobserved', and those that have not been seen for >3 months are termed 'inactive'." It is important to note that these are provisional timescales and the category thresholds may be altered in the future once the dynamics of lineage generation and extinction are better understood.
2. Page 6. The date of the genome downloading from GISAID should be April 9, 2020, if the information shown in the method section is correct.
Response: We have updated the dates to reflect the currently available data.
3. Page 7. The authors suggest that the status of the currently circulating lineages should be assessed at regular intervals and unobserved or inactive lineages should be no longer labeled. Does this mean that the number of names will be changed regularly to reflect the current epidemiological situation? I think changing the number of names would cause confusion.
Response: The number of active names will increase as new lineages appear and are described. We will not reuse lineage names but some may simply go out of use, because the viruses in question are no longer circulating.

Reviewer #2 (Remarks to the Author):

Rambaut et al., proposed a nomenclature of SARS-CoV-2 lineages to assist research on epidemiology and decision making during the COVID-19 pandemic. It is the non-ordinary proposal at the extraordinary times, as stated the clearly by the authors, who include leaders in

the field. They listed major challenges, which include a global and fast expanding scale of the pandemic, an unprecedented pace of genome sequencing, and the lack of a template for the lineage nomenclature to emulate. To address these challenges, the authors offered the nomenclature that is based on the virus phylogeny and designated to provide a real-time birdview
perspective on the diversity of hundreds thousands genome sequences collected worldwide. To do so, the authors developed a set of rules to produce a hierarchical 4-level nomenclature of labels that is flexible and dynamic. It focuses on viruses of the most abundant and currently circulating lineages that facilitates a timely overview of the pandemic at global and local scales. The authors further proposed this nomenclature be compatible with other nomenclatures.

1. The proposed nomenclature is the root-dependent that is evolutionary sensible. However, this makes stability of the entire nomenclature system dependent on stability of internal nodes of the virus tree. Although major sequencing effort may be directed toward analyzing most recent cases of the SARS-CoV-2 pandemic (terminal nodes), we may expect researchers go also after archive specimens. Their analysis might lead to revision of the already delineated lineages and their nomenclatures, including those for the basal A and B lineages. How could this complication be addressed within the proposed system?

Response: The reviewer raises an interesting point, although one that is largely academic. There is a remote possibility that further sampling of human cases will lead to the identification of SARS-CoV-2 genomes do not share the same most recent common ancestor (MRCA) as the 35,000 genomes sampled so far, but instead share an older ancestor (i.e. they will be outgroups to the pandemic). If such genomes exist and are found then they will represent chains of infection that have remained very rare, or have likely gone extinct, and hence ***are not the direct ancestors of the pandemic****. In that sense, they are of no relevance to our scheme, which is intended to push forward through time and to represent the evolution of the pandemic itself. So, in this event, we would not change the root of our classification scheme - it would remain between lineages A and B - because that is clearly the starting point of the pandemic. The outgroup genomes would be simply labelled as "pre-pandemic" strains.*

We do not propose discussing this in the manuscript as it is a very unlikely event that will have no effect on our nomenclature system. Unfortunately, the issue of the pandemic origin is hyper-sensitised and politicised. Even the suggestion of prepandemic ancestors, no matter how improbable, will likely be misrepresented and draw attention away from the true purpose of our manuscript.

2. One way to deal with the above complication is producing fixed dated releases of the lineage classification and nomenclature, fully independent from prior releases. Is this what the authors want to do? If not, the authors should also explain how comparison between results obtained and published at different time points will be ensured.

Response: Please see our above response to the Editor's comment #2 for a description of how dated releases will be generated and archived.

3. According to my reading Table 1 and Figure 1, lineage B may be most sensitive to future updates. Its monophyly is poorly supported, and it has a long internal branch leading to a major subset including B1.1-B1.20 lineages. The very delineation of this lineage seems to be

an example of an expert decision which was made using a special provision in the rules for exceptions, whose rationale and consequences were only briefly commented on (p. 7, a sentence in the first paragraph). It is plausible that the future expanded sampling of the virus may lead to the split of the B lineage, close to its root. This will affect the nomenclature of all descending lineages. If the authors agree with this concern, they may want to use this example to illustrate their reasoning regarding exceptions to the rules and its applicability to other possible cases.

Response: Currently, lineage B is defined by two mutations. Should any sequences come to light that only has one or other of these two mutations then these would be classified as A. The long branch to the B.1.1 - B.1.20 subset is unusual because it seems to be a simultaneous mutation at three adjacent nucleotides (causing two amino acid changes). Because of the distinctive nature of this mutation we have decided to call it a lineage (B.1.1: sub-lineages are pushed down a level).

4. Who will oversee application of the nomenclature to newly sequenced genomes, especially if the label designation will involve a rule exception? Could authors of those new sequences designate lineage labels on their own or they must consult the authors of this manuscript for assistance?

Response: For the implementation of the system we will utilise the facilities afforded by the GitHub repository we are using. This provides a 'versioning' system for all files, tracking changes and who made them, the ability to suggest changes through 'pull requests', an 'Issue' tracker by which users can suggest new lineages, highlight possible rule exceptions etc. While we provide clear links to this nascent online community in our paper, we are keen to keep the practicalities of implementing the system independent of this description of how we envisage it working.

5. Further to the above and especially if the authors' ambition is to make their nomenclature a global standard, I would encourage the authors to contact John Ziebuhr, Chair of the Coronaviridae Study Group, regarding this proposed nomenclature before it is published. Although formally there is no authority overseeing classification and nomenclature matters below species rank of the coronavirus taxonomy, the CSG has effectively served this function over many years, which is in line with a prevailing practice in virology.

Response: Please see our above response to the Editor's comment #3: the ICTV do not deal with genotypic-level classifications.

1. Title: consider including "lineages" after SARS-CoV-2.

Response: Good suggestion – done.

2. Text: literature sources cited must be expanded considerably in the manuscript, especially regarding classification issues.

Response: As requested, we have added additional citations regarding classification and the position of ICTV on within-species taxonomy and naming.

3. Text: subsection headings to separate background, different parts of the proposal and its discussion would improve reading experience.

Response: Done.

4. Text: indicate a virus evolution time-frame or a divergence range which is going to be covered by this nomenclature.

Response: The reviewer raises an interesting point. We expect the "dynamic" phase of

the nomenclature to last for the duration of the global pandemic, which is likely to be 2-3 years. After that time, and assuming that SARS-CoV-2 will not be globally eliminated, the virus will become an endemic (and likely seasonal) infection. It is most likely that only a small number of lineages will survive to become endemic, and will at that point in time, have become genetically divergent. As discussed in our paper, under such circumstances a typical (and less dynamic) "subtype" based classification is entirely adequate. The remaining endemic lineages can simply retain their names from our dynamic nomenclature system, which is designed for the pandemic phase.
This is now explained in the Discussion section.
5. p. 2 and elsewhere: could lineage be 'reactivated', even retroactively, due to expanded genome sequencing effort?
Response: Yes. The inactive label simply highlights that no viruses in this lineage have been sampled recently. This is now explained in the text as follows: "Hence it is possible for lineages that were previously classified as inactive or unobserved to be re-labelled as active."
6. p. 2: please specify the taxonomy of SARS-CoV-2.
Response: We have now added the current taxonomy of SARS-CoV-2.
7. p. 8. Methods. Due to practical considerations, the authors have accepted incomplete genome sequences to the analysis. It would be highly informative to demonstrate, if it is feasible, a relationship between a classification that is based on the incomplete genome sequences and another one that is based on a subset that includes the complete genome sequences. Basically, the question is about assessing an impact of the variation in the trimmed terminal sequences and the non-resolved positions which were excluded from the analysis.
Response: It is very difficult to assess the correctness of the sequencing of the 5' and 3' ends of the virus genome and the lengths of these vary considerably depending on the primers being used and the effort expended to capture the full genome. Because of this we made the decision to only use the coding region of the genome to designate the lineages and then to assign them. There may be some SNPs outside the coding region that are informative about some lineages, but we are only losing ~1.5% of the genome length (493 of 29903 nt) (and many of the sequences generated are missing some of this anyway). We have added some additional text about this.
8. Table 1 and Figure 1: there is a mismatch between the lineages detailed and labelled in these two illustrations; this should be addressed.
Response: We have updated the tables and figure to provide the latest lineage designation.
9. Table 1, and Figure 1: please indicate 'active' and 'inactive' lineages. Are 'unobserved' and 'inactive' synonyms?
Response: We have clarified these terms (see response to Reviewer #1) and now highlight these categories in the table and figure.
10. Figure 1: scale bar is missing. Not clear what "triangles" stand for.
Response: We have added a scale bar and clarified what the triangles denote.

Reviewer #3 (Remarks to the Author):

This is a sensible and timely proposal. I have just a few points to consider:
1. Introductory paragraph: The term "global pandemic" is redundant in my books. Seems to

have gained wide use and there are more important things to worry about. But maybe consider changing to “pandemic”?
Response: Good point – done.
2. Agree a nomenclature system is needed and this one seems the best thought through. It is practical and epidemiologically relevant, it should be flexible and useful for a long time to come.
Response: Many thanks for this comment.
3. Regarding including recombinants in the system: "if recombinant lineages arise, exhibit onward spread, and satisfy the requirements for lineage designation outlined above, then they will be assigned the next available alphabetical prefix irrespective of what their parent lineages are”: I can imagine that satisfying the 70% bootstrap value criterion might not be reliable or meaningful for recombinant lineages. Any thoughts on that and whether that particular criterion might need to be relaxed or modified for recombinants?
Response: We agree that recombination presents many difficulties with regards to phylogenetic analysis. However, our approach is advantageous because it can explicitly handle recombination. We do not attempt to define sets of nested clades, but rather common ancestors and their descendants. A recombination event creates a new common ancestor that will be the root of a lineage and as we state in the text, because it is the product of more than one ancestral lineage, we would assign these a new letter.
4. I couldn't see the scale bar in the figure.
Response: Thank you. We have added a scale bar.
Nice system!
Response: Thanks again!

Decision Letter, first revision:
Dear Eddie,

Thank you for your patience while your manuscript "A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology" was under peer review at Nature Microbiology. It has now been seen by our referees, and in the light of their advice I am delighted to say that we can in principle offer to publish it. First, however, we would like you to revise your paper to address the points made by the reviewers, and to ensure that it is in Nature Microbiology format.

Referee #2 had still some concerns that addressed in the comments to the editor. In order to solve this, please, clearly note in the paper that the emergence of more sequences will likely cause changes to their current proposal, as well as potential skewing towards certain countries due to availability of genome data. Also, you should explain how the tree topology will be revised going forward when more genomes are available (This does not need to be a full explanation some outline will be useful). Please, also clarify in the manuscript text that ICTV only deals with changes at species rank and above and, please, let me know if you had contacted the ICTV about this work.

The referees’ remaining comments are clear, and should not be difficult to implement. Editorially, we will need you to make some changes so that the paper complies with our Guide to Authors at http://www.nature.com/nmicrobiol/info/gta.

I appreciate this email is long and recommend that you print it and use it as a checklist, reading it carefully to the end, in order to avoid speed the process to final acceptance and thus avoid delays to publication.

Nature Microbiology offers a transparent peer review option for new original research manuscripts submitted from 1st December 2019. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. Please state in the cover letter 'I wish to participate in transparent peer review' if you want to opt in, or 'I do not wish to participate in transparent peer review' if you don't. Failure to state your preference will result in delays in accepting your manuscript for publication.
Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our FAQ page.

Specific points:

In particular, while checking through the manuscript and associated files, we noticed the following specific points which we will need you to address:

1. ORCID. As mentioned previously, we ask all corresponding authors to provide their unique ORCID identifiers at the time of final submission. This information is currently missing for Dr. Pybus. Please see below for additional information on how to sign up to ORCID and link your account to the one in our manuscript tracking system.

2. Author Contributions. Please provide a more detailed and specific author contributions statement. A good example can be found at the end of the following article

http://www.nature.com/nature/journal/v532/n7599/full/nature17433.html

3. Code Availability. Please provide a Code Availability statement and deposit any custom code to GitHub.

4. Data availability. Please include a data availability section at the end of the methods - see below for additional details on how to format this section. Note that this section should include all accession codes for sequences deposited to databases. Please note that all accession codes must be live by the time of publication of the piece.

5. Reporting checklist. Note that a final version of the reporting checklist will be published with your manuscript. Therefore, please revise this document according to the instructions found in the annotated PDF attached to this message.

General points:

Please read carefully through all of the following general formatting points when preparing the final version of your manuscript, as submitting the manuscript files in the required format will greatly speed

the process to final acceptance of you work.

We estimate that your manuscript currently exceeds our normal length limit for Articles of about 3,000 words. We have some flexibility, and can allow a revised manuscript at 3,500 words, but please consider this a firm upper limit. You could achieve some shortening by moving some details to the Methods section that should follow the main text (the length of the Methods section is unlimited and does not count towards the main text length).

Titles should give an idea of the main finding of the paper and ideally not exceed 90 characters (including spaces). We discourage the use of active verbs and do not allow punctuation.

The paper's summary paragraph (about 150-200 words; no references) should serve both as a general introduction to the topic, and as a brief, non-technical summary of your main results and their implications. It should start by outlining the background to your work (why the topic is important) and the main question you have addressed (the specific problem that initiated your research), before going on to describe your new observations, main conclusions and their general implications. Because we hope that scientists across the wider microbiology community will be interested in your work, the first paragraph should be as accessible as possible, explaining essential but specialised terms concisely. We suggest you show your summary paragraph to colleagues in other fields to uncover any problematic concepts.

We strongly support public availability of data. Please place the data used in your paper into a public data repository, if one exists, or alternatively, present the data as Source Data or Supplementary Information. If data can only be shared on request, please explain why in your Data Availability Statement, and also in the correspondence with your editor. For some data types, deposition in a public repository is mandatory - more information on our data deposition policies and available repositories can be found at https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data.

Please include a data availability statement as a separate section after Methods but before references, under the heading "Data Availability". This section should inform readers about the availability of the data used to support the conclusions of your study. This information includes accession codes to public repositories (data banks for protein, DNA or RNA sequences, microarray, proteomics data etc…), references to source data published alongside the paper, unique identifiers such as URLs to data repository entries, or data set DOIs, and any other statement about data availability. At a minimum, you should include the following statement: "The data that support the findings of this study are available from the corresponding author upon request", mentioning any restrictions on availability. If DOIs are provided, we also strongly encourage including these in the Reference list (authors, title, publisher (repository name), identifier, year). For more guidance on how to write this section please see:
http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf

Please supply the figures as vector files - EPS, PDF, AI or postscript (PS) file formats (not raster or bitmap files), preferably generated with vector-graphics software (Adobe Illustrator for example). Try to ensure that all figures are non-flattened and fully editable. All images should be at least 300 dpi resolution (when figures are scaled to approximately the size that they are to be printed at) and in RGB colour format. Please do not submit Jpeg or flattened TIFF files. Please see also 'Guidelines for Electronic Submission of Figures' at the end of this letter for further detail.
Please view http://www.nature.com/authors/editorial_policies/image.html for more detailed guidelines.

We will edit your figures/tables electronically so they conform to Nature Microbiology style. If necessary, we will re-size figures to fit single or double column width. If your figures contain several parts, the parts should be labelled lower case a, b, and so on, and form a neat rectangle when assembled.

Please check the PDF of the whole paper and figures (on our manuscript tracking system) VERY CAREFULLY when you submit the revised manuscript. This will be used as the 'reference copy' to make sure no details (such as Greek letters or symbols) have gone missing during file-transfer/conversion and re-drawing.

All Supplementary Information must be submitted in accordance with the instructions in the attached Inventory of Supporting Information, and should fit into one of three categories:

1. EXTENDED DATA: Extended Data are an integral part of the paper and only data that directly contribute to the main message should be presented. These figures will be integrated into the full-text HTML version of your paper and will be appended to the online PDF. There is a limit of 10 Extended Data figures, and each must be referred to in the main text. Each Extended Data figure should be of the same quality as the main figures, and should be supplied at a size that will allow both the figure and legend to be presented on a single legal-sized page. Each figure should be submitted as an individual .jpg, .tif or .eps file with a maximum size of 10 MB each. All Extended Data figure legends must be provided in the attached Inventory of Accessory Information, not in the figure files themselves.

2. SUPPLEMENTARY INFORMATION: Supplementary Information is material that is essential background to the study but which is not practical to include in the printed version of the paper (for example, video files, large data sets and calculations). Each item must be referred to in the main manuscript and detailed in the attached Inventory of Accessory Information. Tables containing large data sets should be in Excel format, with the table number and title included within the body of the table. All textual information and any additional Supplementary Figures (which should be presented with the legends directly below each figure) should be provided as a single, combined PDF. Please note that we cannot accept resupplies of Supplementary Information after the paper has been formally accepted unless there has been a critical scientific error.

All Extended Data must be called you in your manuscript and cited as Extended Data 1, Extended Data 2, etc. Additional Supplementary Figures (if permitted) and other items are not required to be called out in your manuscript text, but should be numerically numbered, starting at one, as Supplementary Figure 1, not SI1, etc.

3. SOURCE DATA: We strongly encourage you to provide source data for your figures whenever possible. Full-length, unprocessed gels and blots must be provided as source data for any relevant figures, and should be provided as individual PDF files for each figure containing all supporting blots and/or gels with the linked figure noted directly in the file. Numerical source data that underlie graphs are required for in vivo experiments and strongly encouraged generally. They should be provided in Excel format, one file for each relevant figure, with the linked figure noted directly in the file. They should be clearly labelled such that individual experiments and/or animals are labelled (for example, across a time course if applicable). For imaging source data, we encourage deposition to a relevant repository, such as figshare (https://figshare.com/) or the Image Data Resource (https://idr.openmicroscopy.org).

Please include any references for the Methods at the end of the reference list. Any citations in the

Supplemental Information will need inclusion in a separate SI reference list.

It is a condition of publication that you include a statement before the acknowledgements naming the author to whom correspondence and requests for materials should be addressed.

Finally, we require authors to include a statement of their individual contributions to the paper -- such as experimental work, project planning, data analysis, etc. -- immediately after the acknowledgements. The statement should be short, and refer to authors by their initials. For details please see the Authorship section of our joint Editorial policies at http://www.nature.com/authors/editorial_policies/authorship.html

We will not send your revised paper for further review if, in the editors' judgement, the referees' comments on the present version have been addressed. If the revised paper is in Nature Microbiology format, in accessible style and of appropriate length, we shall accept it for publication immediately.

Please resubmit electronically

* the final version of the text (not including the figures) in either Word or Latex.

* publication-quality figures. For more details, please refer to our Figure Guidelines, which is available here: https://mts-nmicrobiol.nature.com/letters/Figure_guidelines.pdf

* Extended Data & Supplementary Information, as instructed

* a point-by-point response to any issues raised by our referees and to any editorial suggestions.

* any suggestions for cover illustrations, which should be provided at high resolution as electronic files. Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of Nature Microbiology.

Please use the following link to access your home page:

[REDACTED]

* This url links to your confidential homepage and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this e-mail to co-authors, please delete this link to your homepage first.

Please also send the following forms as a PDF by email to microbiology@nature.com.

* Please sign and return the Licence to Publish form .

* Or, if the corresponding author is either a Crown government employee (including Great Britain and Northern Ireland, Canada and Australia), or a US Government employee, please sign and return the Licence to Publish form for Crown government employees, or a Licence to Publish form for US government employees.

* Should your Article contain any items (figures, tables, images, videos or text boxes) that are the

same as (or are adaptations of) items that have previously been published elsewhere and/or are owned by a third party, please note that it is your responsibility to obtain the right to use such items and to give proper attribution to the copyright holder. This includes pictures taken by professional photographers and images downloaded from the internet. If you do not hold the copyright for any such item (in whole or part) that is included in your paper, please complete and return this Third Party Rights Table, and attach any grant of rights that you have collected.

For more information on our licence policy, please consult http://npg.nature.com/authors.

AUTHORSHIP

ORCID

Nature Microbiology is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS) prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. For more information please visit http://www.springernature.com/orcid

For all corresponding authors listed on the manuscript, please follow the instructions in the link below to link your ORCID to your account on our MTS before submitting the final version of the manuscript. If you do not yet have an ORCID you will be able to create one in minutes. https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research

IMPORTANT: All authors identified as 'corresponding author' on the manuscript must follow these instructions. Non-corresponding authors do not have to link their ORCIDs but are encouraged to do so. Please note that it will not be possible to add/modify ORCIDs at proof. Thus, if they wish to have their ORCID added to the paper they must also follow the above procedure prior to acceptance.

To support ORCID's aims, we only allow a single ORCID identifier to be attached to one account. If you have any issues attaching an ORCID identifier to your MTS account, please contact the Platform Support Helpdesk.

Nature Research journals encourage authors to share their step-by-step experimental protocols on a protocol sharing platform of their choice. Nature Research's Protocol Exchange is a free-to-use and open resource for protocols; protocols deposited in Protocol Exchange are citable and can be linked from the published article. More details can found at www.nature.com/protocolexchange/about.

We hope to hear from you within two weeks; please let us know if the revision process is likely to take longer.

Yours sincerely,
Paula

****
Reviewer Comments:

Reviewer #1 (Remarks to the Author):

The authors responded to my comments appropriately.

Reviewer #2 (Remarks to the Author):

27: …flagging for de-labelling virus lineages that become unobserved or inactive.
43-44: …SARS-CoV-2, a clade within the species Severe acute respiratory syndrome-related virus, subgenus Sarbecovirus, genus Betacoronavirus, subfamily Coronavirinae, family Coronaviridae [Ref: Gorbalenya et al., Nat Microbiol 5: 536–544 (2020)], is urgently…
181-186: consider adding that no C level lineage is observed at the time of writing.
195: …predominant known global lineage…
212-214: consider adding that names of unobserved and inactive lineages will not be reassigned and this provision will facilitate comparative analysis of lineages, regardless of their circulating status.

Reviewer #3 (Remarks to the Author):

Great responses not only to my points but the other points as well.

Author Rebuttal, first revision:
Editorial comments:

Referee #2 had still some concerns that addressed in the comments to the editor. In order to solve this, please, clearly note in the paper that the emergence of more sequences will likely cause changes to their current proposal, as well as potential skewing towards certain countries due to availability of genome data. Also, you should explain how the tree topology will be revised going forward when more genomes are available (This does not need to be a full explanation some outline will be useful). Please, also clarify in the manuscript text that ICTV only deals with changes at species rank and above and, please, let me know if you had contacted the ICTV about this work.

Response: All these statements (that more data may change the proposal, that there may be a skew in sampling toward specific countries, and the role of the ICTV) have been added as requested. In addition, have now contacted the ICTV.

Nature Microbiology offers a transparent peer review option for new original research manuscripts submitted from 1st December 2019. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. **Please state in the cover letter 'I wish to participate in transparent peer review' if you want to opt in, or 'I do not wish to participate in transparent peer review' if you don't.** Failure to state your preference will result in delays in accepting your manuscript for publication.

Response: Done.

Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our FAQ page.

Specific points:

In particular, while checking through the manuscript and associated files, we noticed the following specific points which we will need you to address:

1. ORCID. As mentioned previously, we ask all corresponding authors to provide their unique ORCID identifiers at the time of final submission. This information is currently missing for Dr. Pybus. Please see below for additional information on how to sign up to ORCID and link your account to the one in our manuscript tracking system.

Response: This has been added as requested. Dr. Pybus's OrcID number = 0000-0002-8797-2667

2. Author Contributions. Please provide a more detailed and specific author contributions statement. A good example can be found at the end of the following article

http://www.nature.com/nature/journal/v532/n7599/full/nature17433.html

Response: These have been expanded slighted.
3. Code Availability. Please provide a Code Availability statement and deposit any custom code to GitHub.

Response: A code statement has been added as requested.

4. Data availability. Please include a data availability section at the end of the methods - see below for

additional details on how to format this section. Note that this section should include all accession codes for sequences deposited to databases. Please note that all accession codes must be live by the time of publication of the piece.

Response: This statement has been added as requested.

5. Reporting checklist. Note that a final version of the reporting checklist will be published with your manuscript. Therefore, please revise this document according to the instructions found in the annotated PDF attached to this message.

Response: This has been revised as suggested.

We estimate that your manuscript currently exceeds our normal length limit for Articles of about 3,000 words. We have some flexibility, and can allow a revised manuscript at 3,500 words, but please consider this a firm upper limit. You could achieve some shortening by moving some details to the Methods section that should follow the main text (the length of the Methods section is unlimited and does not count towards the main text length).

Titles should give an idea of the main finding of the paper and ideally not exceed 90 characters (including spaces). We discourage the use of active verbs and do not allow punctuation.

The paper's summary paragraph (about 150-200 words; no references) should serve both as a general introduction to the topic, and as a brief, non-technical summary of your main results and their implications. It should start by outlining the background to your work (why the topic is important) and the main question you have addressed (the specific problem that initiated your research), before going on to describe your new observations, main conclusions and their general implications. Because we hope that scientists across the wider microbiology community will be interested in your work, the first paragraph should be as accessible as possible, explaining essential but specialised terms concisely. We suggest you show your summary paragraph to colleagues in other fields to uncover any problematic concepts.

Response: Our paper confirms to all these requirements.

We strongly support public availability of data. Please place the data used in your paper into a public data repository, if one exists, or alternatively, present the data as Source Data or Supplementary Information. If data can only be shared on request, please explain why in your Data Availability Statement, and also in the correspondence with your editor. For some data types, deposition in a public repository is mandatory - more information on our data deposition policies and available repositories can be found at https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data.

Response: Done.

Please include a data availability statement as a separate section after Methods but before references, under the heading "Data Availability". This section should inform readers about the availability of the data used to support the conclusions of your study. This information includes accession codes to public repositories (data banks for protein, DNA or RNA sequences, microarray, proteomics data etc...), references to source data published alongside the paper, unique identifiers such as URLs to data repository entries, or data set DOIs, and any other statement about data availability. At a minimum, you should include the following statement: "The data that support the findings of this study are available from the corresponding author upon request", mentioning any restrictions on availability. If DOIs are provided, we also strongly encourage including these in the Reference list (authors, title, publisher (repository name), identifier, year). For more guidance on how to write this section please see: http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf

Response: Done.

Please supply the figures as vector files - EPS, PDF, AI or postscript (PS) file formats (not raster or bitmap files), preferably generated with vector-graphics software (Adobe Illustrator for example). Try to ensure that all figures are non-flattened and fully editable. All images should be at least 300 dpi resolution (when figures are scaled to approximately the size that they are to be printed at) and in RGB colour format. Please do not submit Jpeg or flattened TIFF files. Please see also 'Guidelines for Electronic Submission of Figures' at the end of this letter for further detail.
Please view http://www.nature.com/authors/editorial_policies/image.html for more detailed guidelines.

Response: Done.

All Supplementary Information must be submitted in accordance with the instructions in the attached Inventory of Supporting Information, and should fit into one of three categories:

Response: Not applicable.

Please include any references for the Methods at the end of the reference list. Any citations in the Supplemental Information will need inclusion in a separate SI reference list.

Response: Done.

It is a condition of publication that you include a statement before the acknowledgements naming the author to whom correspondence and requests for materials should be addressed.

Response: Done.

Finally, we require authors to include a statement of their individual contributions to the paper -- such as experimental work, project planning, data analysis, etc. -- immediately after the acknowledgements. The statement should be short, and refer to authors by their initials. For details please see the Authorship section of our joint Editorial policies at http://www.nature.com/authors/editorial_policies/authorship.html

Response: Done.

Reviewer Comments:

Reviewer #1 (Remarks to the Author):

The authors responded to my comments appropriately.

Response: We thank the reviewer for this comment.

Reviewer #2 (Remarks to the Author):

27: ...flagging for de-labelling virus lineages that become unobserved or inactive.
43-44: ...SARS-CoV-2, a clade within the species Severe acute respiratory syndrome-related virus, subgenus Sarbecovirus, genus Betacoronavirus, subfamily Coronavirinae, family Coronaviridae [Ref: Gorbalenya et al., Nat Microbiol 5: 536–544 (2020)], is urgently...
181-186: consider adding that no C level lineage is observed at the time of writing.
195: ...predominant known global lineage...
212-214: consider adding that names of unobserved and inactive lineages will not be reassigned and this provision will facilitate comparative analysis of lineages, regardless of their circulating status.

Response: All these minor changes have been made as suggested.

Reviewer #3 (Remarks to the Author):

Great responses not only to my points but the other points as well.

Response: We thank the reviewer for this comment.

Final Decision Letter:

Dear Eddie,

I am pleased to accept your Article "A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology" for publication in Nature Microbiology. Thank you for having chosen to submit your work to us and many congratulations.

Before your manuscript is typeset, we will edit the text to ensure it is intelligible to our wide readership and conforms to house style. We look particularly carefully at the titles of all papers to ensure that they are relatively brief and understandable.

The subeditor may send you the edited text for your approval. Once your manuscript is typeset you will receive a link to your electronic proof via email within 20 working days, with a request to make any corrections within 48 hours. If you have queries at any point during the production process then please contact the production team at rjsproduction@springernature.com. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see www.nature.com/nmicrobiolate/authors/gta/content-type/index.html). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication date (the day on which it is uploaded onto our website).

The Author's Accepted Manuscript (the accepted version of the manuscript as submitted by the author) may only be posted 6 months after the paper is published, consistent with our self-archiving embargo. Please note that the Author's Accepted Manuscript may not be released under a Creative Commons license. For Nature Research Terms of Reuse of archived manuscripts please see: http://www.nature.com/authors/policies/license.html#terms
If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

We welcome the submission of potential cover material (including a short caption of around 40 words) related to your manuscript; suggestions should be sent to Nature Microbiology as electronic files (the image should be 300 dpi at 210 x 297 mm in either TIFF or JPEG format). Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content, and that colour images work better than black and white or grayscale images. Please do not try to design a cover with the Nature Microbiology logo etc., and please do not submit composites of images related to your work. I am sure you will understand that we cannot make any promise as to whether any of

your suggestions might be selected for the cover of the journal.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

With kind regards,
Paula