Longest contig problem - sars-cov-2

11 messages

Robert Karlsson <robertkarlss.se@gmail.com> Wed, Dec 7, 2022 at 16:48
To: j.quick@bham.ac.uk

Dear Joshua Quick,

My name is Robert Karlsson. I am biochemist from Sweden.

I have a question for you about the SARS-CoV-2 reference genome Wuhan-Hu-1 because it is used for Artic Protocol.

Why is not possible to reproduce the longest contig (30,474) with raw reads associated with the description of itsown study (Fan Wu et al.)?

Probably, raw reads are not raw reads?

Thanks for your help.

Best regards,

Robert

Joshua Quick <J.Quick@bham.ac.uk> Wed, Dec 7, 2022 at 16:58
To: robertkarlss.se@gmail.com <robertkarlss.se@gmail.com>

Hi Robert,

You'd have to go back to the original paper and look at the methods. The MN908947.3 reference was improved with the addition of some 3' RACE I think.

Thanks

From: robertkarlss.se@gmail.com <robertkarlss.se@gmail.com>
Sent: 07 December 2022 15:48
To: Joshua Quick (Biosciences) <J.Quick@bham.ac.uk>
Subject: Longest contig problem - sars-cov-2

[Quoted text hidden]

Robert Karlsson <robertkarlss.se@gmail.com> Wed, Dec 7, 2022 at 17:09
To: Joshua Quick <J.Quick@bham.ac.uk>

Hi Joshua,

Thank you for the quick response.

Yes, they used RACE after the primers are designed in the second "step":

1. Non-target based amplification + Ilumina MiniSeq generated 56,565,928 reads and the longest contig is 30,474 nt

2. Target-based amplification + Sequencer shouldn't generate the same 56,565,928 million reads and the longest contig is 29,802 nt

The reads from the second step are somehow inserted in the raw data (1st step), but number of the reads shouldn't be 56,565,928. They did it backwards, some reads are removed and some reads are inserted in original first file.

Probably they were in hurry to show perfect contig to the public. I tried to reproduce experiment but I can't solve the problem.

Could you try to reproduce the experiment? I use galaxy community hub. You can get the result fast, in few hours. In few clicks.

Maybe I am wrong? Is it usual practice to change original reads, but again, there shouldn't be the same 56 million reads.

Thanks for your help.

Best,
Robert
[Quoted text hidden]

Robert Karlsson <robertkarlss.se@gmail.com> Thu, Dec 8, 2022 at 17:33
To: Joshua Quick <J.Quick@bham.ac.uk>

Hi Joshua,

Maybe you could ask prof. Yong-Zhen Zhang? I tried few months ago but he didn't respond.

I am very worried about this.

Is it allowed to remove "unwanted" sars-cov-2 reads from the original file and replace them with "better" reads? If this is not allowed, then the next explanation would be manipulation.

I urgently need some good explanation. If the data is manipulated then everyone should stop using MN908947.

I hope I am wrong.

Please respond as soon as possible, within 7 days at the latest.

Best regards,

Robert
[Quoted text hidden]

Joshua Quick <J.Quick@bham.ac.uk> Fri, Dec 9, 2022 at 12:04
To: robertkarlss.se@gmail.com <robertkarlss.se@gmail.com>

Hi Robert,

I expect there was a mis-assembly in the first version or something. The MN908947 genome is correct though, I designed primers off it and they amplify the viral genome perfectly, it also matches all the early patient samples sequenced in the UK.

Josh

From: robertkarlss.se@gmail.com <robertkarlss.se@gmail.com>
Sent: 07 December 2022 16:09

To: Joshua Quick (Biosciences) <J.Quick@bham.ac.uk>
Subject: Re: Longest contig problem - sars-cov-2

[Quoted text hidden]

Robert Karlsson <robertkarlss.se@gmail.com> Fri, Dec 9, 2022 at 14:30
To: Joshua Quick <J.Quick@bham.ac.uk>

Hi Joshua,

Thank you for the answer.

So, PCR-positive patient samples (sars-cov-2 positive) could generate sars-cov-2 whole-genome assembly.

Did you or your colleagues in UK try to exclude these unexpected possibilities:

1. Did you try to amplify any other virus (other than sars-cov-2, for example: zika virus with zika primers) from the covid-19 patient sample (sars-cov-2 PCR-positive nasopharyngeal swab sample) using amplicon-based WGS?

2. Did you try amplicon-based WGS approach for sars-cov-2 genome with sars-cov-2 primers but with RNA from cells and from supernatant of uninfected cell cultures (example: human epithelial cell cultures because swabs are in contact with the cells during sampling) treated in the same way (treatment for detection of CPE) as infected cell cultures but without any virus?

3. Did you try amplicon-based WGS approach for construction (assembly) of sars-cov-2 genome from sars-cov-2 negative sample (example: PCR-negative nasopharyngeal sample) using sars-cov-2 primers.

4. Have you tried to generate sars-cov-2 whole-genome assembly using healthy controls (healthy persons) or people suffering from some other disease (other than covid-19) ?

Best regards,

Robert

[Quoted text hidden]

Joshua Quick <J.Quick@bham.ac.uk> Fri, Dec 9, 2022 at 15:33
To: robertkarlss.se@gmail.com <robertkarlss.se@gmail.com>

Hi Robert,

I don't really understand questions 1 and 2. If SC2 positive cases were co-infected with another virus then you would be able to sequence that virus too. We only attempt sequencing on positive samples below Ct30 because sequencing is a lot more expensive than qPCR testing.

It is standard practice to runs positive and negative controls for every amplicon sequencing run, this is necessary to detect errors and possible amplicon contamination. There is a very high correlation between diagnostic Ct value and sequencing performance, either amplicon for RNA-Seq which is expected as they are directly linked to the titre of viral RNA in the sample.

Thanks

From: robertkarlss.se@gmail.com <robertkarlss.se@gmail.com>
Sent: 09 December 2022 13:30
[Quoted text hidden]

[Quoted text hidden]

Robert Karlsson <robertkarlss.se@gmail.com> Sat, Dec 10, 2022 at 11:17
To: Joshua Quick <J.Quick@bham.ac.uk>

Hi Joshua,

I'm sorry to bother you.

You use a negative control for sequencing. Can you describe that control? You probably mean water or buffer to exclude contamination?

The question 2 is proper negative control.

More info about question 2:

When a sterile swab comes into contact with the human mucosa, the consequence of the contact is the introduction of epithelial cells on the swab. This means that the negative control should contain RNA extracted from human epithelial cells. VTM (virus transport medium with antibiotics and antifungals) should be used in the generation of the negative control also. Considering that human epithelial cells are contaminated with human microbiome, it is necessary to use RNA from epithelial cell cultures. Epithelial cell cultures are virus-free. So, this control is very useful to exclude generation of sars-cov-2 genome without the presence of sars-cov-2 genome. I hope you could understand it now?

Why is this control important? Because the sequenced nucleic acids from the mixed-sample cannot be assigned to a specific structure or particle. For example, "de novo" approach would be logical in the case of isolated, purified, concentrated virus particles. So, the control would be very interesting. Maybe you don't agree with me? :)

I wish you a great weekend

Best,
Robert
[Quoted text hidden]

Robert Karlsson <robertkarlss.se@gmail.com> Mon, Dec 19, 2022 at 12:37
To: Joshua Quick <J.Quick@bham.ac.uk>

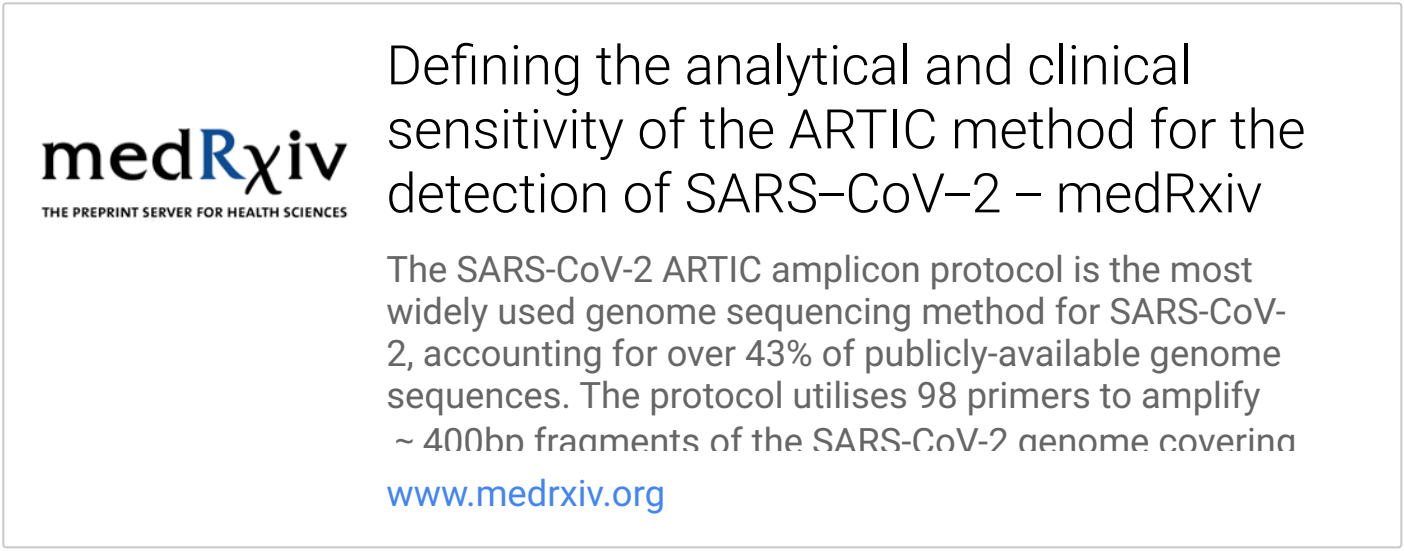
Reminder
[Quoted text hidden]

Joshua Quick <J.Quick@bham.ac.uk> Wed, Jan 4, 2023 at 14:50

To: robertkarlss.se@gmail.com <robertkarlss.se@gmail.com>

Hi Robert,

Yes there are different kinds of negative control you could use, a full end-to-end negative sample is common practise in 16S metagenomic sequencing. What I'm describing is a PCR negative control in which we use buffer instead of extracted RNA. The purpose of that control is to detect any amplicon contamination. If you see any amplicons in your PCR negative then you know you need to repeat the plate because amplicons have got into the reaction. You can identify a lot of issues purely using this method which we published as a preprint a while ago; https://www.medrxiv.org/content/10.1101/2021.10.09.21264695v1



Thanks

From: robertkarlss.se@gmail.com <robertkarlss.se@gmail.com>
Sent: 19 December 2022 11:37
[Quoted text hidden]

[Quoted text hidden]

Robert Karlsson <robertkarlss.se@gmail.com> Wed, Jan 4, 2023 at 16:21
To: Joshua Quick <J.Quick@bham.ac.uk>

Dear Joshua,

Thanks for the detailed explanation. I noticed that all scientists in viroligy use controls to rule out contamination, but no one thinks to rule out the possibility of fabrication by experiment

setup itself.

"The second general approach is to perform negative controls: to repeat the experiment under conditions in which it is expected to produce a null result and verify that it does indeed produce a null result. Several strategies are employed to design negative controls, such as:

• Leave out an essential ingredient."

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3053408/

Essentiall ingredient is a virus. So no one has tried to extract RNA from inside uninfected cells of human epithelial cell cultures. These RNAs should be treated with sars-cov-2 primers, i.e. with an amlicon-based WGS protocol.

There are currently about 300,000 sars-cov-2 studies and no one has attempted this logical control. Maybe I'm wrong, but I've read thousands of studies and no one has done (conducted and documented) the necessary control. Maybe there are unpublished results somewhere.

If you find that negative control in any study, please let me know. (Vero cells or any appropriate cell cultures can be used instead of human epithelial cell cultures)

Than you very much!

Best regards

Robert
[Quoted text hidden]