

From: Lau, Matthew K. matthewklau@fas.harvard.edu  
Subject: Re: ant sequences NCBI errors  
Date: May 24, 2017 at 10:39 AM  
To: Jim Bochicchio jboch@broadinstitute.org  
Cc: Caroline Cusick ccusick@broadinstitute.org, Aaron M Ellison aellison@fas.harvard.edu



Hi Jim and Caroline, looking forward to getting something on the calendar.

Thanks!

Matt

On May 22, 2017, at 8:25 PM, Jim Bochicchio  
<[jboch@broadinstitute.org](mailto:jboch@broadinstitute.org)> wrote:

Hi Matt

We'll check with Jenn our admin to set up a meeting. She's currently out of the office but should be back on Wednesday.

Thanks  
Jim

On Thu, May 18, 2017 at 4:49 PM, Lau, Matthew K.  
<[matthewklau@fas.harvard.edu](mailto:matthewklau@fas.harvard.edu)> wrote:

Hi Caroline and Jim, I know your analysts are still processing, but I figured it would be great to schedule an in-person meeting sometime within the next few weeks. These meetings can be hard to schedule for us in the summer time, so it's best to get something on the books now.

Does Thursday June 8 in the mid-to-late afternoon work for you? If not, does Monday June 12 (any time) work?

Also, I came across this presentation on Data Sharing at Broad (<http://genomicinfo.broadinstitute.org/acton/attachment/13431/f-0132/1/-/-/-/ASHG16%20Workbench%20Presentation%20-%20Final.pdf?sid=TV2:QKwcswckd>).

Per my request for the code are your analysts starting to use

For my request for the code, are your analysts starting to use these Workbench methods? It seems like it would make this a whole lot easier.

Thanks!

Matt

*Postdoctoral Fellow*  
Harvard Forest  
Harvard University  
324 N Main St.  
Petersham, MA 01366  
[978-756-6165](tel:978-756-6165)

"Knowledge is knowing a tomato is a fruit. Wisdom is not putting it in a fruit salad." — Miles Kingston

On May 17, 2017, at 6:25 AM, Lau, Matthew K.  
<[matthewklau@fas.harvard.edu](mailto:matthewklau@fas.harvard.edu)> wrote:

Hi Caroline, great thanks. Also, please ask your analyst what the ETA is. I would also like to see the code for the pipeline. I had asked for this back in December, but Terry wasn't able to get it after the analysis had already been done.

Best,

Matt

Postdoctoral Research Fellow  
Harvard Forest  
Harvard University  
324 N Main St  
Petersham, MA 01366

Office: [\(978\) 756-6165](tel:978-756-6165)

Knowledge is knowing that a tomato is a fruit, wisdom is not putting it in a fruit salad.

-- Miles Kingston

On May 16, 2017, at 9:17 PM, Caroline Cusick  
<[ccusick@broadinstitute.org](mailto:ccusick@broadinstitute.org)> wrote:

Hi Matt,

Our analyst is going through the comments that you sent from NCBI now, and I will pass along any updates/questions he has during this and definitely keep you updated!

thanks,  
Caroline

On Mon, May 15, 2017 at 8:44 AM, Lau, Matthew K.  
<[matthewklau@fas.harvard.edu](mailto:matthewklau@fas.harvard.edu)> wrote:

Hi Caroline and Jim, happy Monday!

What's the status on the assemblies? And per my last email, are you re-processing from the raw reads?

Cheers,

Matt

On May 12, 2017, at 12:14 PM, Lau, Matthew K.  
<[matthewklau@fas.harvard.edu](mailto:matthewklau@fas.harvard.edu)> wrote:

Great, thanks Caroline!

All 7 were flagged with errors, but only 5 have reports. I've concatenated the errors that I could get from NCBI (see attached). Each report starts with a header like this: "Contamination.txt\_AZXXM\_filtered\_scaffolds.txt". Let me know if you have any questions.

A question though. It seems like this means that there were adapters left in the sequences used for the assemblies. Are you going to go back to the raw sequences, remove them and then re-do the assemblies?

Just in case it's helpful to know now, your FTP server will work fine for the data transfer.

Cheers,

Matt

On May 12, 2017, at 11:58 AM, Caroline Cusick  
<[ccusick@broadinstitute.org](mailto:ccusick@broadinstitute.org)> wrote:

Hi Matt,

We're going to go through and screen the assemblies for adapter, paying specific attention to the sites flagged by NCBI, and then set up a site for that data handoff. Once we get a sense of how long this will take, I can follow up with an ETA.

Our analyst had a question about whether all 7 assemblies were submitted to NCBI, or just this one (SM-AZXXM)? If the other six were returned with errors also, if you could pass that along and he will make sure to confirm that we've removed all the sites flagged by NCBI.

let me know if you have any questions!

thanks,  
Caroline

On Fri, May 12, 2017 at 11:21 AM, Lau, Matthew K.  
<[matthewklau@fas.harvard.edu](mailto:matthewklau@fas.harvard.edu)> wrote:

Morning Caroline, have you had a chance to catch up with your data analysts yet? I have a meeting to discuss the state of this project after lunch and it would be good to have an idea of what you guys are doing on your end.

Thanks,

Matt

Postdoctoral Research Fellow  
Harvard Forest  
Harvard University  
324 N Main St  
Petersham, MA 01366

Office: [\(978\) 756-6165](tel:(978)756-6165)

Knowledge is knowing that a tomato is a fruit, wisdom is not putting it in a fruit salad.

-- Miles Kington

On May 11, 2017, at 4:55 PM, Caroline Cusick  
<[ccusick@broadinstitute.org](mailto:ccusick@broadinstitute.org)> wrote:

Hi Matt,

You are not forgotten! I'm just finalizing a few things with our analyst and I'll have an update for you tomorrow.

thanks!  
Caroline

On Thu, May 11, 2017 at 11:31 AM, Lau, Matthew K.  
<[matthewklau@fas.harvard.edu](mailto:matthewklau@fas.harvard.edu)> wrote:

Hi Caroline, just pinging you again to make sure I don't fall off the radar.

Thanks!

Matt

On May 10, 2017, at 10:13 AM, Lau, Matthew K.  
<[matthewklau@fas.harvard.edu](mailto:matthewklau@fas.harvard.edu)> wrote:

Hi Caroline, please keep me posted. A quick question, if there are primers still in the assemblies, what are the potential issues with the estimates of coverage, GC content, contaminant removal, etc.?

Thanks, Jim, have a good few days out of the office.

Best,

Matt

*Postdoctoral Fellow*  
Harvard Forest  
Harvard University  
324 N Main St.  
Petersham, MA 01366  
[978-756-6165](tel:978-756-6165)

"Knowledge is knowing a tomato is a fruit. Wisdom is not putting it in a fruit salad." — Miles Kingston

On May 9, 2017, at 3:36 PM, Jim Bochicchio  
<[jboch@broadinstitute.org](mailto:jboch@broadinstitute.org)> wrote:

Hi Matt

Our analysts are still looking into this. I'm going to be out of the office for the next few days so Caroline is going to keep an eye on this.

Thanks  
Jim

On Tue, May 9, 2017 at 10:32 AM, Lau, Matthew K.  
<[matthewklau@fas.harvard.edu](mailto:matthewklau@fas.harvard.edu)> wrote:

Hi Jim, just following up on my last couple of emails. Were you able to find out anything about the sequence labeling issue? Also, is there a good solution to removing the primers from the assemblies?

Best,



Matt

*Postdoctoral Fellow*  
Harvard Forest  
Harvard University  
324 N Main St.  
Petersham, MA 01366  
[978-756-6165](tel:978-756-6165)

"Knowledge is knowing a tomato is a fruit. Wisdom is not putting it in a fruit salad." — Miles Kingston

--

Jim Bochicchio  
Sr, Product Manager  
Broad Technology Labs  
Broad Institute  
75 Ames Street  
Cambridge MA 02142, USA

Phone: [617-714-8513](tel:617-714-8513)

Website: <https://www.broadinstitute.org/btl>

--

Caroline Cusick  
Product Coordinator  
Broad Institute of MIT and Harvard  
75 Ames St | Cambridge MA 02142 USA

Caroline Cusick  
Product Coordinator  
Broad Institute of MIT and Harvard  
75 Ames St | Cambridge MA 02142 USA  
<ncbi\_contam\_screen.txt>

--

Caroline Cusick  
Product Coordinator  
Broad Institute of MIT and Harvard  
75 Ames St | Cambridge MA 02142 USA

--

Jim Bochicchio  
Sr, Product Manager  
Broad Technology Labs  
Broad Institute  
75 Ames Street  
Cambridge MA 02142, USA

Phone: 617-714-8513

Website: <https://www.broadinstitute.org/btl>







