Lab Notebook for {blank}

MONDAY, 4/20/2020

Aims:

The aim of this session is to get you used to interacting with sequenced data, simple annotation of a genome, and testing and designing primers. As well as becoming familiar with some essential web-tools for molecular biology & using an electronic lab notebook (ELN).

Below is a checklist of tasks to carry out, our aim is to complete this in small groups and submit this lab page at the end of the session.
You might breeze through all of these tasks in this session, or you may find you only get partway through. Don't worry either way - you will have permanent access to this resource, so you can always come back and complete this later.

- Get stuck? No problem, you have your classmates and me to help you through - if you get really stuck there is a link here to the video tutorials for each task - but try to have a go on your own first.

- MESS UP REAL BAD? That's ok with Benchling you are a master of time! The clock on the right hand toolbar let's you turn back to any previous version of the file - that accidental deletion never happened!

Note: Where possible include hyperlinks or attachments of files in the Notebook (see Benchling tutorials). This adds reproducibility to your notes.

Task One:

- [ ] Set up your own Project & copy this notebook and all of the contents from the Inventory section of COVID-19 to your Project.
  - All files should now be editable by you & you now have permanent access to this data - check in permissions that your sequences are now unlocked for editing.
  - Make your notebook readable only to you e.g. private to thr 4013Y group - this will prevent it clogging up the menu space with >200 identical projects!!!

Task Two:

- [ ] Identify the reference genome for SARS-COV2.

In the SARS-COV-2 folder in the inventory, you should find that there are five annotated genomes. Can you use the metadata & descriptions for these files to identify which of these is the reference genome (assembled as a representative example of a species' set of genes)?

Q1. What is the accession number for the reference genome? :
Include a hyperlink (use @ - the type your filename or drag and drop from the left hand tool bar to this location :
https://benchling.com/tutorials/43/referencing-files
Q2. Is there anything different about the amount of annotation of the sequence map for the reference genome compared to the others?

Q3. Using the reference genome as the template - set up an alignment with any one other SARS-COV-2 genome and complete the table below:

Alignment info

	Reference genome	Other genome	Number of nt subsitutions between the genomes
1			
2			

Task Three:

- [ ] Annotate your reference genome copy with the 29 protein coding genes. (Start with ORF1ab - and complete at least five annotations)

https://benchling.com/tutorials/32/creating-annotations-on-sequences
Using the link or attached pdf read up on the different proteins identified.
Inside the Coronavirus Genome

Inside the Coronavirus Genome.pdf

These genes should already be annotated on your reference genome - though they won't have been labelled e.g. NSP1-10. Not very useful - can you edit the annotations to give them their proper names?

- Check that the sequence of each annotation is correct (you can copy the sequences directly from the PDF and CTRL-F on your genome).
- Try and edit the annotations to update them with the correct names.

Task Four:

Translate your genome

https://benchling.com/tutorials/32/creating-annotations-on-sequences

- Time to have a go at translating your genes into proteins! Find the Spike protein gene (S) on your reference genome, and translate your gene sequence into a protein sequence (Right-click on the gene and look at the options). Remember don't just rely on the annotation - your sequence should start with an ATG and end with a TAA sequence, so that your protein will start with an M(ethionine) and end with a *

Q4. What is the protein sequence for S here:

Q5. Refer back to Inside the Coronavirus Genome - what is the **protein sequence** for the insertion of 12 genetic letters that may help the spikes bind more tightly to human cells.

PRRA

Bonus task - complete this now or come back to it later:

In the folder SARS-COV you will find the reference genome of the SARS virus from the outbreak in 2003 - try and run an alignment between SARS-COV and SARS COV-2. You should find they are quite different to each other at the nucleotide level.

Now for something a bit more advanced - don't worry if you get stuck: Save your copy of the Spike protein from SARS-COV-2 as an amino acid file - copy translation/ then go to the create button top-left/ AA sequence and CTRL+V (ignore any error message).
Q5b. Repeat this for the Spike protein of SARS-COV. Are the amino acid sequences more or less divergent than the nucleotide sequences?
(Note - using the @ function you can add files from your inventory here!)
Q5c. Has the amino acid sequence equally diverged across the length of the protein? Can you think of a reason for what you have observed? Have a look for the where the protein sequence diverges at the at that 12 genetic letter insertion. What has that done to the protein sequence at this point?

Task Five:

Testing Primers

On the reference genome, you should find sets of primers already attached to the sequence file. These primers are also located in two folders EU primer sets (2 in total) and CDC primer sets (3 in total) in the Inventory section. If you navigate to the primers section on the right-hand side of the genome sequence you will find a tab marked pairs - here you can see each forward and reverse primer pair.

Q6. Using the information on that section please complete this table for each forward and reverse primer.

Primers summary

	Developed by	Primer Name	Binds to positive or negative sense strand?	Binding location	Melting temperature
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

Now let's look at the PCR products a primer pair should produce -
Here is a checklist for "good" primer design - let's investigate these designed primers

For the Primer pair - Forward and Reverse

- [ ] **PCR product size:** 70 – 200 base pairs.
- [ ] **Primer length:** 18 – 22 nucleotides (optimal: 20).
- [ ] **GC% content:** 50 – 60%. Primers should have similar GC%.
- [ ] **Contain a GC clamp:** A GC clamp is having either a G or a C base in the last 5 bases of the primer. Since G and C bases have a stronger binding affinity, this will ensure the 3' end of the primer anneals correctly to the cDNA sequence.
- [ ] **Tm:** 59 – 65^{o}C (optimal: 62^{o}C). Primers should have similar Tm that are within 5^{o}C of each other.
- [ ] **Avoid nucleotide repeats:** Avoid repeats of 4 or more adjacent repeats (e.g. GTTTT) and/or dinucleotide repeats (e.g. GCGCGCGC) completely when using oligo(dT)s, therefore by having primers that bind to the 3' end the efficiency should improve.
- [ ] **Low self-complementarity score ΔG:** Self-complementarity is the likelihood that the primer will bind to itself and to the other primer in the pair rather than the template DNA. - a strong delta G (-9kcal/mol or more negative) indicates this oligo could be problematic

For the oligo probe

- [ ] **Location:** Ideally, the probe should be in close proximity to the forward or reverse primer, but should not overlap with a primer-binding site on the same strand. Probes can be designed to bind to either strand of the target.
- [ ] **Melting temperature (T_m):** Preferably, probes should have a T_m 6–8°C higher than the primers. If the melting temperature is too low, the percentage of probe bound to target will be low. In this case, the primers may amplify a product, but sensitivity may be compromised as all target sites are not saturated with probe resulting in reduced fluorescence signal that does not truly represent the true amount of target present in the sample.
- [ ] **Annealing temperature (Ta):** The annealing temperature should be set no more than 5°C below the lower primer Tm. Use this as a general guideline, but note that optimization may still be necessary.
- [ ] **GC content:** As with primer sequences, aim for a GC content of 35–65% and avoid a G at the 5' end to prevent quenching of the 5' fluorophore.

Q7. Can you write a brief summary for each of the five primer pairs designed for RT-qPCR testing - how many of the criteria does each Primer pair and probe fit?
N1.
N2.
N3.
E_Sarbeco
RdRP_SARS

Task Six:

- [ ] Primer BLAST - try the primer sets in Primer BLAST

1. Go to the Primer BLAST submission form. (use hyperlink)
2. Enter both forward and reverse primer sequences in the Primer Parameters section of the form. Enter the accession number of the SARS-COV-2 reference genome (see Q1).
3. In the Primer Pair Specificity Checking Parameters section, **make sure this is blank (it likes to default to Human)** so that it will check all reference genomes for you. For broadest coverage, choose the nr database and do not specify an organism.
4. Click the "Get Primers" button to submit the search and retrieve template and specificity information.

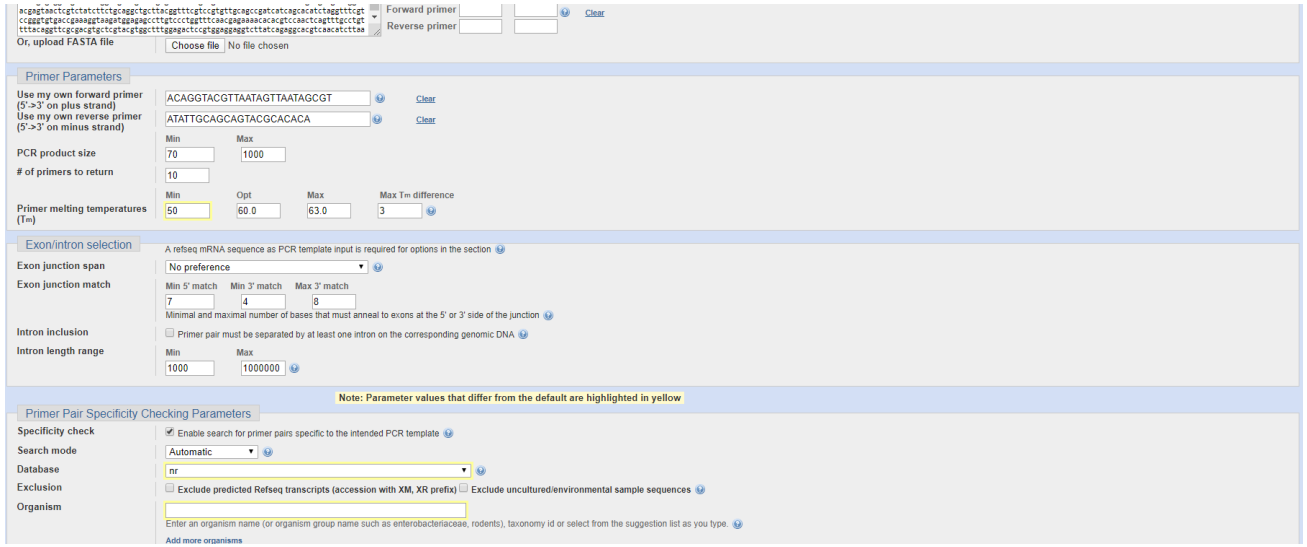
Q8. How specific to SARS-COV-2 are the different primer sets according to Primer BLAST? - scroll through the list of genome hits and broadly summarise

N.B. The primer sets N1-N3 designed by the CDC are supposed to be explicitly designed to be unique to the SARS-COV-2 virus
The EU primer sets RdRP and E are designed to detect a broader range of SARS related viruses - with only the oligo probe of the RdRP panel designed to be unique to SARS-COV-2.

N1.
N2.
N3.
E_Sarbeco
RdRP_SARS

There may be some genomes labelled as Bat Coronavirus (the species SARS-COV-2 might have jumped from) or isolated from the Wuhan Seafood market (likely outbreak point). How would you go about checking whether these were examples of SARS-COV-2?

image.png



Task Seven:

- [ ] Can you do a better job than the CDC or WHO? 👍

Good primer design does not guarantee that your reaction will work - but it can definitely help!
Repeat the Primer BLAST step as before, but this time leave the Primer boxes empty - Primer 3 is going to do the hardwork for you.

MAKE SURE THE PARAMETERS FOR TESTING ARE SET RIGHT!!! For example PCR product size should be **<200bp.**

- [ ] What are your new Forward and Reverse primers?
- [ ] Can you find them in the reference genome and attach them to the sequence?
- [ ] Can you include hyperlinks use (@) to add links from files here!

Q8. Complete a table and summary for your new primer set as you did for Task Five

Q9. How does your primer pair compare to the published primer sets?

THAT'S IT! CONGRATULATIONS!!!!!!!
Before you go please do the following:

1. **Go to Print - top left hand corner to print your lab notebook - and send it to me p.leftwich (at) uea.ac.uk so I can see how you got on!!!**