iChips, nanopores and virtual collaboration spaces

Prof JAMES CHONG
Royal Society Industry Fellow
Department of Biology

james.chong@york.ac.uk
@insanity_one
https://goo.gl/c17LxT

Belfast 7 April 2019
# Outline

## 1. GENOME ASSEMBLY & ANNOTATION

1. SOIL SAMPLE
   - a. PLATING SOIL AND ICHIP WELLS
   - c. DNA EXTRACTION
   - d. 16S rRNA PCR
   - e. AMPLICON SEQUENCING

2. IDENTIFY CLEARING ZONES
   - b. OVERLAY WITH ESKAPE PROXIES
   - REPICA PLATING

3. DATA ANALYSIS WITH QIIME
   - a. FINAL REPORT

## 2. FURTHER BIOINFORMATICS PROJECTS

10^8-10^9 cells per g soil
= 2x10^3-8x10^5 species
The great plate count anomaly

This approach used to identify teixobactin

Growing the ungrowable
iChips

Field work in a field...
Student data

Student data

Nanopore sequencing

512 nanopore channels
10-20 Gb / 48 hours
Nanopore sequencing

19 contigs (7 > 200kb)

1.4 Mb contig!
## Testing MinION

### MicrobesNG (Illumina)

<table>
<thead>
<tr>
<th>Sample id</th>
<th># contigs</th>
<th>Total length</th>
<th>Largest contig</th>
<th>GC%</th>
<th>Most frequent species</th>
</tr>
</thead>
<tbody>
<tr>
<td>3_PL</td>
<td>70 (68)</td>
<td>5,917,814</td>
<td>605,897</td>
<td>35.2</td>
<td><em>Bacillus weihenstephanensis</em></td>
</tr>
<tr>
<td>6_PL</td>
<td>106 (81)</td>
<td>5,079,274</td>
<td>467,491</td>
<td>47.2</td>
<td><em>Paenibacillus terrae</em></td>
</tr>
<tr>
<td>8_PL</td>
<td>51 (33)</td>
<td>6,139,014</td>
<td>1,541,288</td>
<td>45.6</td>
<td><em>Paenibacillus polymyxa</em></td>
</tr>
<tr>
<td>21_PL</td>
<td>907 (821)</td>
<td>8,773,359</td>
<td>117,348</td>
<td>72.0</td>
<td><em>Streptomyces venezuelae</em></td>
</tr>
<tr>
<td>M4_SW*</td>
<td>21 (17)</td>
<td>3,823,503</td>
<td>1,902,398</td>
<td>41.2</td>
<td><em>Bacillus pumilus</em></td>
</tr>
</tbody>
</table>

Pool these species (5:1:2) and run on MinION.
# Testing MinION

MicrobesNG (Illumina)

<table>
<thead>
<tr>
<th>Sample id</th>
<th># contigs</th>
<th>Total length</th>
<th>Largest contig</th>
<th>GC%</th>
<th>Most frequent species</th>
</tr>
</thead>
<tbody>
<tr>
<td>3_PL</td>
<td>70 (68)</td>
<td>5,917,814</td>
<td>605,897</td>
<td>35.2</td>
<td><em>Bacillus weihenstephanensis</em></td>
</tr>
<tr>
<td>6_PL</td>
<td>106 (81)</td>
<td>5,079,274</td>
<td>467,491</td>
<td>47.2</td>
<td><em>Paenibacillus terrae</em></td>
</tr>
<tr>
<td>21_PL</td>
<td>907 (821)</td>
<td>8,773,359</td>
<td>117,348</td>
<td>72.0</td>
<td><em>Streptomyces venezuelae</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample id</th>
<th># contigs</th>
<th>GC%</th>
<th>Gemeente species</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>75</td>
<td>35.2</td>
<td><em>Bacillus weihenstephanensis</em></td>
</tr>
<tr>
<td>Y</td>
<td></td>
<td>47.2</td>
<td><em>Paenibacillus terrae</em></td>
</tr>
<tr>
<td>Z</td>
<td>GC% 50</td>
<td>71.8</td>
<td><em>Streptomyces venezuelae</em></td>
</tr>
</tbody>
</table>

6.4 Gb
Biology projects @ York
• 2 term projects
• Supervised
• Lab or non-lab based
• Expected 2.5 days of work per week
• Students learn new skills
• Work can be collaborative (but this is rare and difficult to encourage)
• Individual independent report for assessment

Bioinformatics downsides
• Limited / no experience with technical aspects and bioinformatics tools
• Students are trained in underlying biology, but not bioinformatics specifically
• Command-line programming challenging for students with little / no prior training

Skills learning...
• Tends to be one-to-one
• Timetabling of meetings / supervision can be a challenge
• Little / no benefit to / from peers
• Propagation of solutions is slow
A virtual collaboration space

- Google Team Drive
  - Shared digital and updatable protocol
  - Easy to use

- Slack

Both tools are platform agnostic
Genome assembly & annotation

Lassopeptide
Slack-ers

#undergrads

Friday, February 9th

11:42 AM
Don't worry it's been solved! Downloaded a program called WinSCP to move files across and it actually worked.

8:03 PM
uploaded and commented on this image: circos.png

64 Get Antismash clusters on to the circos graph - can't get rid of the weird grey lines yet though (or label them)

8:27 PM
That looks awesome!

insanity_one 8:27 PM
Well done! That's really impressive!

8:28 PM
It does look awesome. What's happening at 1.5 Mb?

7:31 PM
I'm in nano text editor for my megahit assembly and when I try to save my file it says 'Error writing: ' permission denied: permission?

7:34 PM
Just to double check, have you written the nano script and saved it? Are you editing one you've received?

7:37 PM
Insanity_one 7:37 PM
and which file are you in? ie. what does your nano command look like when you're not in nano?

7:38 PM
I've written it from scratch but adapted it from the 'Bioinformatics project student guidelines' with minor modifications. I'm in /c2d2/groups/chong/undergrads/2017seq.

7:40 PM
Ahh you probably don't have permission to save it in the same directory as the one you've already saved. You'll need to move into your directory and then save

insanity_one 7:40 PM
Yup. cd - to get to your own directory then write: 'cd -'

7:42 PM
You'll also need to copy the sequence you want to plot to that directory too. You can probably write your script to copy the sequences to the directories, but it might be easier to do it the first time.

7:51 PM
Ah perfect thank you, just submitted it

7:51 PM
Insanity_one
I've shown you how to point to the sequence files in your directories so that you don't need to copy the file. Having more space on the server unnecessarily, so try not to avoid it.

7:51 PM
Great. Let us know how it goes...

Nov 27th, 2018 at 1:29 PM
Is this the way I'm supposed to be organizing these figures, before I go on to make all of them?

Kmers_AAAA.png

2 replies

2 months ago
Looks good to me. @insanity_one?

2 months ago
Insanity_one
Awesome. Yes, happy with this!
Improving student confidence

Student Confidence with Various Aspects of the Project

- use Unix
- use BI tools
- use Slack
- annotate a genome
- be in a peer-learning group
Feedback

‘We all came to the point where we all had the same problem.’

‘We were working as a team; although we have different projects, we have the same problems and were solving them as a team.’

‘We were encouraged that there were no silly questions. The [Slack] undergraduate channel was for low-level questions. I had a lot of questions!’
Conclusions

Benefits

- Developing transferable skills
- Less reliance on instructor-led learning
- More ownership of the project and independent learning
- Peer learning
- Critical thinking
- Reduces working hours time pressure on instructors and allows students to make progress out of hours

Disadvantages

- Developing the student's confidence in using the medium takes a little time
- Raised expectations of immediate and frequent contact
Thanks

Anna Alessi  Richard Walker  Peter Ashton
Kim Barnes  Setareh Chong  Sally James
Annabel Cansdale  Kelly Redeker  John Davey
Katherine Newling  Emma Barnes

Maggie Smith

UoY / Department of Biology Undergraduates