In-house full plasmid sequencing to enable rapid synthetic biology

Daniel Giguere

PhD Candidate, Department of Biochemistry

Slides: https://github.com/dgiguer/lab_resources/presentations

github.com/dgiguer 

@DanielJGiguere 

We are using *Phaeodactylum tricornutum* to produce high-need proteins

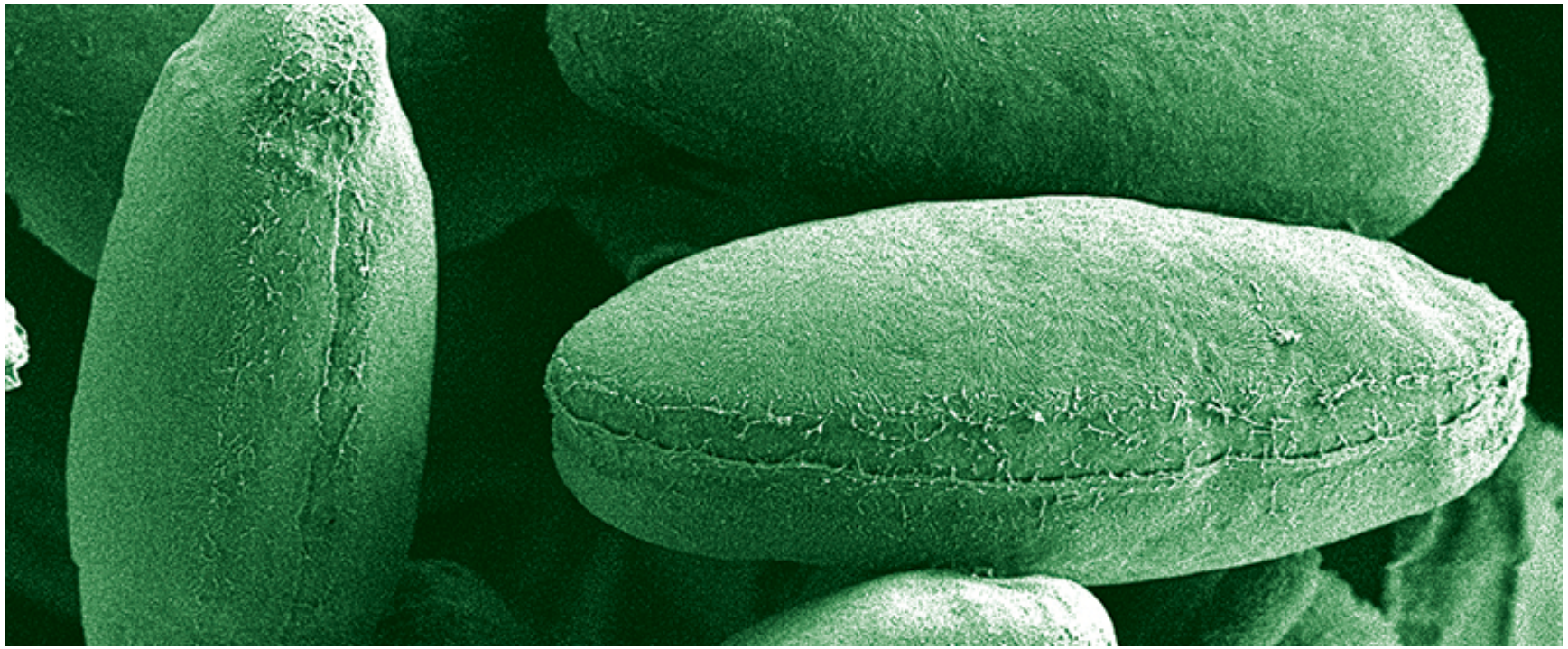


Image taken by Thomas J. Deerinck, National Center for Microscopy and Imaging Research, University of California, San Diego

We are using *Phaeodactylum tricornutum* to produce high-need proteins

- Spike protein
- Nucleocapsid
- Different subunits and stabilizing mutations

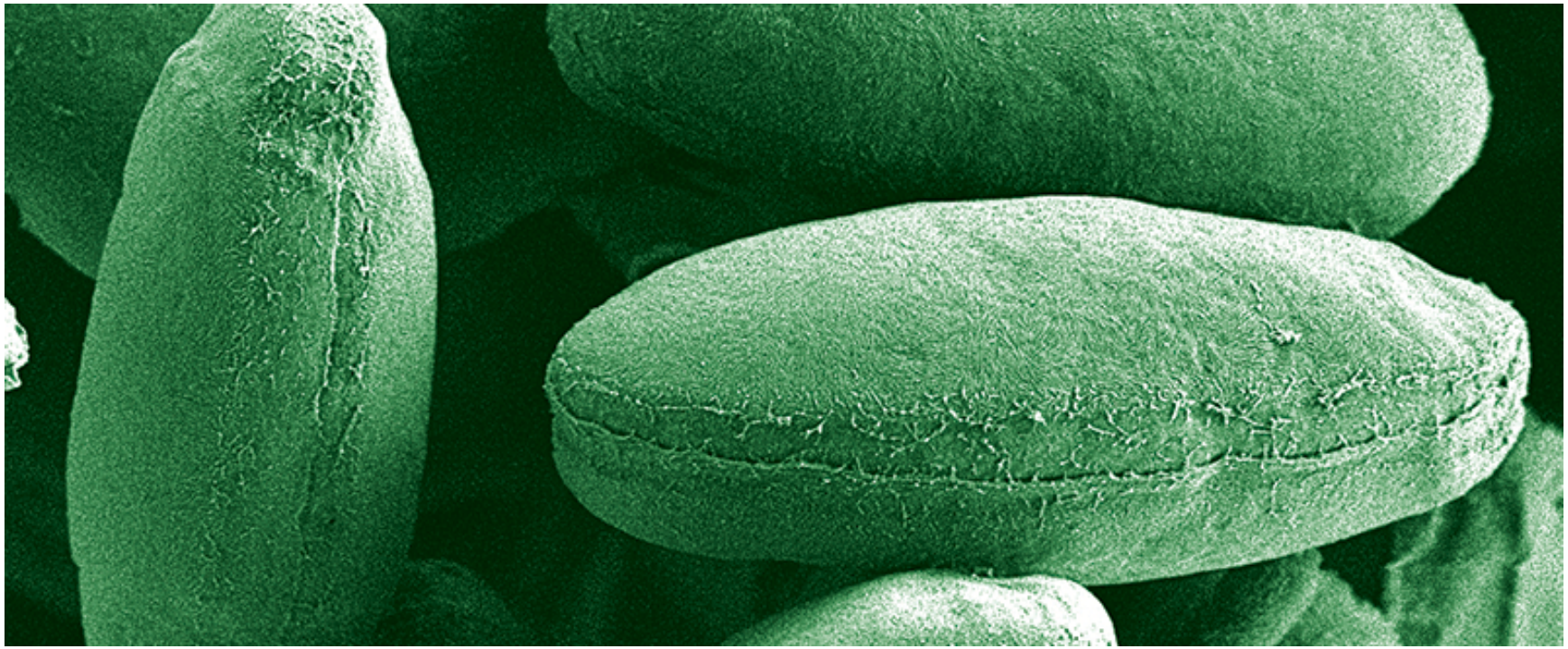
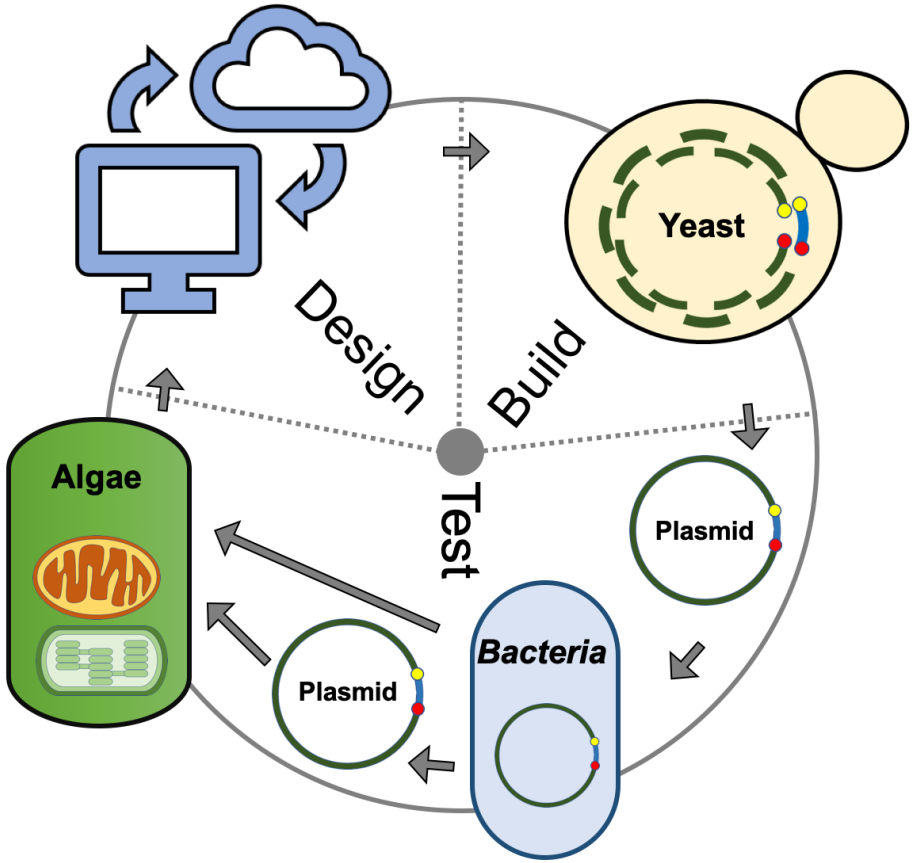


Image taken by Thomas J. Deerinck, National Center for Microscopy and Imaging Research, University of California, San Diego

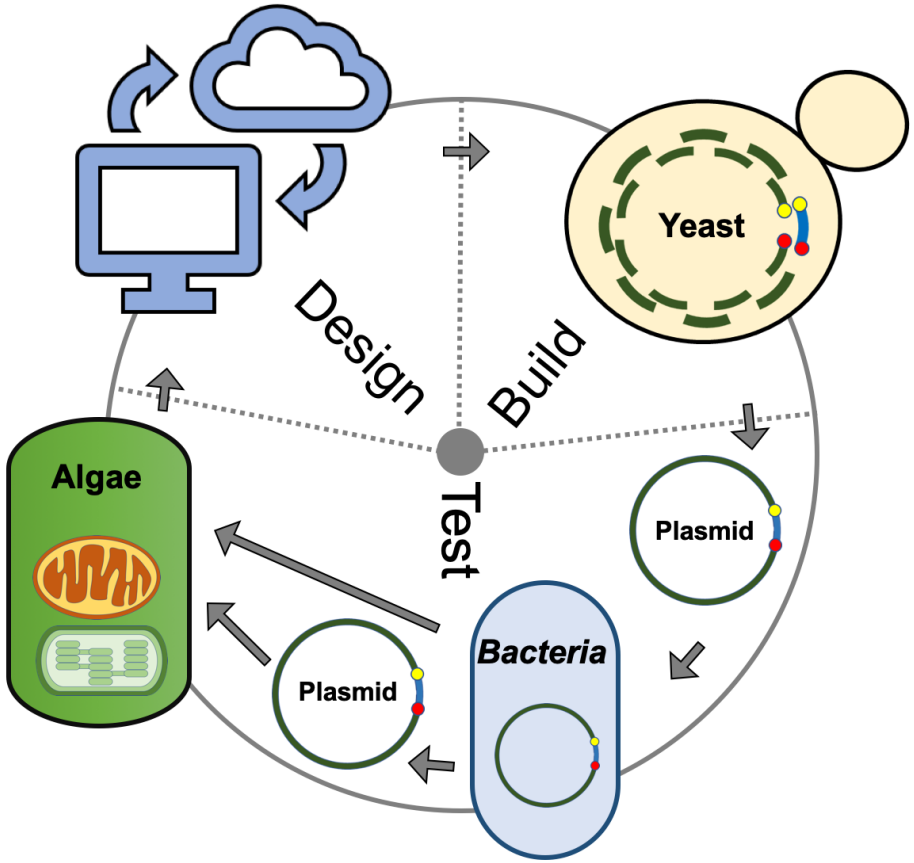
Synthetic biology requires rapid validation



Modified from Cochrane et al., 2019 (bioRxiv) with permission

- Ensure the plasmids don't have mutations before transformation in P. *tricornutum*

Synthetic biology requires rapid validation



Modified from Cochrane et al., 2019 (bioRxiv) with permission

- Ensure the plasmids don't have mutations before transformation in P. *tricornutum*
- Speed is highly important for this platform for SARS-CoV-2 proteins

Plasmid sequencing on Illumina is too slow

- Pre-COVID, validation done at MGH DNA core (Massachusetts)
  - CAD$100 per plasmid + international shipping
  - 2-3 weeks from plasmid DNA to results

Plasmid sequencing on Illumina is too slow

- Pre-COVID, validation done at MGH DNA core (Massachusetts)
  - CAD$100 per plasmid + international shipping
  - 2-3 weeks from plasmid DNA to results
- Local MiSeq run (lrgc.ca):
  - $170 per sample
  - > $1720 sequencing kit

Plasmid sequencing on Illumina is too slow

- Pre-COVID, validation done at MGH DNA core (Massachusetts)
  - CAD$100 per plasmid + international shipping
  - 2-3 weeks from plasmid DNA to results
- Local MiSeq run (lrgc.ca):
  - $170 per sample
  - > $1720 sequencing kit
- Initial shutdown caused even further delays and uncertainty

Research objectives

- Design a cost-effective in-house sequencing pipeline for plasmids

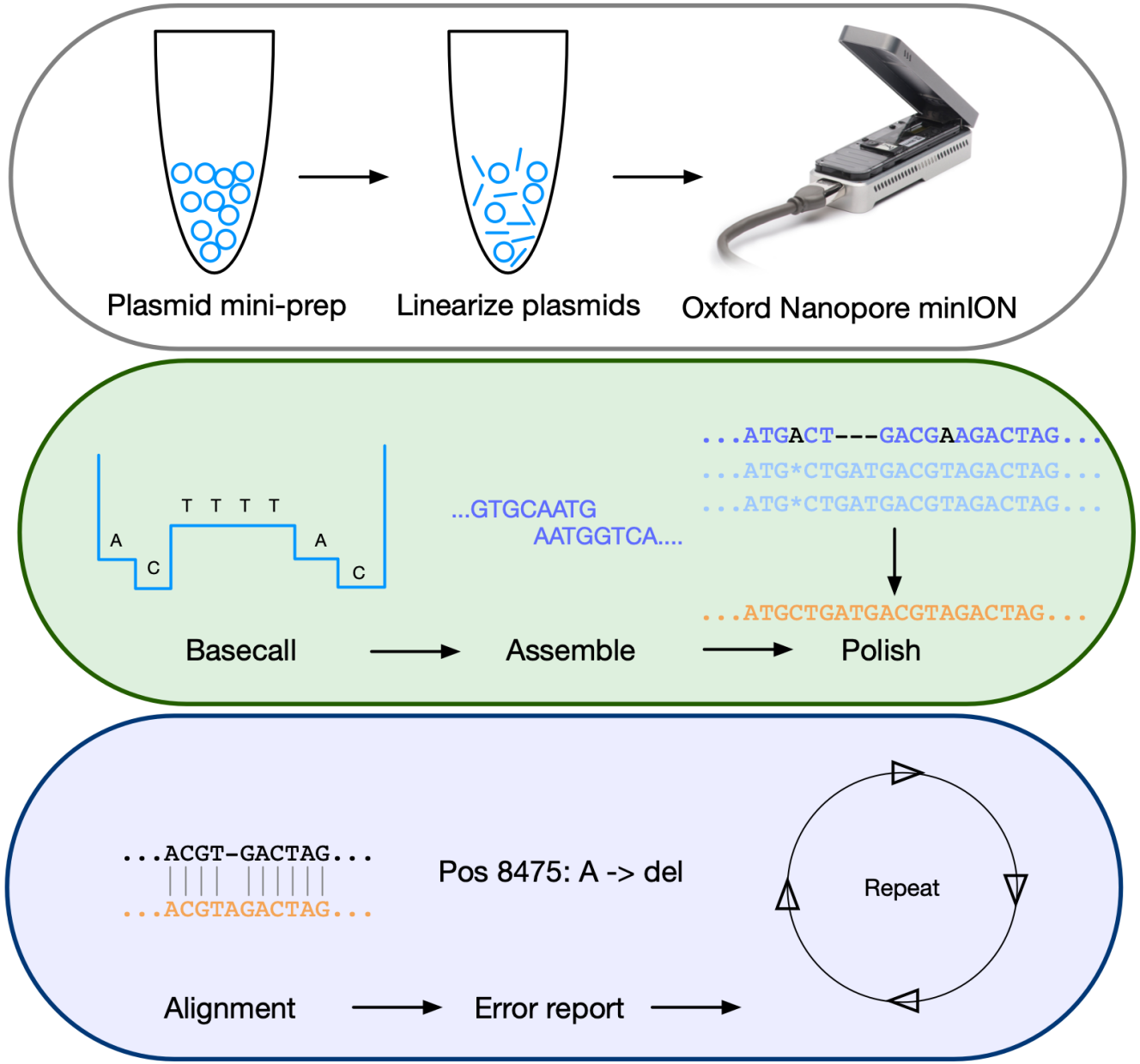
Research objectives

- Design a cost-effective in-house sequencing pipeline for plasmids
- Develop the bioinformatic infrastructure to rapidly analyze the data

Research objectives

- Design a cost-effective in-house sequencing pipeline for plasmids
- Develop the bioinformatic infrastructure to rapidly analyze the data
- Evaluate quality compared to Illumina sequencing

Workflow


Plasmid mini-prep
Linearize plasmids
Oxford Nanopore minION
T T T T
A
C
A
C
...GTGCAATG
AATGGTCA....
...ATGACT---GACGAAGACTAG...
...ATG*CTGATGACGTAGACTAG...
...ATG*CTGATGACGTAGACTAG...
...ATGCTGATGACGTAGACTAG...
Basecall
Assemble
Polish
...ACGT-GACTAG...
...ACGTAGACTAG...
Pos 8475: A -> del
Repeat
Alignment
Error report

What does the raw data actually look like?

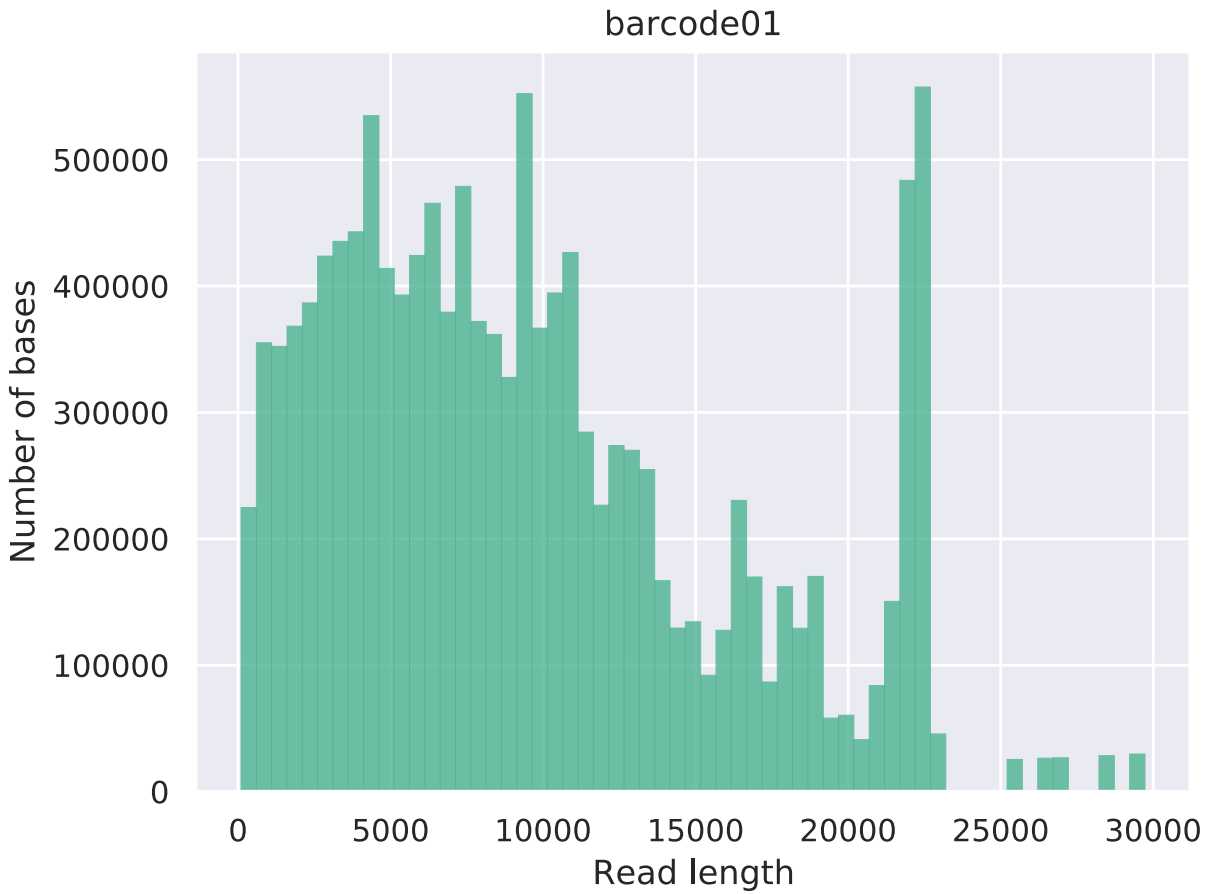
We obtain mostly full length plasmid reads

Version 1: 12 plasmids

We obtain mostly full length plasmid reads

Version 1: 12 plasmids

Version 2: 12 plasmids

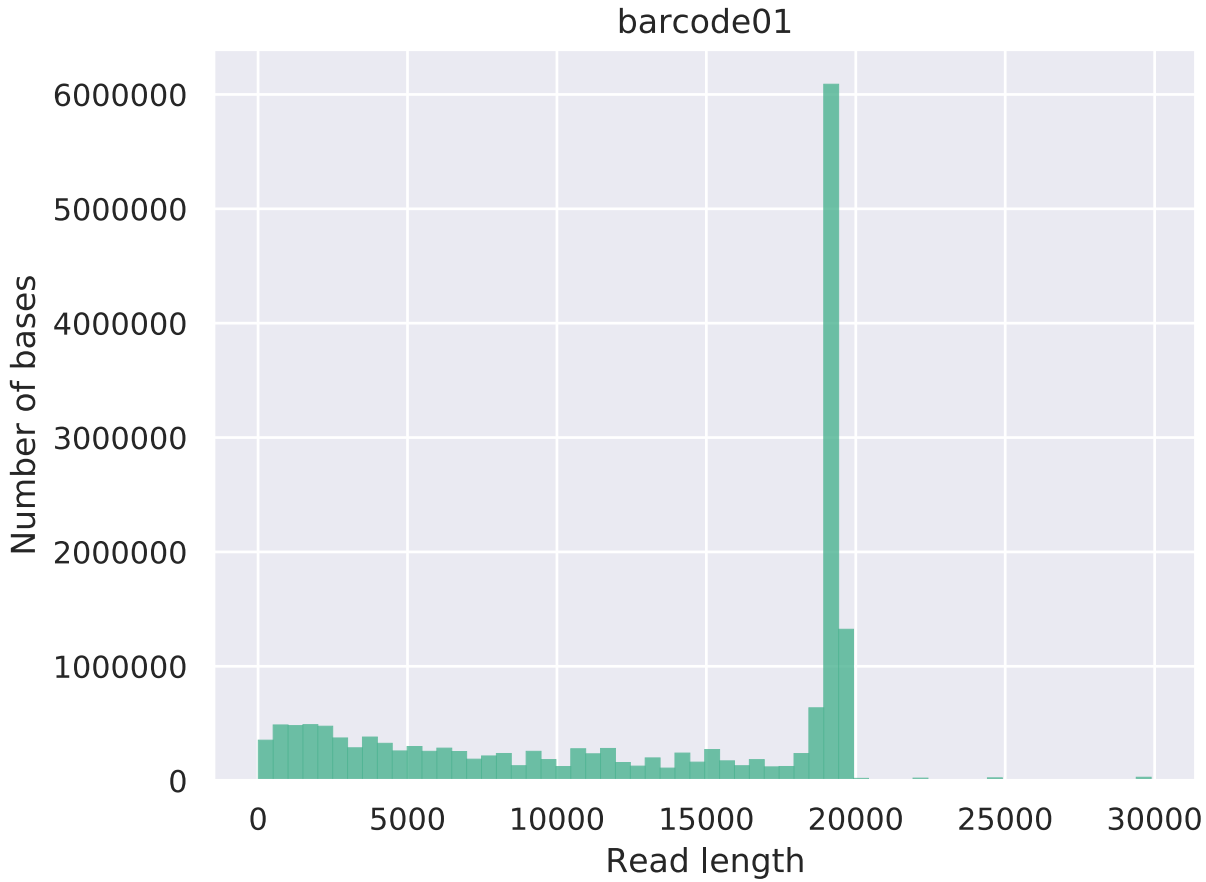


N50: 10,470.0

Version 1: single plasmid

N50: 10,470.0

Version 1: single plasmid



N50: 16,682.0

Version 2: single plasmid

How rapid is the data analysis?

- In initial stages, analysis would take an entire day for 12 plasmids

How rapid is the data analysis?

- In initial stages, analysis would take an entire day for 12 plasmids

- Now:

```
Start of analysis: Thu Dec  3 09:02:57 EST 2020
End of analysis: Thu Dec  3 09:13:57 EST 2020
```

How rapid is the data analysis?

- In initial stages, analysis would take an entire day for 12 plasmids

- Now:

```
Start of analysis: Thu Dec  3 09:02:57 EST 2020
End of analysis: Thu Dec  3 09:13:57 EST 2020
```

  - Less than 1 minute of analysis per plasmid
  - *Caveat*: you need lots of RAM and a high-end graphics processing unit!

Aren’t Nanopore reads poor quality?

Aren’t Nanopore reads poor quality?

- Raw read Q-score ~ 11 = 92% accuracy

Aren’t Nanopore reads poor quality?

- Raw read Q-score ~ 11 = 92% accuracy
- Final Q-score of assembly after pipeline for small plasmids > 40

	Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	thomas_1		11784	19802	100%	0.0	100.00%	10726	Query_11729

	Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	thomas_2		17078	19803	100%	0.0	99.99%	10727	Query_4733

Aren’t Nanopore reads poor quality?

- Raw read Q-score ~ 11 = 92% accuracy
- Final Q-score of assembly after pipeline for small plasmids > 40

- 1 homopolymer error in 21454 bases between two plasmids

	Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	thomas_2		17078	19803	100%	0.0	99.99%	10727	Query_4733

Thanks Thomas and Dalton!

Aren’t Nanopore reads poor quality?

- Similar accuracy in larger circular assemblies of 99.96% = Q34
- All at large homopolymers

	Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	NC_008588.1 Phaeodactylum tricornutum chloroplast, complete genome		1.343e+05	2.423e+05	100%	0.0	99.96%	117369	Query_11943

Thanks Emma!

Applying this platform for plasmid engineering for producing SARS-CoV-2 proteins

- 133 plasmids over 16 runs to date

Applying this platform for plasmid engineering for producing SARS-CoV-2 proteins

- 133 plasmids over 16 runs to date

- Deletion of His-tags from protein
- Multiple SNPs at multiple cloning site
- SNPs outside coding regions (common)
- SNPs inside promoters
- Correcting the reference sequence

Cost

- MGH costs CAD$100 per plasmid (+ shipping)

Cost

- MGH costs CAD$100 per plasmid (+ shipping)

- Average cost of CAD$97 per plasmid
  - Includes flow cell, third party reagents, library prep kits, flow cell wash

Conclusions

- We developed and validated in-house full plasmid sequencing using the Oxford Nanopore platform for a competitive cost
- In-house sequencing enables faster iterations of design, build, test
  - Library prep and sequencing takes up to 6 hours
  - Bioinformatics pipeline takes less than 1 hour
- Final assembly Q-score of 40 or higher for 2 Illumina-sequenced plasmids

Acknowledgements

Protein production
Sam Slattery
Arina Shrestha
Emily Stuckless
Greg Gloor
David Edgell
Bogumil Karas
Martin Flatley (Suncor)
Others…

Initial plasmid sequencing
Thomas Hamilton
Dalton Ham

Want to quickly validate your plasmids?
dgiguer@uwo.ca


github.com/dgiguer


@DanielJGiguere

Slides: https://github.com/dgiguer/lab_resources/presentations