Optimizing High Molecular Weight gDNA Extraction for Single Molecule Sequencing in the Redwood Genome Project

Rachael Workman
Timp Lab, Johns Hopkins University, Department of Biomedical Engineering
Two species in our redwood genome project

California endemics

Economic, cultural, and conservation value

"Advanced management strategies"

**Sequoia sempervirens**  
Coast redwood  
California – central/north coast

**Sequoiadendron giganteum**  
Giant sequoia  
California – Sierra Nevada
Giant sequoia
*Sequoiadendron giganteum*
*Cupressaceae*

2n = 2x = 18
9 gigabase genome

Occur in groves throughout Sierra Nevada

Slides courtesy of Neale Lab, UCD
Coast redwood
Sequoia sempervirens
Cupressaceae

2n = 6x = 60
30 gigabase genome

Restricted to coastal fog belt
Climbing & collection by Steve Sillett + team

DNA extraction
Neale lab

Short read sequencing @ UCD

Long read sequencing @ JHU

Reference genome

Assembly @ JHU

DNA extraction
Timp lab

Short read sequencing @ UCD

Long read sequencing @ JHU

Climbing & collection by Steve Sillett + team

DNA extraction
Timp lab
Nanopore Single Molecule Sequencing

No theoretical upper limit to sequencing read length, practical limit only in preparing long fragment libraries and delivering DNA to the pore intact

Typical user-reported sequencing output 5-15Gb (as of R9.4.1, March 2018)

Disclosure: Timp has two patents (US 2011/0226623 A1; US2012/0040343 A1) licensed to ONT
# Sample requirements for sequencing

<table>
<thead>
<tr>
<th>HMW</th>
<th>Yield</th>
<th>Quality</th>
<th>Reproducibility</th>
<th>Sequencing Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>100kb+ average</td>
<td>&gt;10ug gDNA From 1g tissue</td>
<td>Nanodrop, gel migration in range</td>
<td>Want no Wizards</td>
<td>&gt;5Gb per run</td>
</tr>
</tbody>
</table>
Sample realities before optimization

<table>
<thead>
<tr>
<th>LMW</th>
<th>Low Yield</th>
<th>Poor Quality</th>
<th>Inconsistent</th>
<th>Low Seq Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10kb</td>
<td>&lt;1ug gDNA</td>
<td>Residual polyphenolics</td>
<td>Results varied largely</td>
<td>&lt;1Gb per run</td>
</tr>
<tr>
<td>average</td>
<td>From 1g tissue</td>
<td>And polysaccharides</td>
<td>By sample</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Trials: DNA extraction

Extract Nuclei?
- Yes
  - KIT
  - CTAB
  - Lysis (KIT, CTAB, SDS, SLS, Triton)
  - Agarose embedding?
    - Yes
      - Phenol Chloroform?
        - Yes
          - Alcohol precipitation (NaCl, NaC2H3O2, NH4C2H3O2, None)
        - No
          - Nanobind
    - No
      - Alcohol precipitation (SCODA, Boreal genomics)

- No
  - KIT
  - CTAB
  - Phenol Chloroform?
    - Yes
      - Alcohol precipitation (NaCl, NaC2H3O2, NH4C2H3O2, None)
    - No
      - Alcohol precipitation (SCODA, Boreal genomics)
Top 3 extraction protocols

ZHANG

HEALEY

NANOBIND
Top 3 extraction protocols: Zhang

Tissue ground in LN2

Nuclei: Cell wall lysis -> filtration -> differential centrifugation

DNA: Overnight SLS lysis, phenol chloroform, alcohol precipitation + sodium acetate

Top 3 extraction protocols: Healey

Preferred plant protocol: Skips the differential nuclei extraction

CTAB (cetyl trimethylammonium bromide) – cationic detergent

CTAB extraction, chloroform, alcohol precipitation + NaCl, elute

Healey et al, Plant Methods (2014)
Top 3 extraction protocols: Nanobind

Nanobind: Nuclei isolation as previously described
Nanobind extraction with modified CTAB, EtOH washes, elute
45 minutes total time (after nuclei extract)
Nanostructured silica ($\text{SiO}_2$), essentially.

With or without Iron layer for magnetism
Tentacle Binding Mechanism
Enhances binding capacity and protects DNA from shear forces

- Three material properties needed: low shear, non-porous, high surface area
- DNA tentacles form and extend from substrate to get high binding capacity
- Low shear unlike beads and columns

Top 3 extraction protocols: gDNA Yield/Qual

<table>
<thead>
<tr>
<th>From 1g input</th>
<th>Zhang</th>
<th>Healey</th>
<th>Nanobind</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yield (ug)</td>
<td>2</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>260/280</td>
<td>1.65</td>
<td>1.68</td>
<td>1.76</td>
</tr>
<tr>
<td>260/230</td>
<td>0.34</td>
<td>0.76</td>
<td>1.51</td>
</tr>
</tbody>
</table>

Modified CTAB protocols (Nanobind and Healey) produce highest yield and Nanobind extraction produces the best quality extract with long fragment length.

Nanodrop spectra not full picture of quality – plant samples sometimes retain visible color
Extraction Comparison – Yield

Though same input concentration, nanobind showed much higher pore occupancy, and resulting higher yield than Zhang or Healey.

Nanopore sequencing chemistry: LSK108, R9.4
Extraction Comparison: Read length

<table>
<thead>
<tr>
<th></th>
<th>Zhang</th>
<th>Healey</th>
<th>Nanobind</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reads</td>
<td>201k</td>
<td>195k</td>
<td>500k</td>
</tr>
<tr>
<td>Yield</td>
<td>0.51Gb</td>
<td>1.08Gb</td>
<td>4.72Gb</td>
</tr>
<tr>
<td>N50</td>
<td>7.1kb</td>
<td>8.6kb</td>
<td>12.3kb</td>
</tr>
<tr>
<td>Median</td>
<td>929</td>
<td>5.1kb</td>
<td>9.8kb</td>
</tr>
</tbody>
</table>

Nanobind and Healey seemed to give reasonable read lengths, but doesn’t match with PFGE profile size.

Nanobind sequencing yield increased by 10-fold over Zhang, 5-fold over Healey.

Nanopore sequencing chemistry: LSK108, R9.4
Improving read lengths: Shear Comparison

Currently a trade-off between long reads and high yield

Read length limitations a function of library prep? DNA delivery to pore?

Nanopore sequencing chemistry: LSK108, R9.4
Improving Read Lengths: Size selection

Nanobind as a size selection alternative improves median read length but has lower N50 than Blue pippin size select.

Sheared gDNA gives lower N50, but more reads >5kb. Sequencing approach then depends on your study question.

Is N50 best metric of sequencing success when selecting for long reads?

Nanopore sequencing chemistry: LSK108, R9.4
Improving Read Lengths: Shear * FFPE Repair

FFPE repair increases yield slightly for samples sheared larger, but has no impact on 15kb sheared input – for this specific sample.

Nanopore sequencing chemistry: LSK108, R9.4.1
Improving Read Lengths: Rapid kit RAD004

15 minute protocol

Yield:
3Gb from 150K reads
0.89Gb from >50kb reads

Nanopore sequencing chemistry: RAD003/R9.4; RAD004/R9.4.1
Methodology extensible to other plants

Maize courtesy of B. Vaillancourt and Krystle Wiegert-Rininger of the C. Robin Buell Lab at Michigan State University
Conclusions

HMW
100kb+ average

Yield
>10ug gDNA
From 1g tissue

Quality
Nanodrop, gel migration in range

Reproductibility
Want no Wizards

Sequencing Yield
>5Gb per run

Extractions reproduced by other groups

6-10Gb per run sheared, 3-5 unsheared
Acknowledgments

- Timp Lab
- Stephanie Hao

- Salzberg Lab
- Jennifer Lu

- Neale Lab
- Alison Scott
- Zane Moore

- Circulomics
- Kelvin Liu
- Duncan Kilburn
- Jeffrey Burke
- Renee Fedak

Redwood Genome Project (Neale)

National Human Genome Research Institute
1R01HG009190-01A1 (Timp)
2R44GM109618-02 (Liu)