Nanopore Sequencing in the Redwood Genome Project: DNA Extraction to Sequencing Optimization

Rachael Workman

Timp Lab, Johns Hopkins University, Department of Biomedical Engineering

Sequoia sempervirens
Coast redwood
California – central/north coast

Sequoiadendron giganteum
Giant sequoia
California – Sierra Nevada

Save The Redwoods LEAGUE®

Two species in our redwood genome project

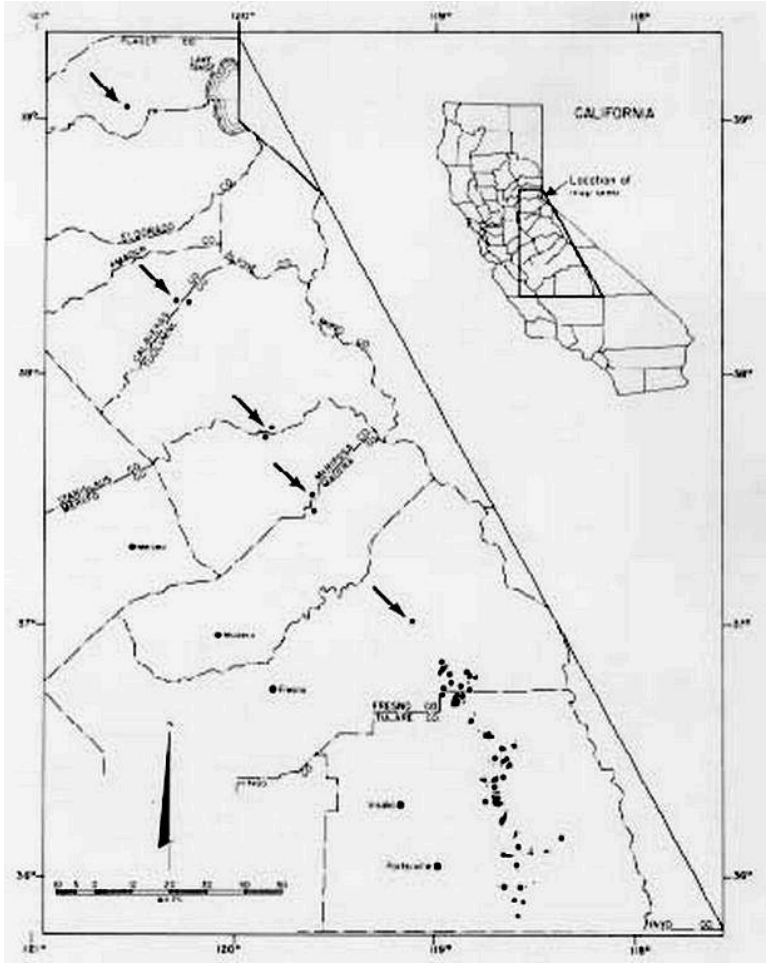
California endemics

Economic, cultural, and conservation value

”Advanced management strategies”


Photo: Harold Hoyer

Giant sequoia
Sequoiadendron giganteum
Cupressaceae

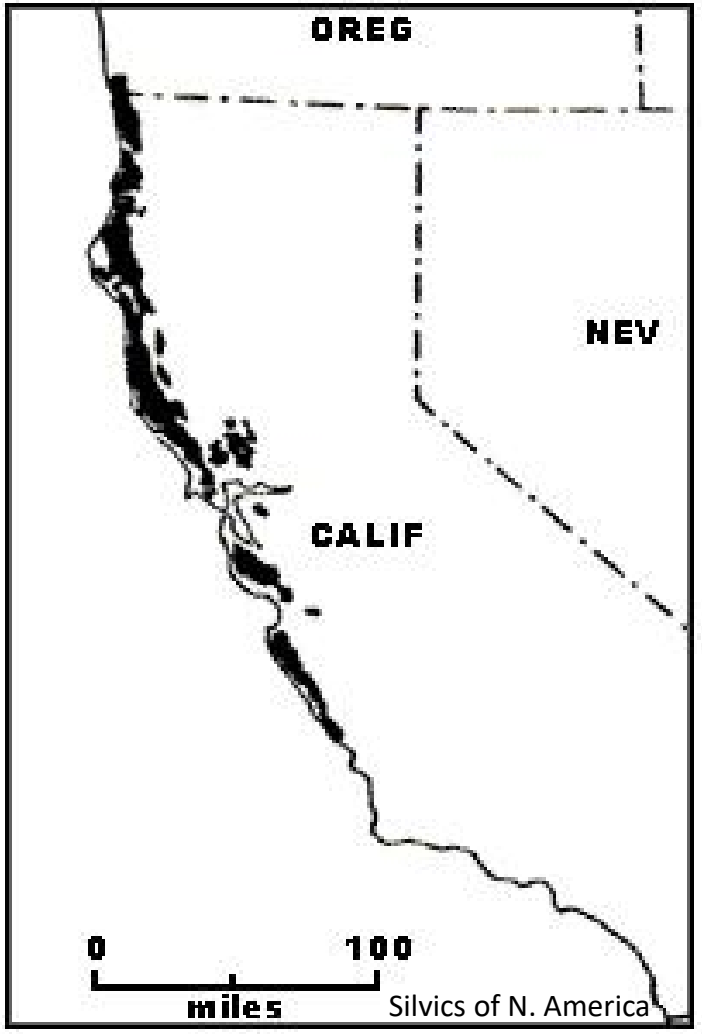


Silvics of N. America

2n = 2x = 18
9 gigabase genome

Occur in groves
throughout Sierra Nevada

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

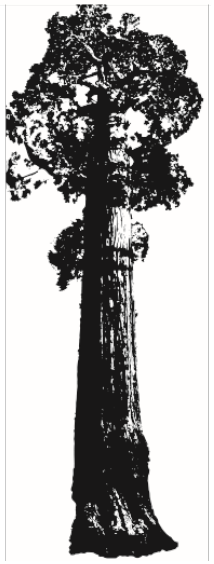
Short read
sequencing @ UCD

DNA extraction
Timp lab

Long read
sequencing @ JHU

Reference
genome

Assembly
@ JHU

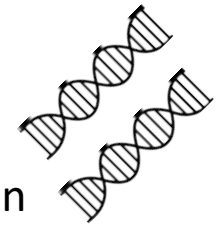


Climbing & collection
by
Steve Sillett + team

DNA extraction
Neale lab

Short read
sequencing @ UCD

DNA extraction
Timp lab



Long read
sequencing @ JHU

Climbing & collection
by
Steve Sillett + team

DNA extraction
Neale lab

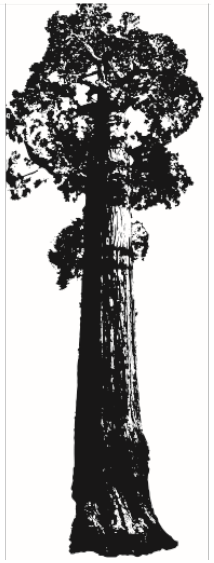
Short read
sequencing @ UCD

DNA extraction
Timp lab

Long read
sequencing @ JHU

Reference
genome

Assembly
@ JHU

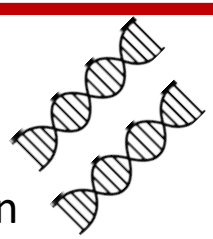


Climbing & collection
by
Steve Sillett + team

DNA extraction
Neale lab

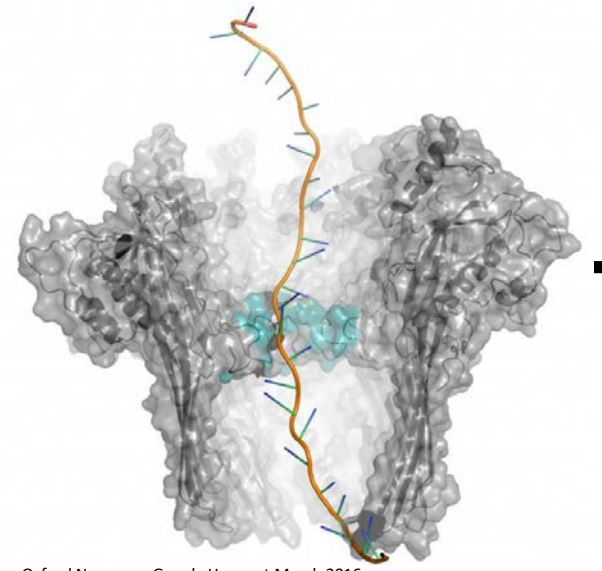
Short read
sequencing @ UCD

DNA extraction
Timp lab

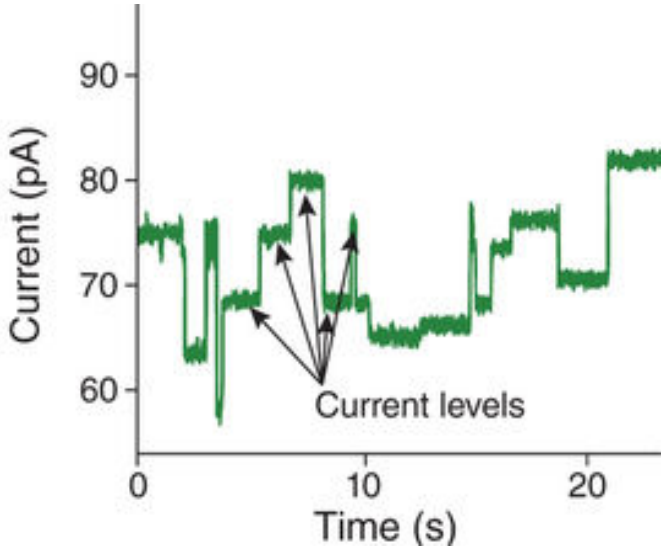


Long read
sequencing @ JHU

Nanopore Single Molecule Sequencing



Oxford Nanopore Google Hangout March 2016



Deamer et al 2016, Nature Biotech

ATCGATCGATAG
TATTAGATACGA
CTAGCGATCAG

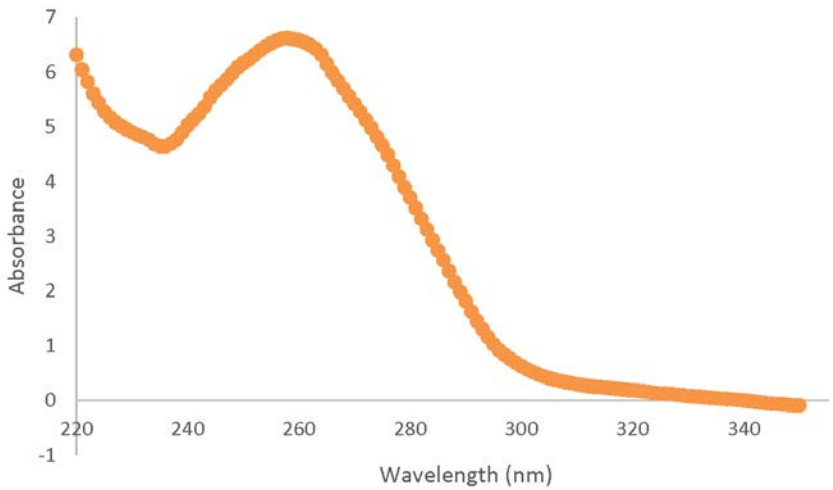
No theoretical upper limit to sequencing read length, practical limit only in preparing long fragment libraries and delivering DNA to the pore intact

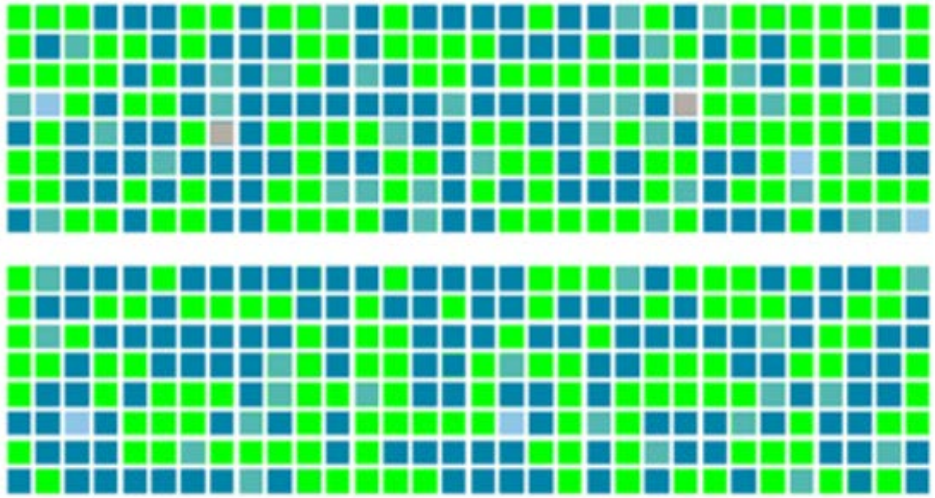
Typical user-reported sequencing output 5-15Gb (as of R9.4.1, March 2018)

Disclosure: Timp has two patents (US 2011/0226623 A1; US2012/0040343 A1) licensed to ONT

Sample requirements for sequencing

HMW	Yield	Quality	Reproducibility	Sequencing Yield
100kb+ average	>10ug gDNA From 1g tissue	Nanodrop, gel migration in range	Want no Wizards	>5Gb per run





Sample realities before optimization

LMW

<10kb
average

Low Yield

<1ug gDNA
From 1g tissue

Poor Quality

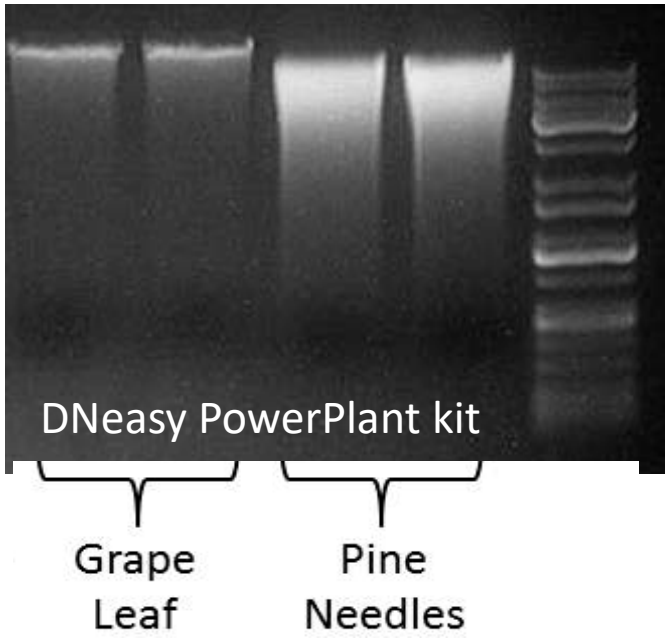
Residual polyphenolics
And polysaccharides

Inconsistent

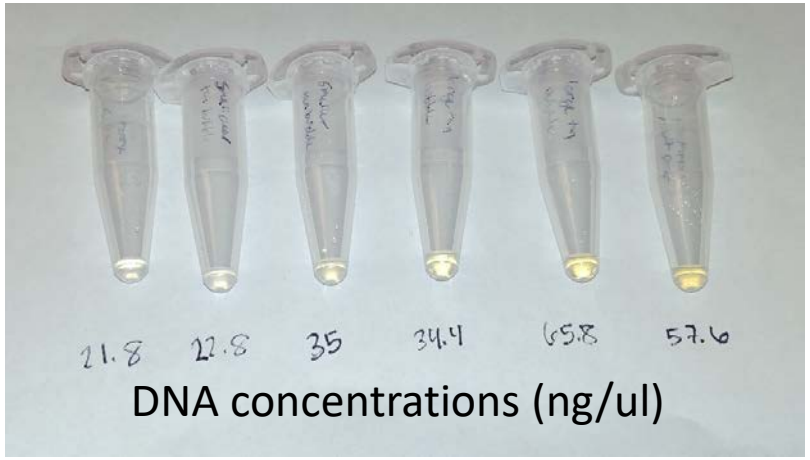
Results varied largely
By sample

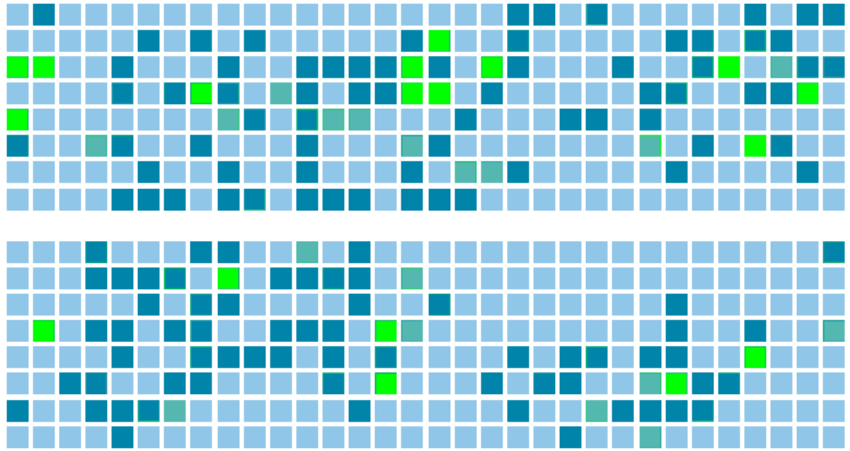
Low Seq Yield

<1Gb per run

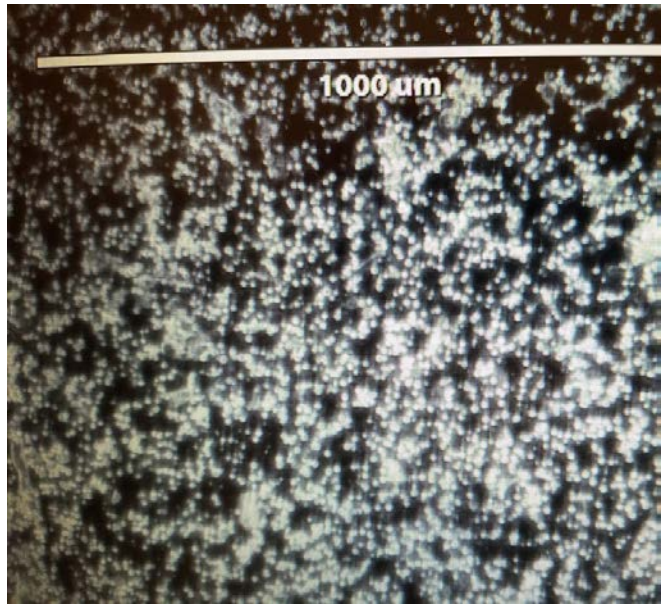


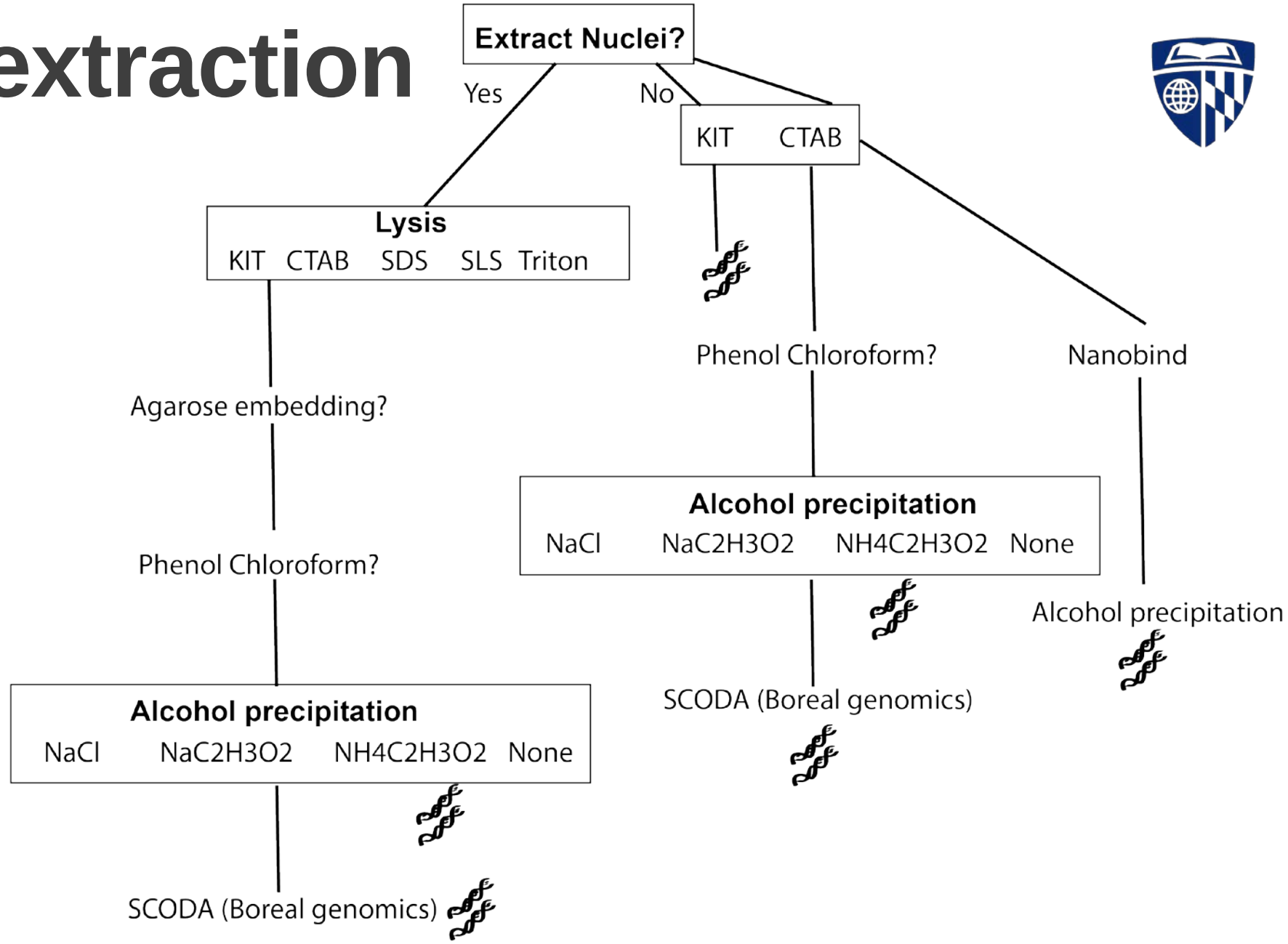
10kb





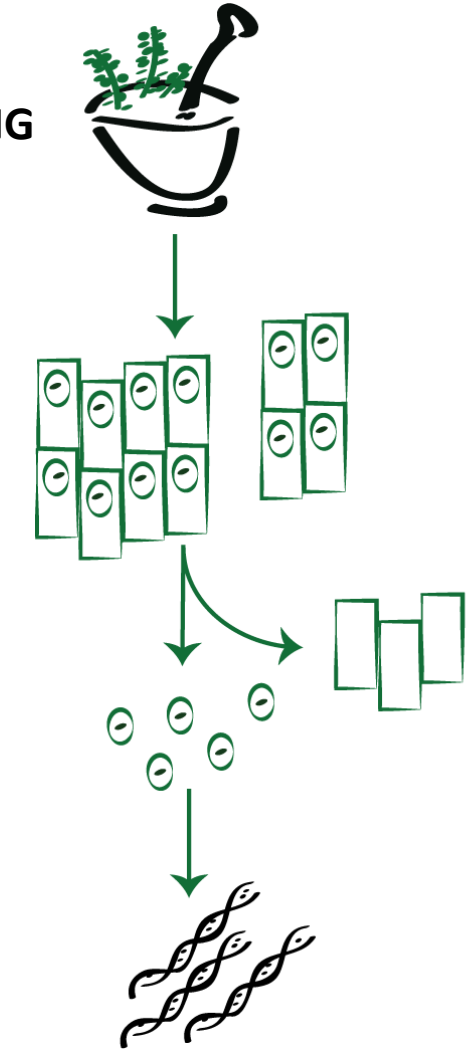
Trials: DNA extraction



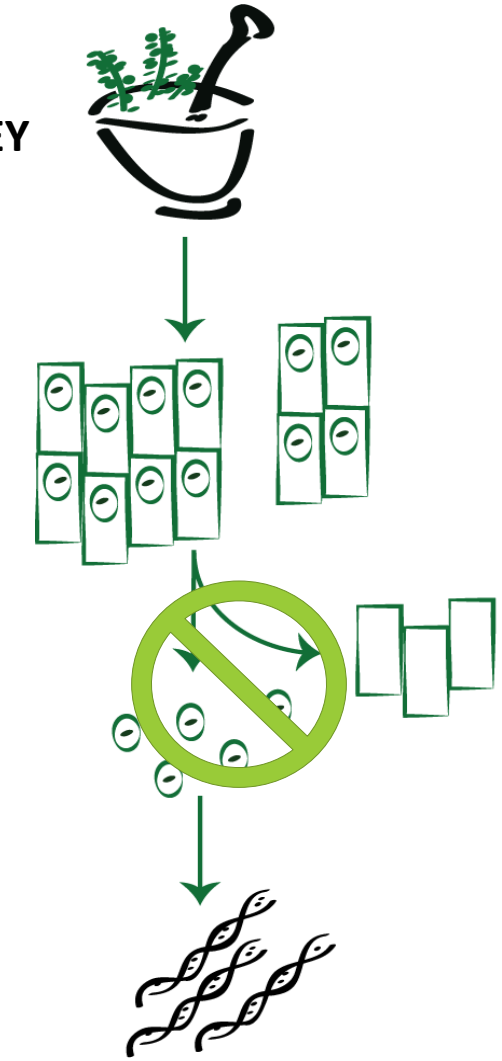


Top 3 extraction protocols

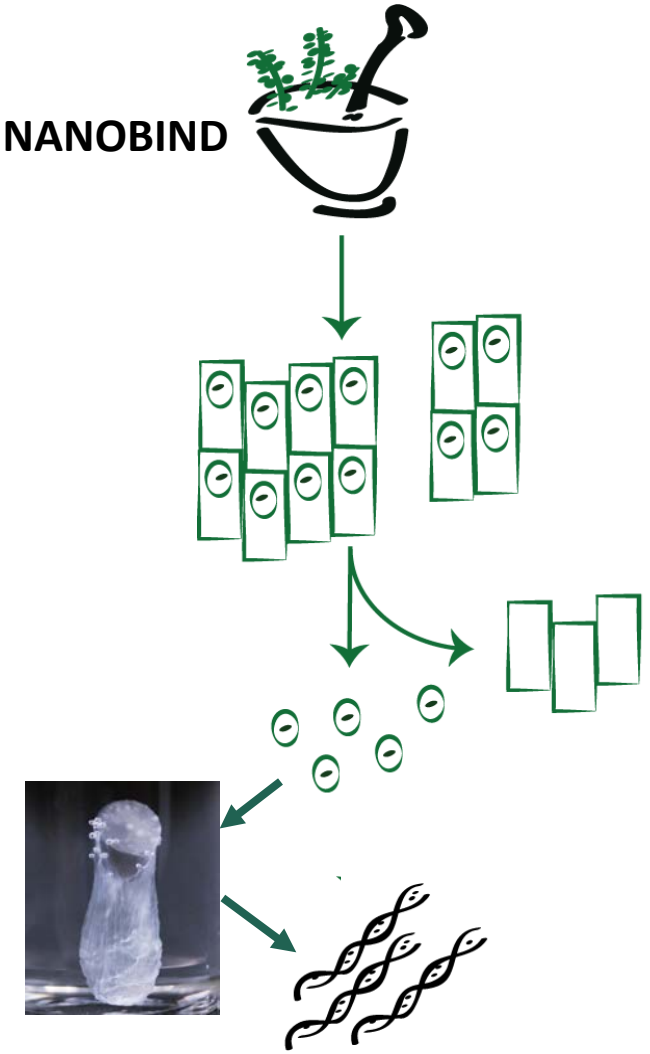
ZHANG

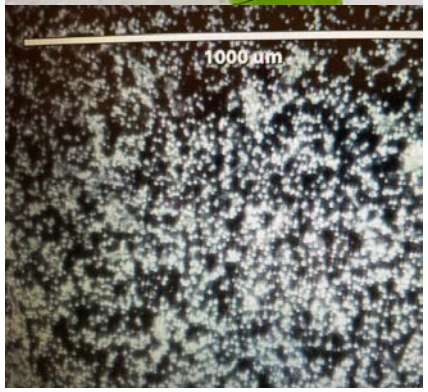


HEALEY

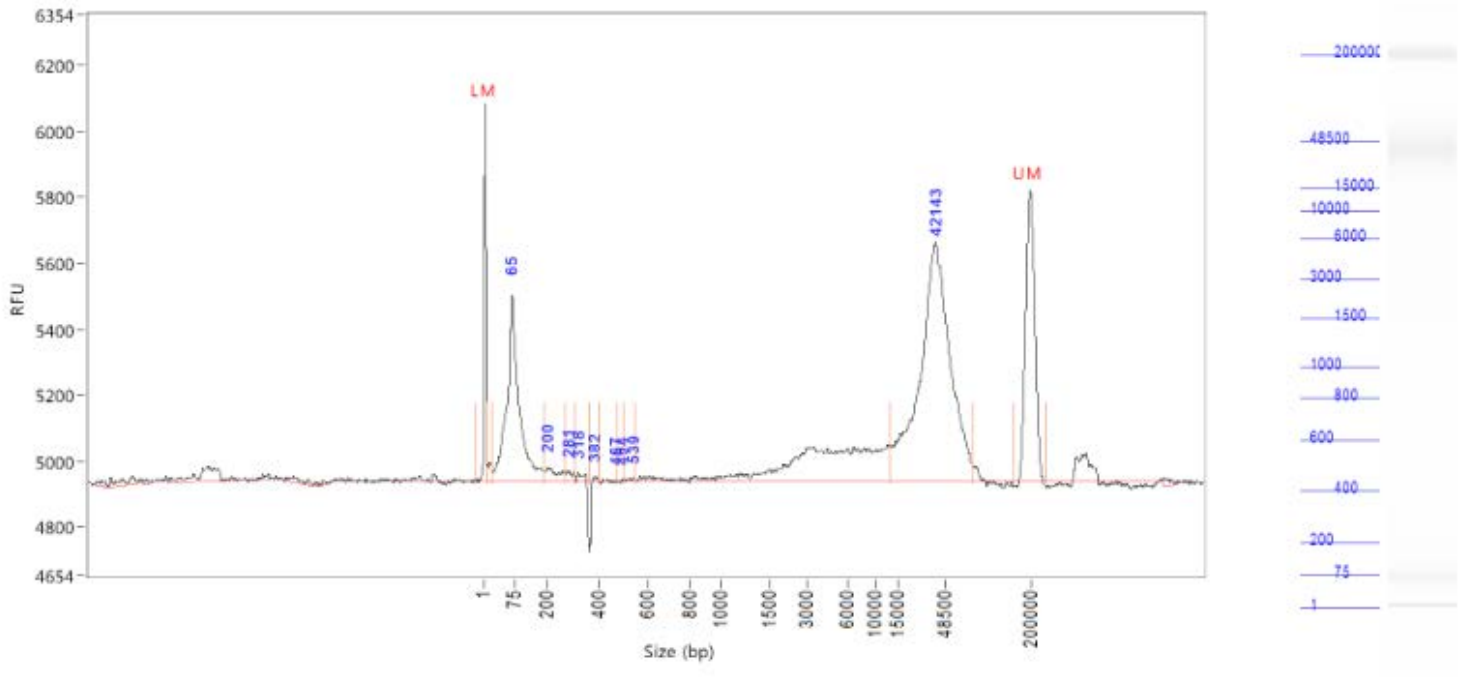


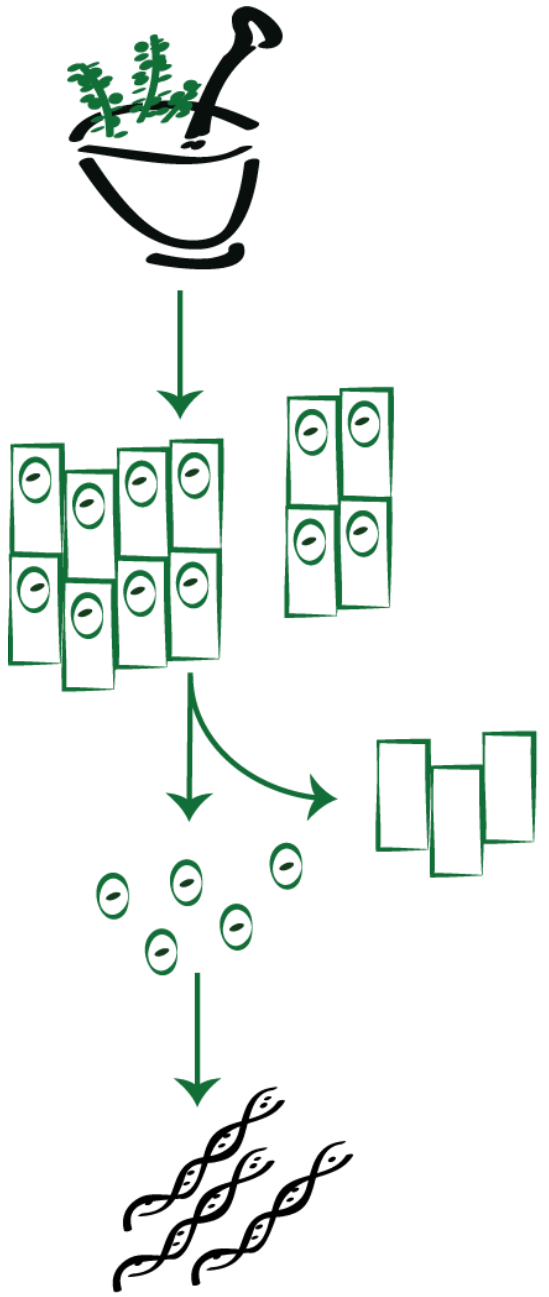
NANOBIND





Top 3 extraction protocols: Zhang





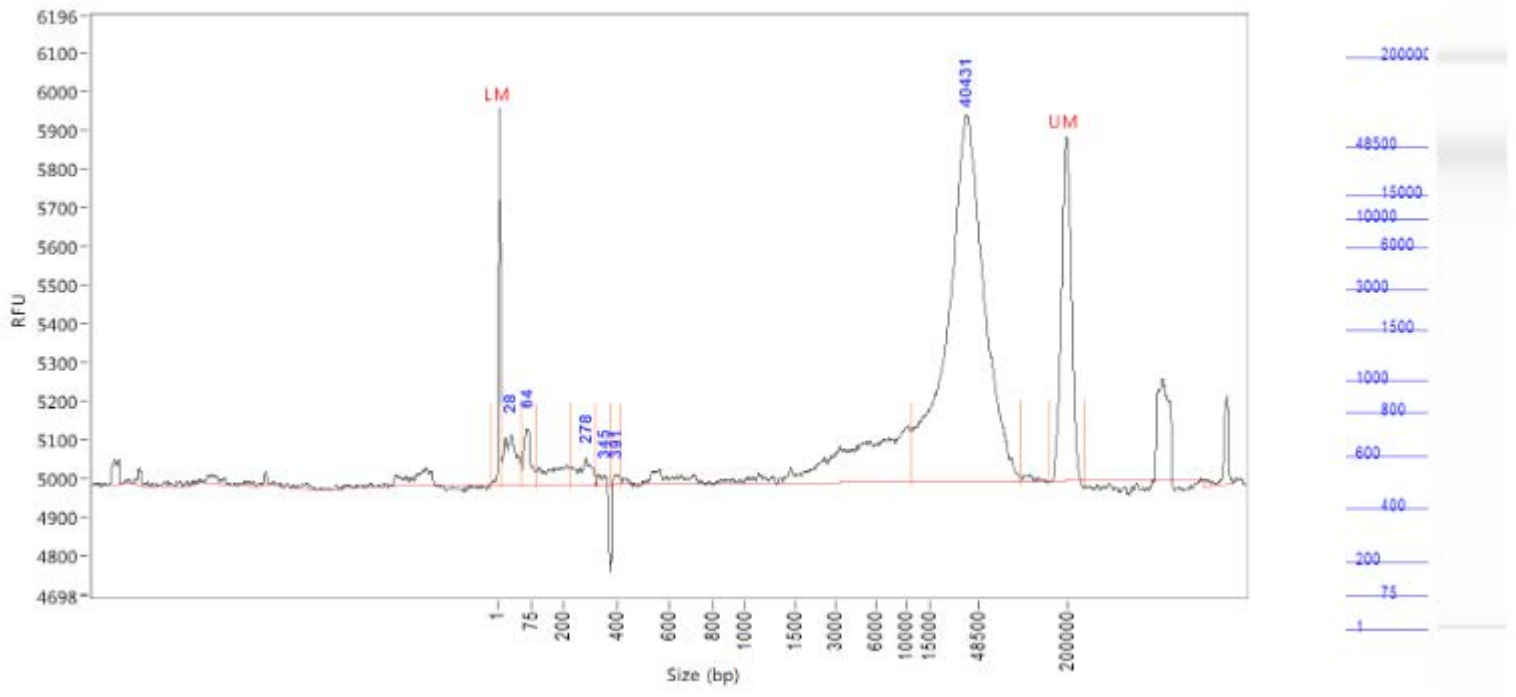
Tissue ground in LN2

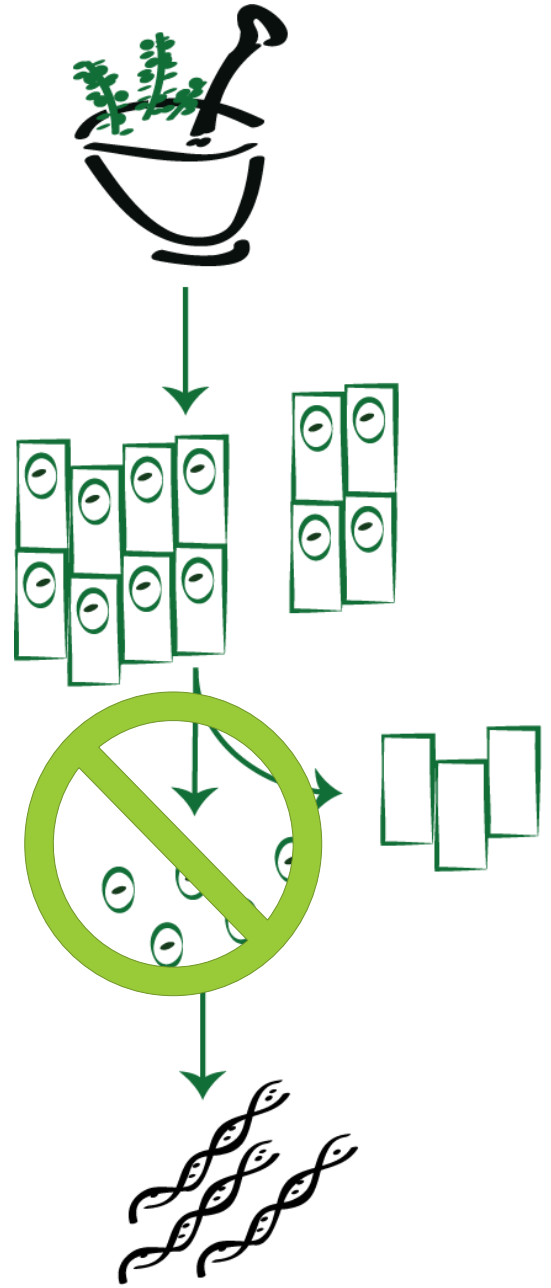
Nuclei: Cell wall lysis -> filtration -> differential centrifugation

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

Zhang et al, Nature Protocols (2012)

Top 3 extraction protocols: Healey





Preferred plant protocol: Skips the differential nuclei extraction

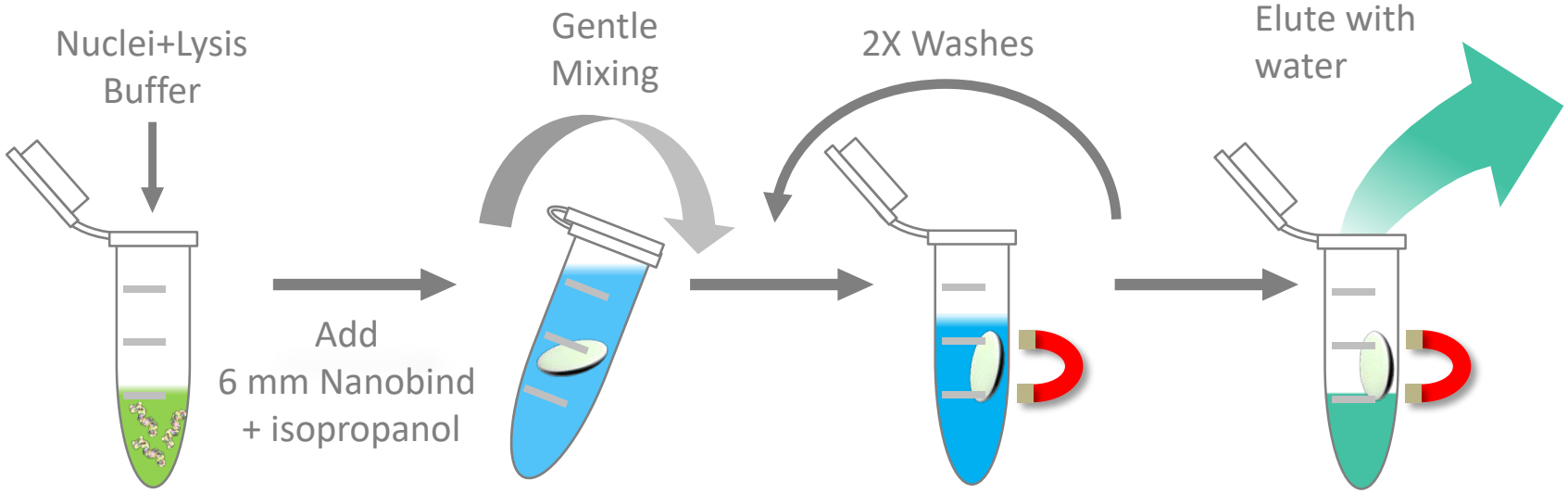
CTAB (cetyl trimethylammonium bromide) – cationic detergent

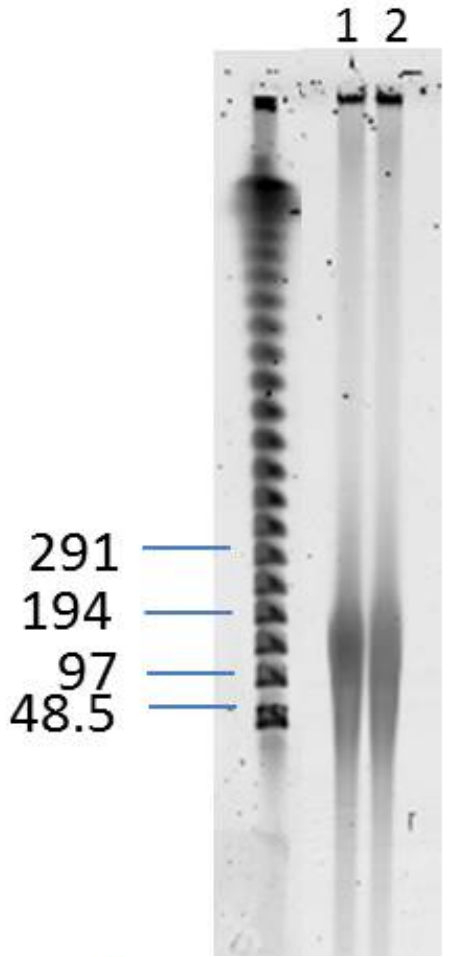
CTAB extraction, chloroform, alcohol precipitation + NaCl, elute

Healey et al, Plant Methods (2014)

Top 3 extraction protocols: Nanobind

Nuclei →





Nanobind: Nuclei isolation as previously described

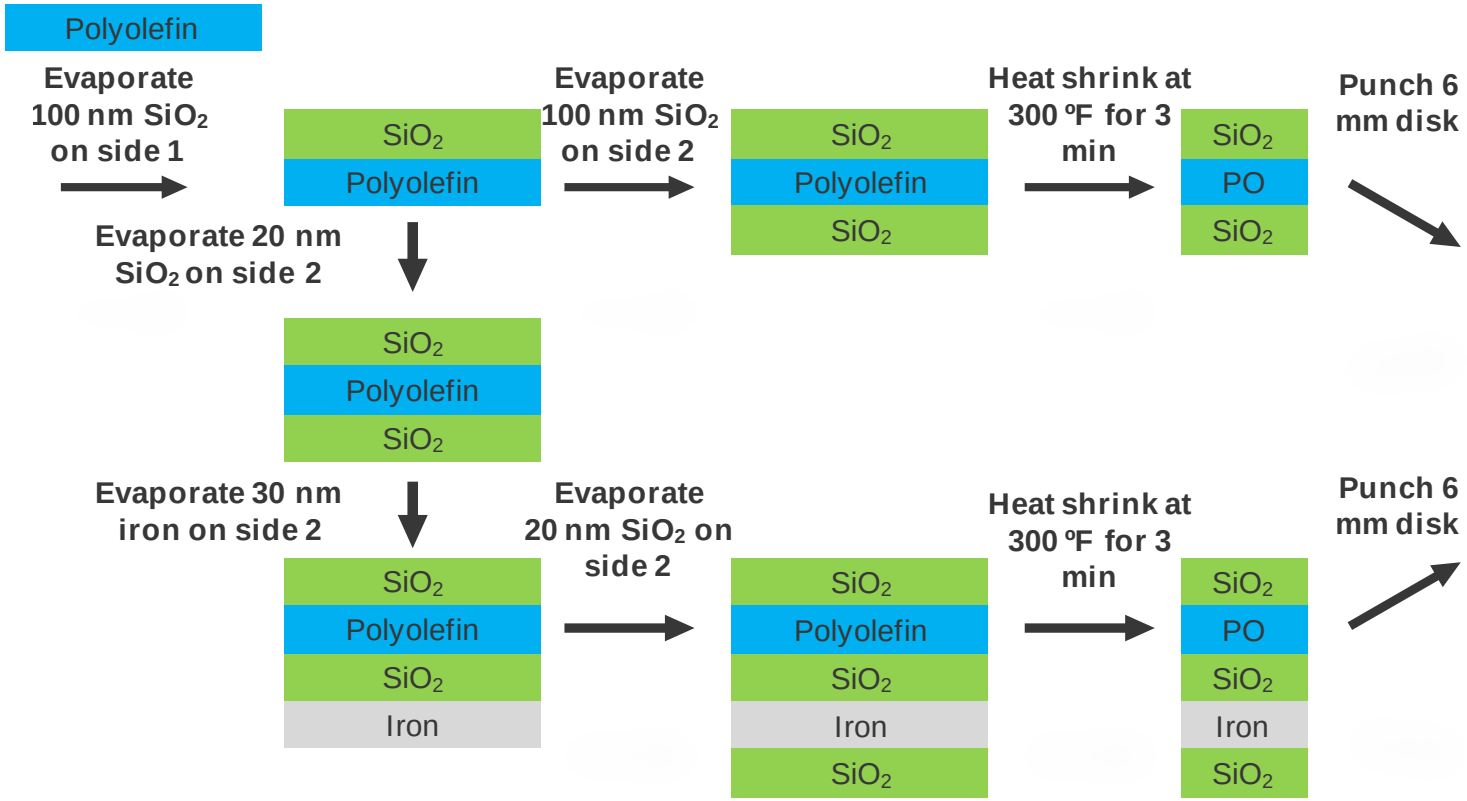
Nanobind extraction with modified CTAB, EtOH washes, elute

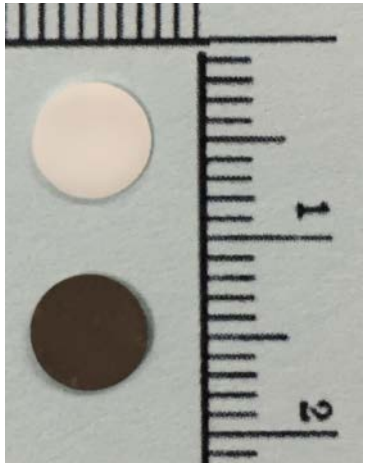
45 minutes total time (after nuclei extract)

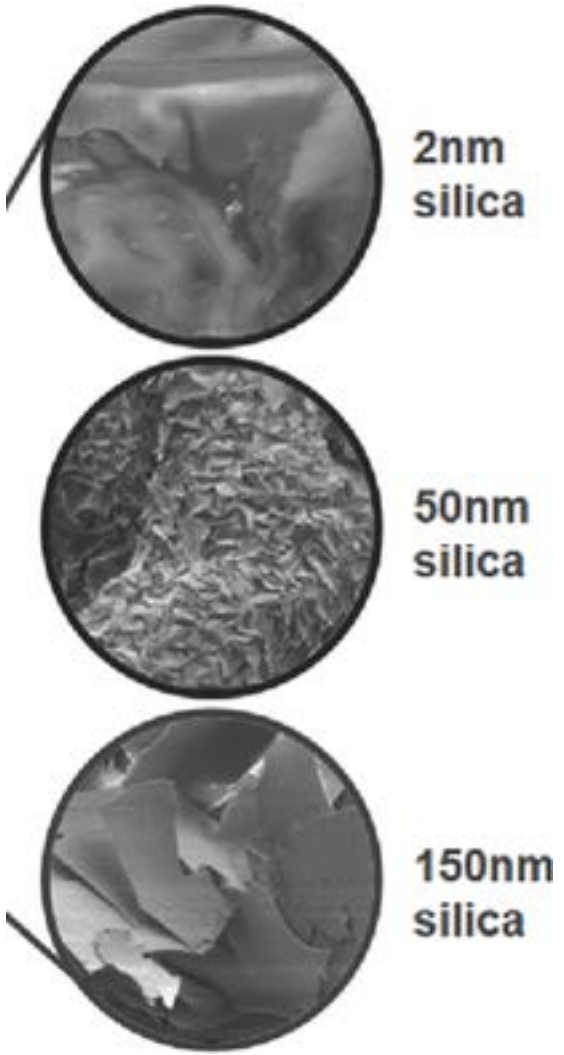
circulomics

Zhang et al. Adv Mat. (2016)

Nanobind: How does it work







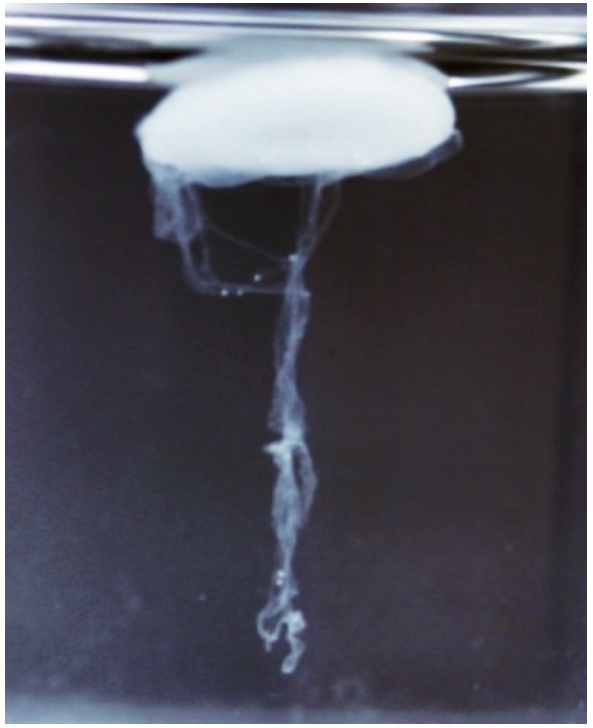
Nanostructured silica (SiO_2), essentially.

With or without Iron layer for magnetism

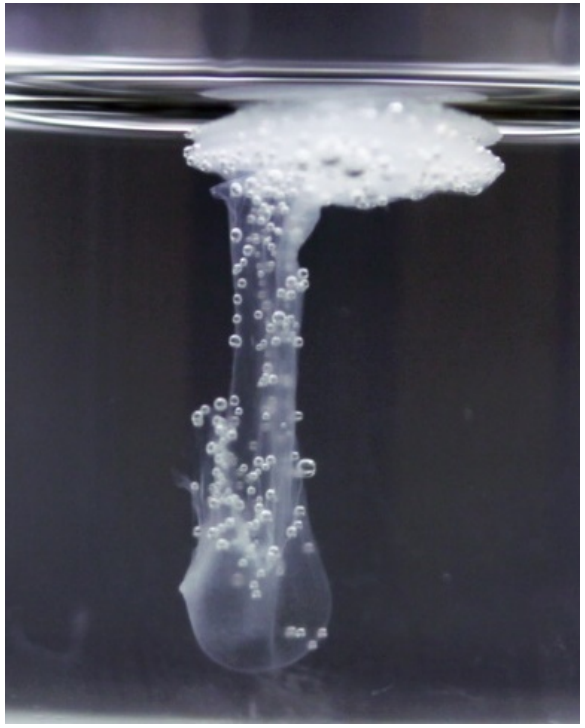
Zhang et al. Adv Mat. (2016)

Tentacle Binding Mechanism

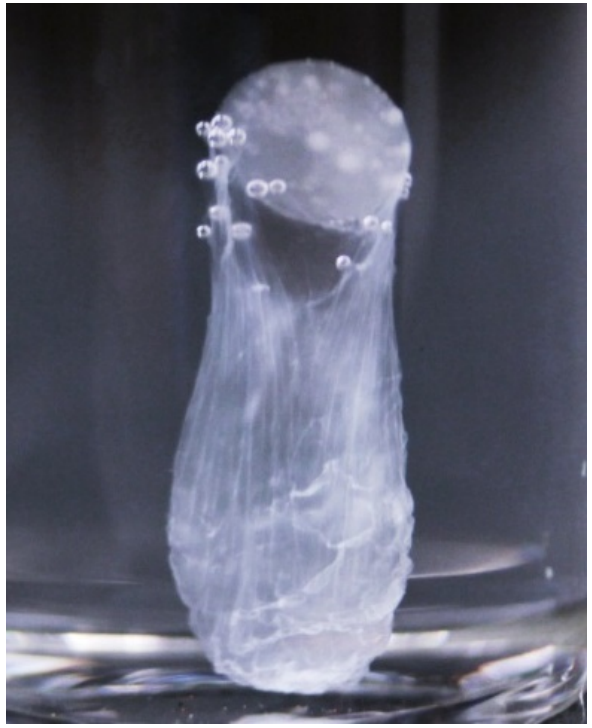
Enhances binding capacity and protects DNA from shear forces



Low Input (10 µg)



Medium Input (50 µg)



High Input (200 µg)

- **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**

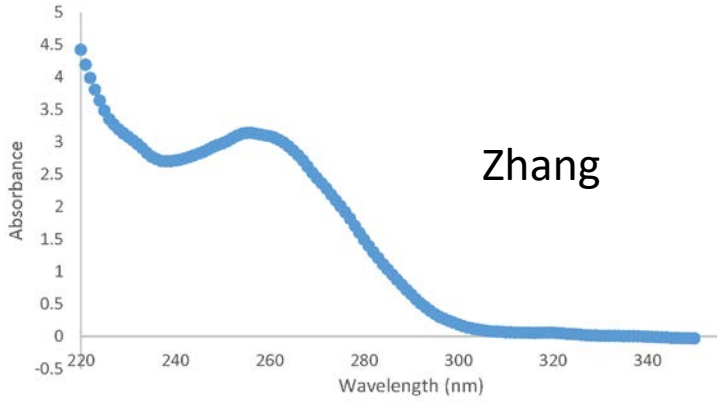
Zhang et al. Adv Mat. (2016)

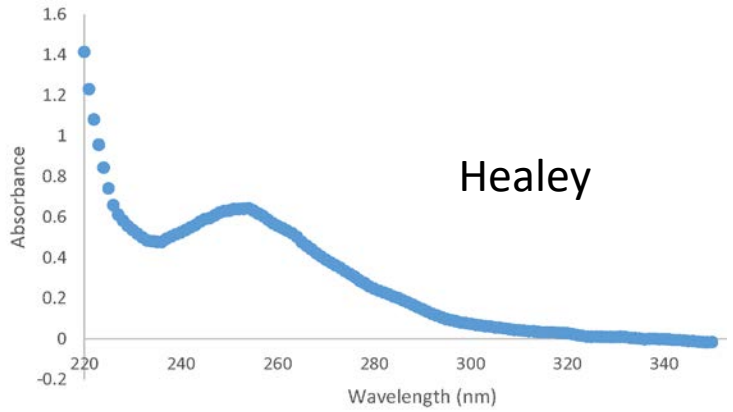
Top 3 extraction protocols: gDNA Yield/Qual

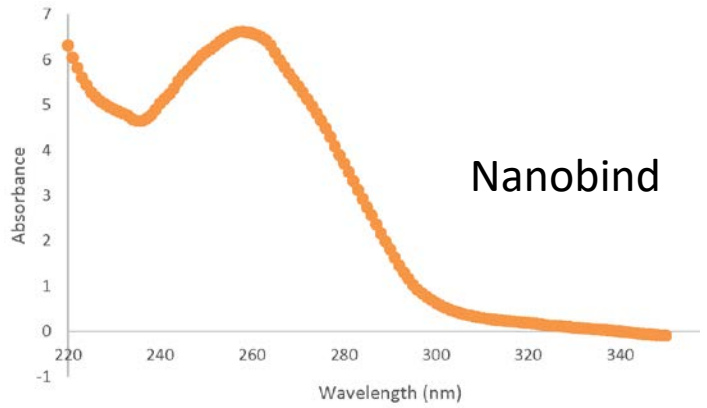
From 1g input	Zhang	Healey	Nanobind
Yield (ug)	**2**	**15**	**20**
260/280	1.65	1.68	1.76
260/230	0.34	0.76	1.51

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

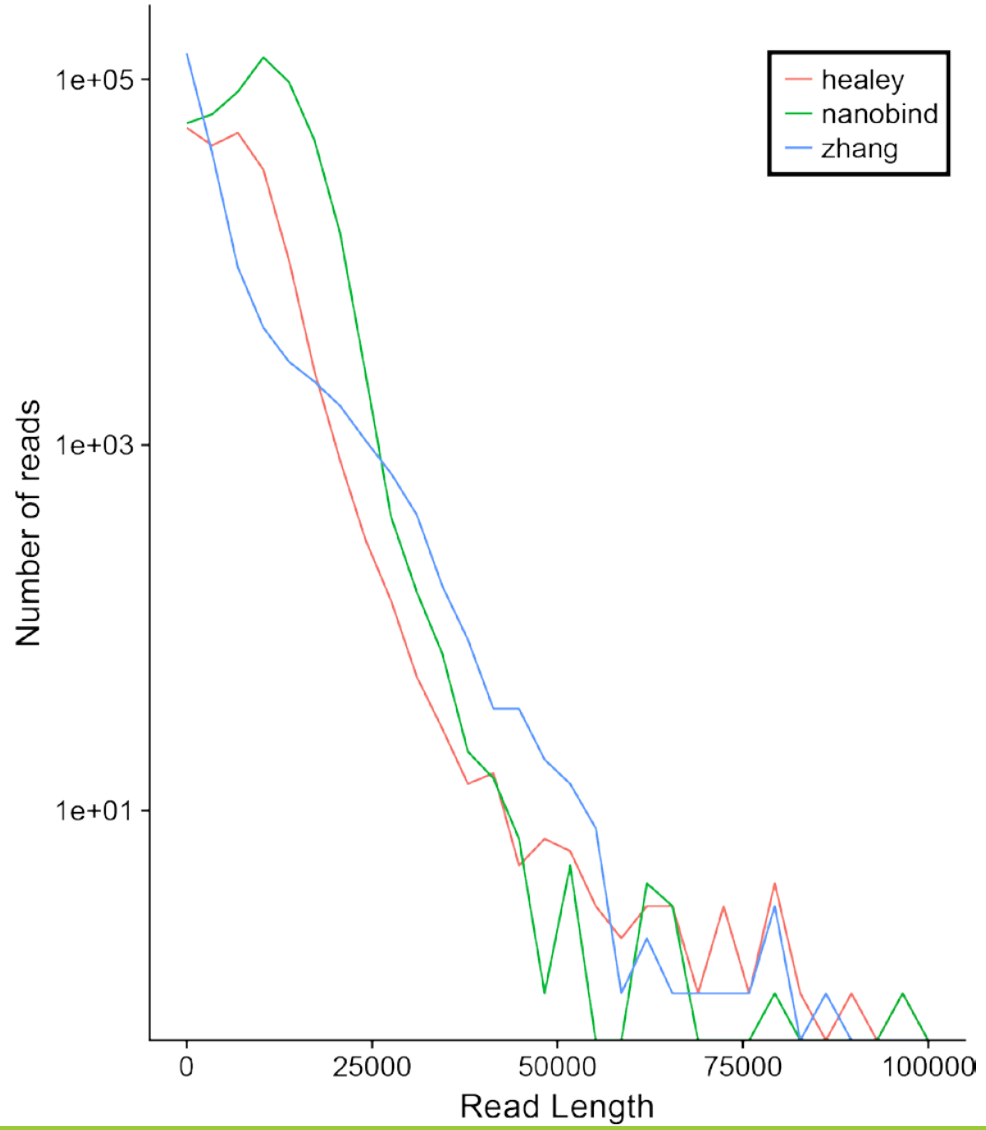
Nanodrop spectra not full picture of quality – plant samples sometimes retain visible color







Extraction Comparison



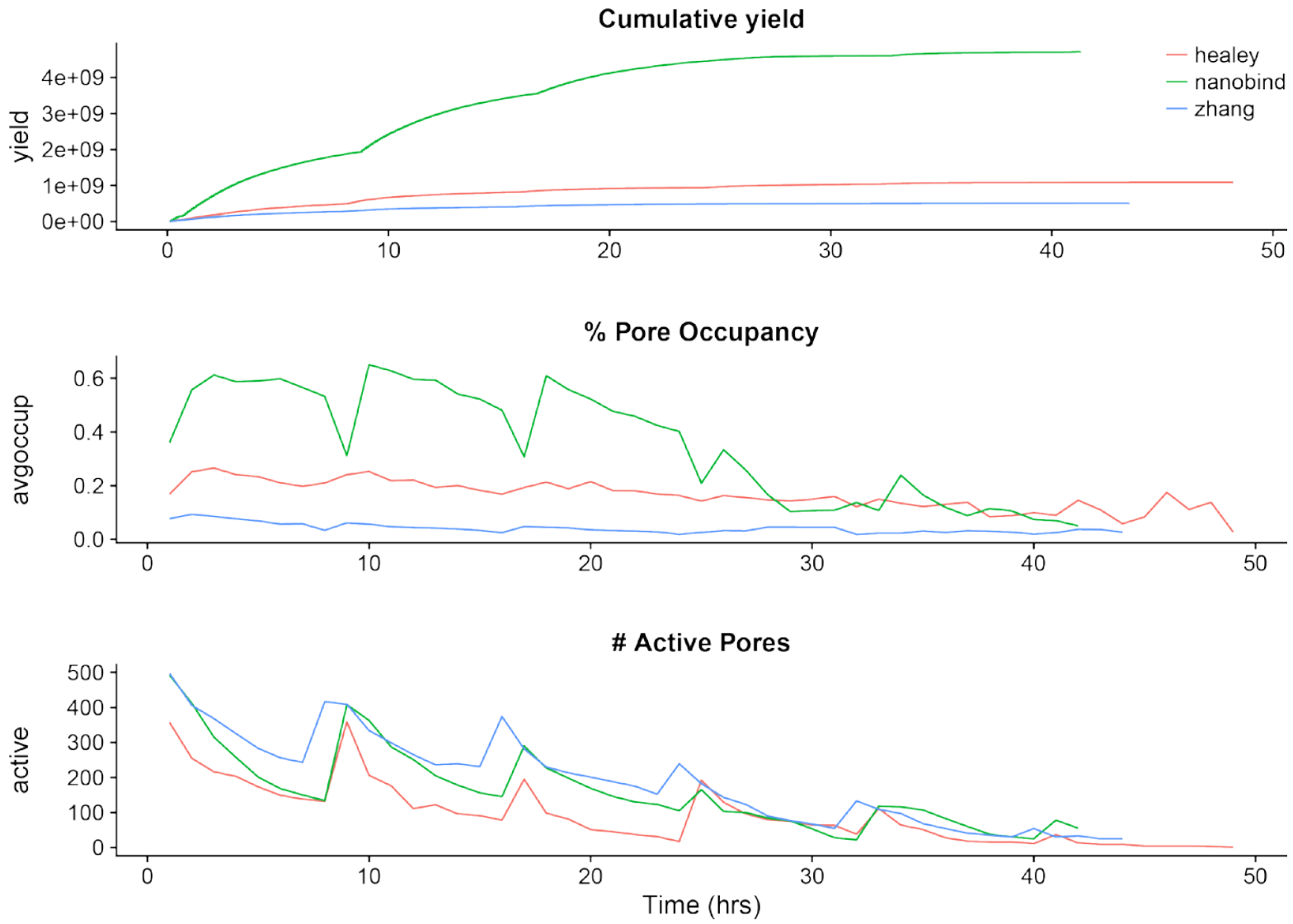
	Zhang	Healey	Nanobind
Reads	201k	195k	500k
Yield	0.51Gb	1.08Gb	4.72Gb
N50	7.1kb	8.6kb	12.3kb
Median	929	5.1kb	9.8kb

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

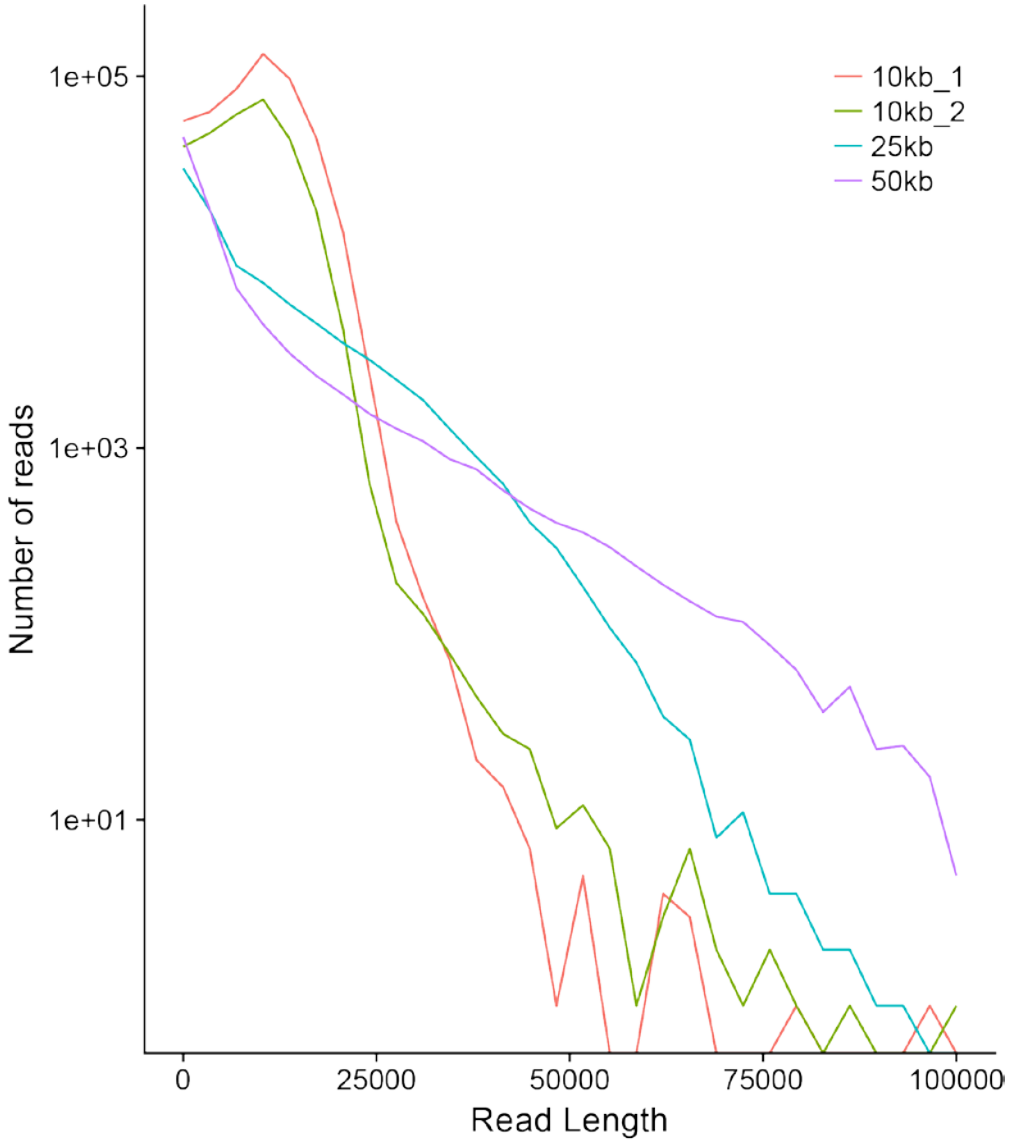
Nanobind sequencing yield increased by 10-fold over Zhang, 5-fold over Healey

Extraction Comparison – Time Plot

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



Shear Comparison

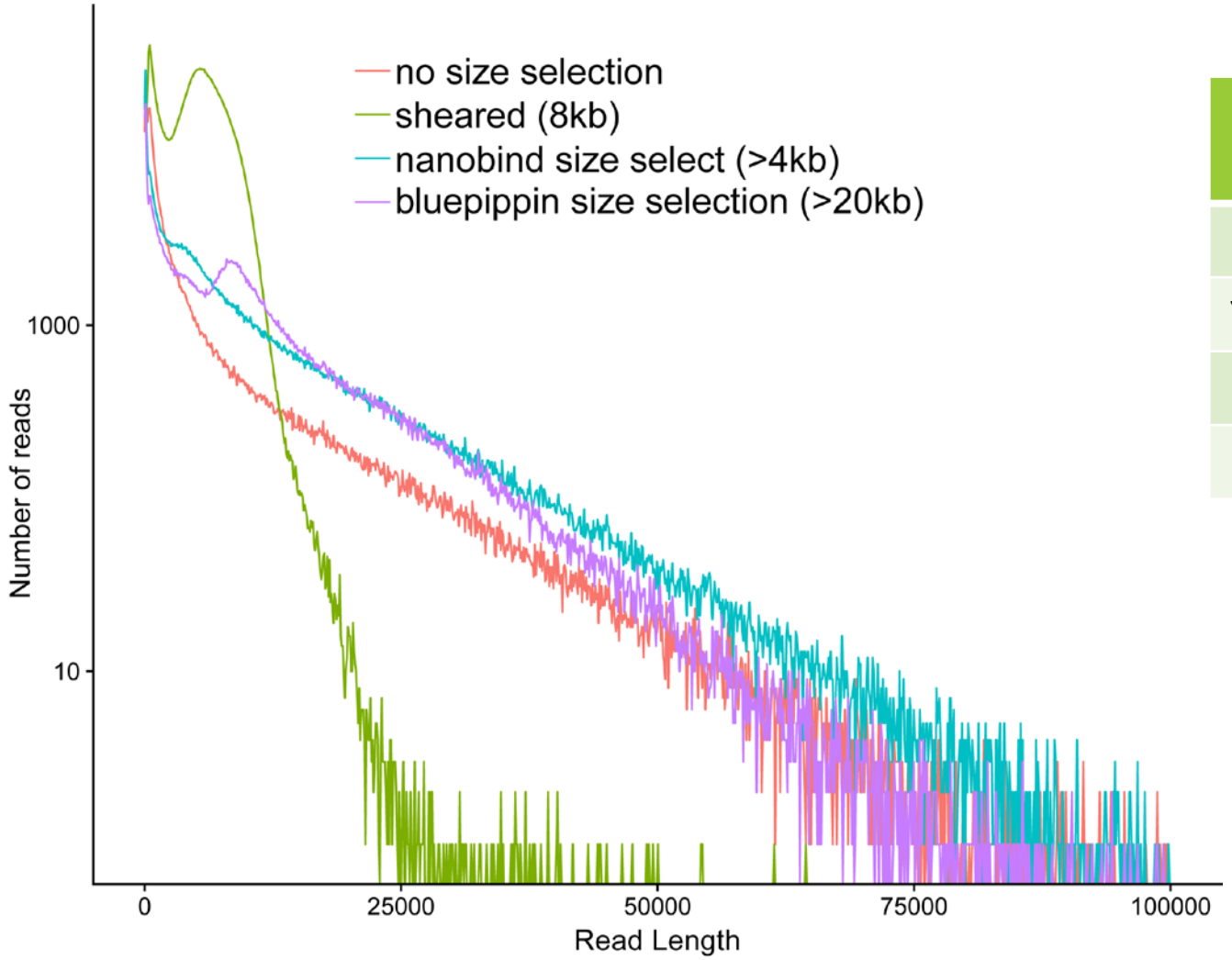


	10kb	10kb_2	25kb	50kb
Reads	500k	299k	93.7k	94.7k
Yield	**4.72Gb**	**2.47Gb**	**0.82**	**0.66**
N50	12.3kb	11.3kb	19.8	24
Median	**9.8kb**	**8.5kb**	**4.1**	**1.7**

Currently a trade-off between long reads and high yield

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

Improving Read Lengths: Size selection



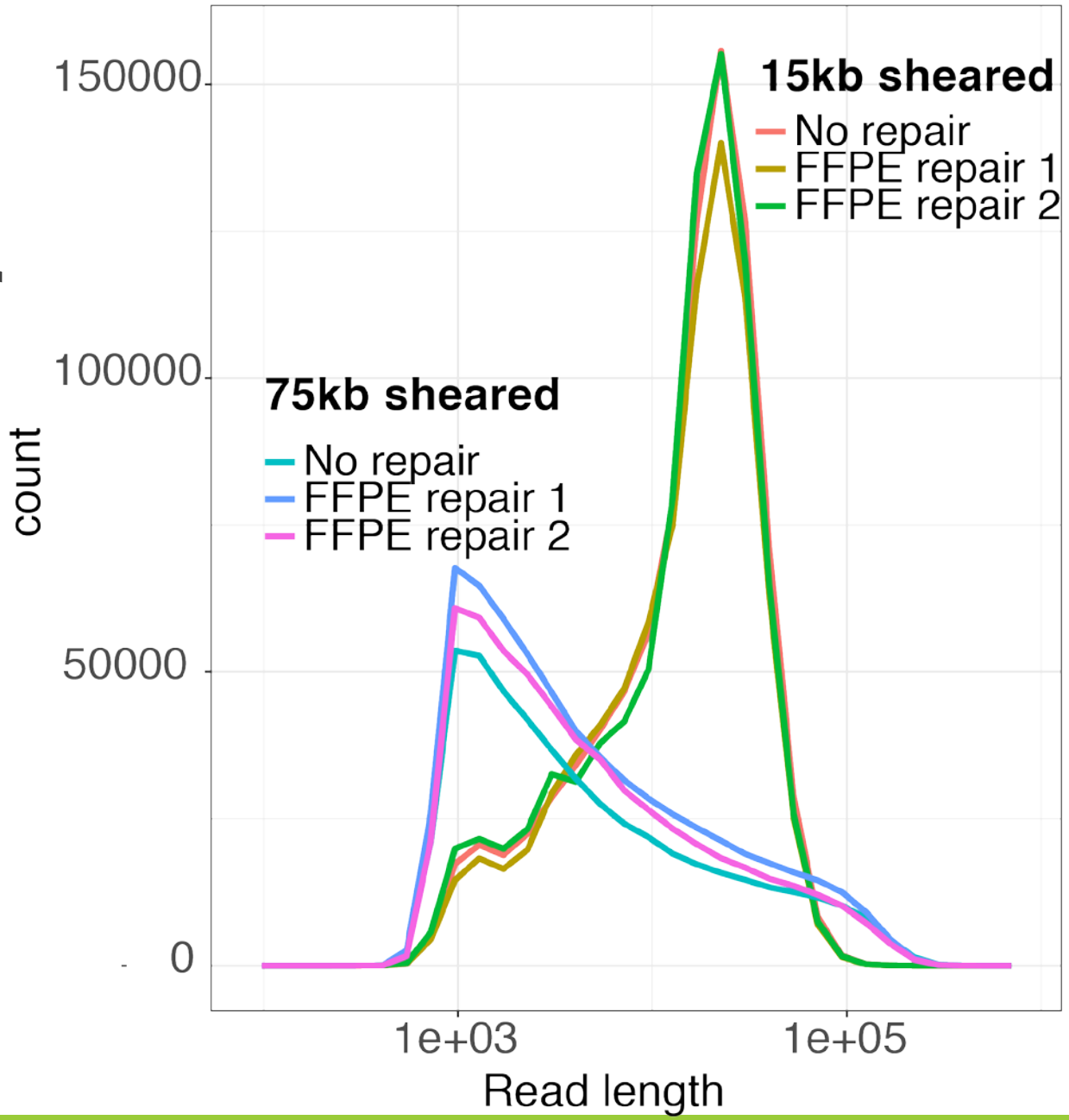
	None	Sheared	NB (4kb)	BP (20kb)
Reads	353k	2060k	400k	435k
Yield	**1.71Gb**	**10.1Gb**	**3.57Gb**	**3.65Gb**
N50	17.3kb	6.6kb	15.7kb	19.0kb
Median	**1.2kb**	**5.1kb**	**6.8kb**	**4.3kb**

Using Nanobind as an AMPure replacement, we get reasonable yield

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?

Improving Read Lengths: Shear * FFPE Repair

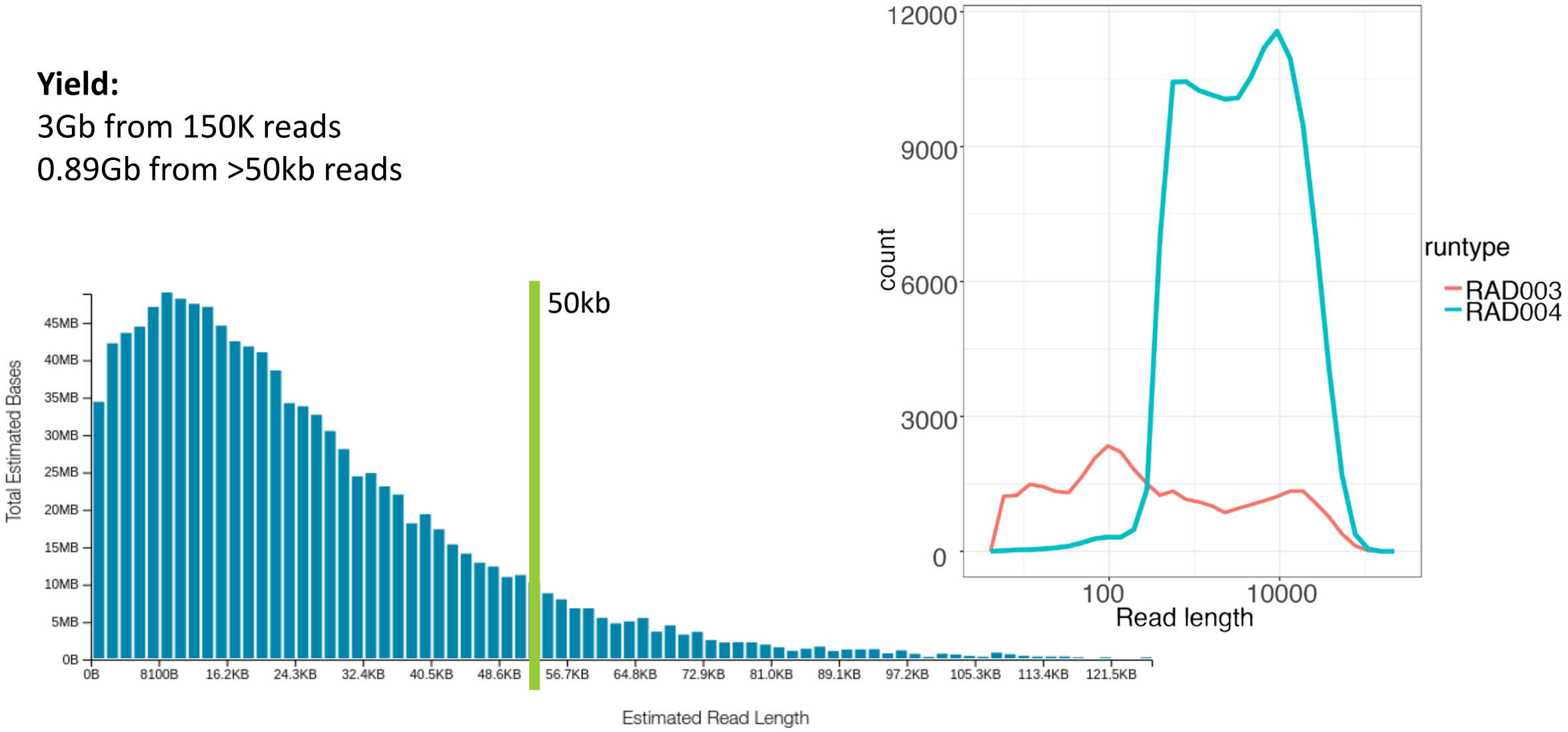


Currently a trade-off between long reads and high yield

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

Improving Read Lengths: Rapid kit RAD004

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



Conclusions

Customized extraction methodologies are required for many organisms

- Improved sequence data quality and quantity through improved extraction

HMW extraction from plants is hard

- Polyphenolics and polysaccharides, additional impurities difficult to eliminate – but this rigorous purification is needed

At the moment, nanopore sequencing yield is maximized with shorter fragment input. This is not merely a molarity issue.

- Yield and median read length decrease drastically with 25kb and 50kb shearing relative to 10kb.
- Our best solution – target 8kb for high yield runs, sprinkle in some long-read runs.




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)