High Molecular Weight Conifer DNA Extraction for Nanopore Sequencing

Rachael Workman[1], Stephanie Hao[1], Renee Fedak[2], Kelvin Liu[2], Steven Salzberg[3], David Neale[4], Winston Timp[1]

[1]Johns Hopkins University Department of Biomedical Engineering
[2]Circulomics, Inc
[3]Johns Hopkins University, Center for Computational Biology Institute for Genetic Medicine and Departments of Biomedical Engineering, Computer Science and Biostatistics
[4]University of California at Davis Department of Plant Sciences

Abstract

The assembly of high quality conifer genomes can benefit many fields of research from conservation and restoration efforts, to disease and stress studies, even evolutionary history. However, these tree genomes present unique assembly challenges; they are large (10-30+ Gb haploid), repetitive, and can have high ploidy. While long read sequencing (Oxford Nanopore and PacBio) can greatly improve assembly contiguity, extracting large amounts of high quality, high molecular weight (HMW) DNA from adult trees presents a unique challenge. Although many extraction methodologies exist for recalcitrant plant species, most either yield DNA of quality “fit for PCR” and not for sensitive, i.e. nanopore, sequencing applications, or DNA too fragmented to obtain sequencing reads of sufficient length to improve assembly contiguity. Obtaining 60kb+ and "nanopore clean" DNA places higher demands on sample extraction and preparation than existing methodology can provide in adult trees.

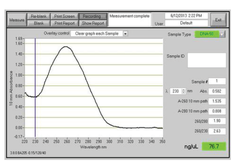
We have combined several techniques to develop HMW, "nanopore clean" extraction methodologies from Gymnosperm species *Sequoiadendron giganteum* (Giant sequoia) and *Sequoia sempervirens* (Coastal redwood), and generated preliminary sequence data on the Oxford Nanopore MinION. Our method integrates nucleic isolation and Nanobind DNA extraction (Circulomics Inc) to improve purity and recovery 10-fold, and reduce extraction time from 2-3 days to a single day. We also detail sequencing library preparation methodology, including the relationship between DNA shearing size and utility for assembly. We found through differential shearing with the Megaruptor (Diagenode) that input fragment lengths of longer than 10kb decrease throughput, but increases read N50 for assembly. Lastly, we specify an analysis pipeline for data QC, filtering and polishing in preparation for genome assembly.

Goals

High Yield

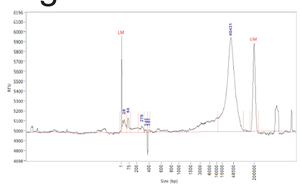
>10µg HMW gDNA from 1g of leaf tissue

High Quality



Spectra within normal range, no aberrant migation, sequenceable

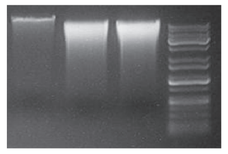
High MW



100kb+ average

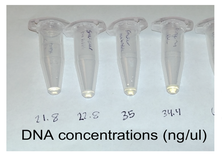
Challenges

Length vs Yield



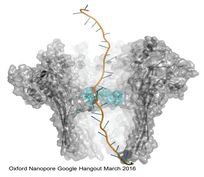
Existing protocols optimized for <20kb

Extract Impurities



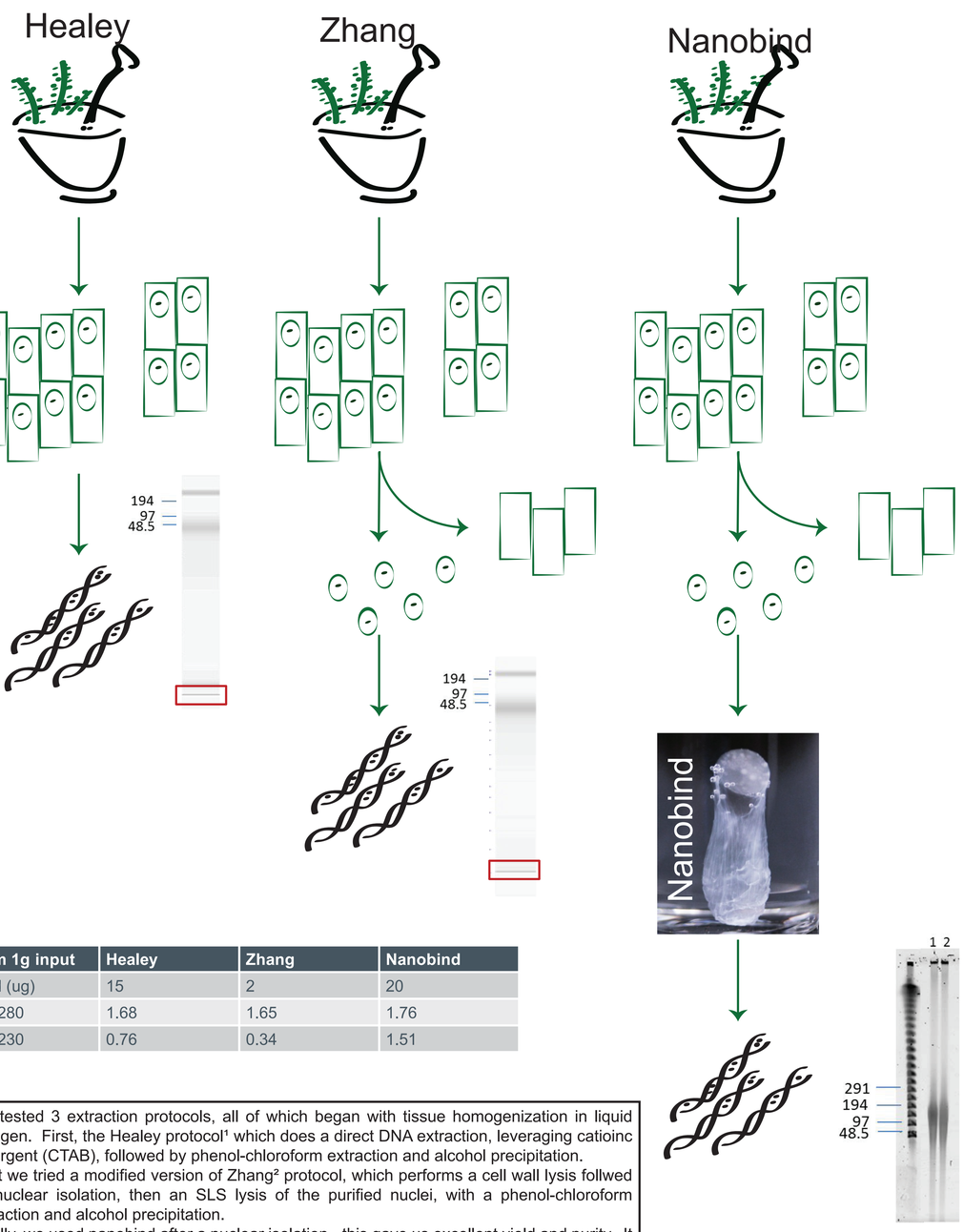
Polyphenolics, polysaccharides co-precipitate with DNA

Sensitive Sequencer



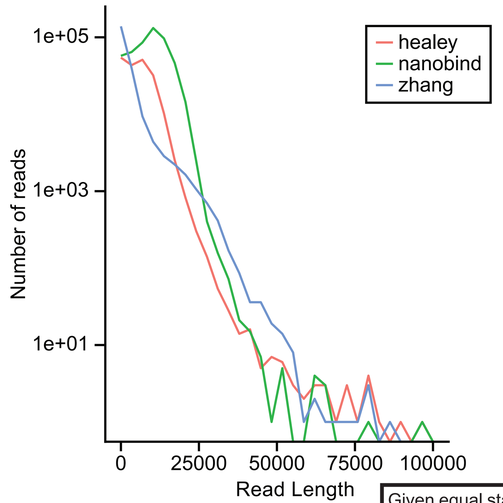
Biological nanopores not robust to residual impurities

Results

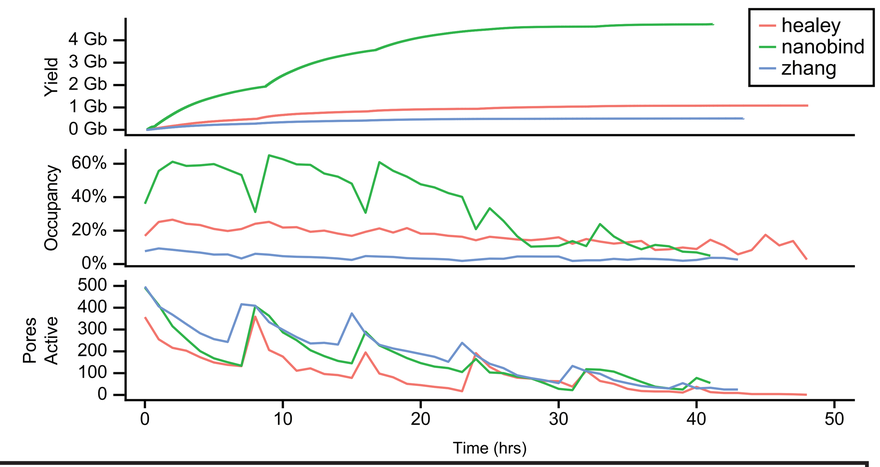


From 1g input	Healey	Zhang	Nanobind
Yield (ug)	15	2	20
260/280	1.68	1.65	1.76
260/230	0.76	0.34	1.51

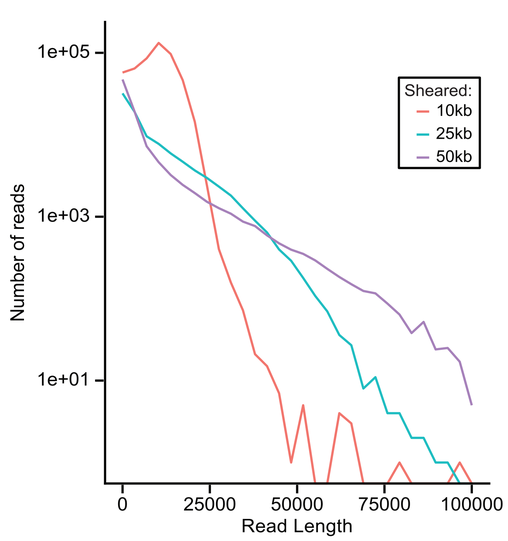
We tested 3 extraction protocols, all of which began with tissue homogenization in liquid nitrogen. First, the Healey protocol[1] which does a direct DNA extraction, leveraging catioinc detergent (CTAB), followed by phenol-chloroform extraction and alcohol precipitation.
Next we tried a modified version of Zhang[2] protocol, which performs a cell wall lysis follwed by nuclear isolation, then an SLS lysis of the purified nuclei, with a phenol-chloroform extraction and alcohol precipitation.
Finally, we used nanobind after a nuclear isolation - this gave us excellent yield and purity. It is important to note that 260/280 and 260/230 are not the complete picture of quality, plant samples may retain contaminants which absorb at different wavelengths.



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

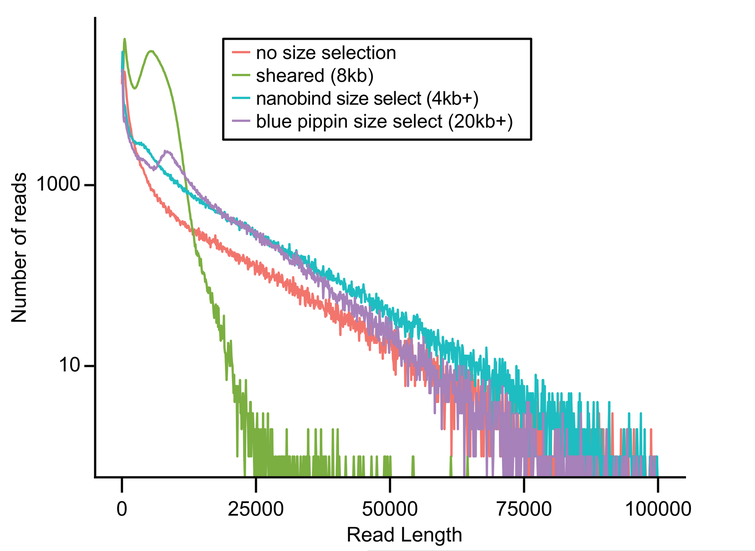


Given equal starting input, sequencing performance varied depending on extraction methodology (top left). First, the Zhang protocol produced highly fragmented reads, with a median read length below 1kb. Both Nanobind and Healey protocols gave reasonable read lengths, but don’t match the profile we observe via electrophoretic sizing. Importantly, the yield from nanobind is ~5X that of Healey. A possible explanation is provided examining the timing of sequencing (above right). We note that more of the pores are occupied with DNA (~60% vs ~20%) for Nanobind versus Healey. We hypothesize that residual impurities from Healey and Zhang protocols is precluding efficient delivery of DNA to the pore - perhaps the library prep or other steps are inhibited by comtaninants.



	10kb	25kb	50kb
Reads	500k	93.7k	94.7k
Yield	4.72Gb	0.82	0.66
N50	12.3kb	19.8	24
Median	9.8kb	4.1	1.7

We then tested how molecular size affects DNA yield (using Nanobind DNA) and discovered that 10kb sheared DNA (Diagenode Megaruptor) outperforms 25kb and 50kb sheared DNA. Our current working theory is that library prep and delivery of long molecules is not optimized



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

To further test this hypothesis, we tried size selection using either BluePippin (Sage) or Nanobind size select. Though both size selection methods acted to improve yield and median read length over unsheared., the yield was still substantially lower than sheared DNA.

References

[1]Healey, Adam, Agnelo Furtado, Tal Cooper, and Robert J. Henry. 2014. “Protocol: A Simple Method for Extracting next-Generation Sequencing Quality Genomic DNA from Recalcitrant Plant Species.” Plant Methods 10 (June):21.
[2]Zhang, Meiping, Yang Zhang, Chantel F. Scheuring, Cheng-Cang Wu, Jennifer J. Dong, and Hong-Bin Zhang. 2012. “Preparation of Megabase-Sized DNA from a Variety of Organisms Using the Nuclei Method for Advanced Genomics Research.” Nature Protocols 7 (3):467–78.

Acknowledgments

We gratefully acknowledge conversations with Daniela Puiu and Alison Scott which contributed to this work and our understanding. Thank you to Jen Lu for sequencing assistance. We are also grateful to Oxford Nanopore for extensive technical support and information. Funded by the Save the Redwoods League and NHGRI [1R01HG009190-01A1 (Timp) and 2R44GM109618-02 (Liu). Disclosure: Winston Timp has two patents licensed to ONT.