



Optimization of DNA Extraction and Size Selection for NGS Sequencing Across Plant Families Found in Dover Memorial Park



Olivia Schwartz, Dr. Jason Comer
Montana State University Billings | Billings, Montana

Research Questions:

- What are the current levels of plant biodiversity in Montana?
- How can NGS workflows be optimized for plant families with high levels of secondary compounds to maximize yields for analysis of biodiversity?

Abstract

Biodiversity can be explored in a variety of ways, from species richness to phylogenetic diversity. This project aims to investigate the plant biodiversity of Montana through analyses of species diversity (floristic collections) and phylogenetic diversity (next-generation sequencing [NGS]). Plants are well known for their secondary metabolites that interfere with downstream applications, such as DNA extraction and sequencing. To investigate phylogenetic diversity, optimized protocols for DNA extraction, fragmentation, and size selection need to be developed first. By optimizing extraction protocols, unique plant family characteristics will minimally affect yields and save time spent troubleshooting downstream applications. Plant specimens collected from Dover Memorial Park over the 2022 growing season were used to optimize a NGS workflow. This study found additional fragmentation of genomic DNA was unnecessary and automated size selection was sufficient to select the optimum fragment size range.

Premise

Global biodiversity is under threat due to the byproducts of human activity (urbanization/climate change) prompting calls for better understanding of current levels of biodiversity. Montana's plant biodiversity is poorly represented, and the overarching goal of this study is to document the floristic diversity of Montana (Fig. 1-3). Next-generation sequencing (NGS) offers high-throughput approaches to generate genomic data for studying phylogenetic diversity, but upstream steps often need optimization. DNA extraction and fragmentation/size selection are critical for NGS applications (Fig. 4). Protocols can be modified to provide higher quality yields, but have only been done for a few families (Snirc et al., 2010). Many NGS platforms require DNA fragmentation to generate DNA fragments of a specific size range for sequencing and must be optimized for the desired size range. This study presents the optimization of DNA extraction and size selection for downstream sequencing with the Oxford Nanopore Technologies (ONT) platform.

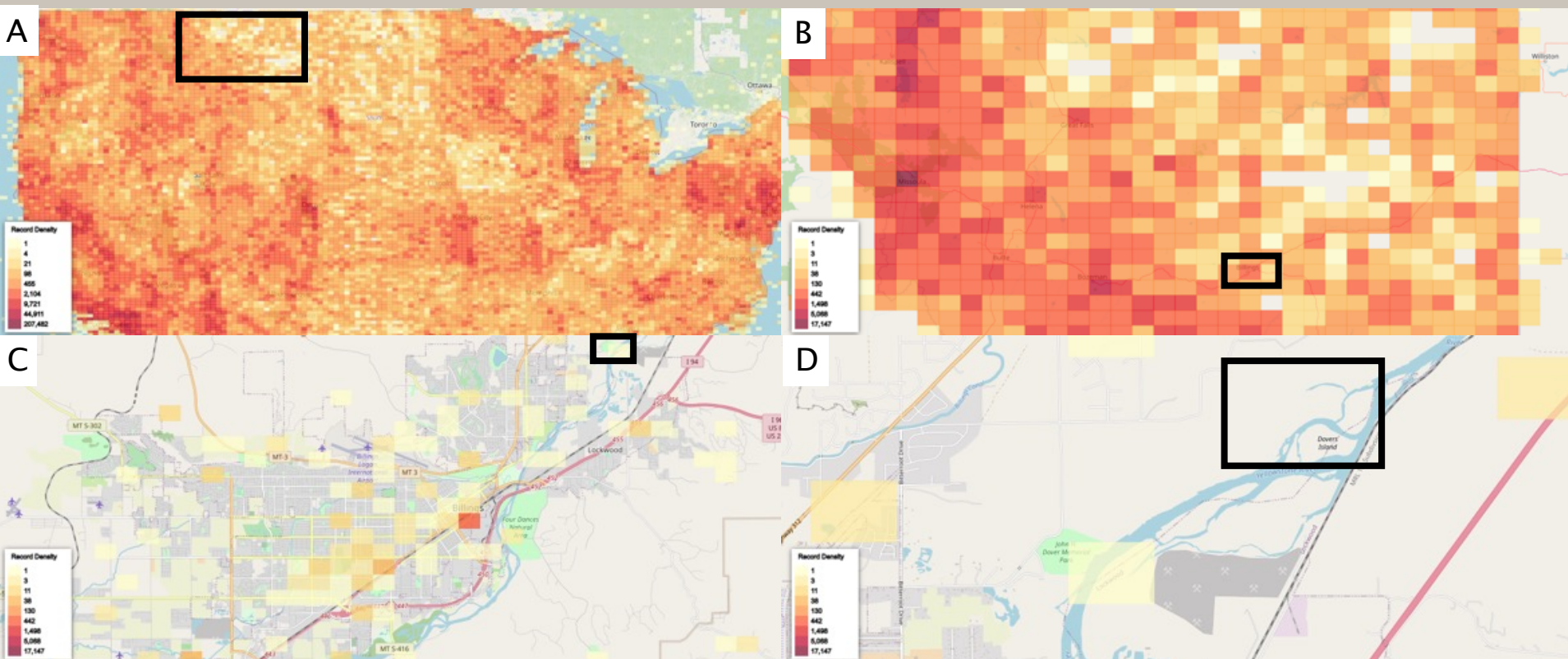


Figure 1: Heat maps showing the sampling intensity of embryophytes based on herbarium records. A) Continental US with Montana's approximate boundaries (black box). B) Montana with Billings approximate boundaries (black box). C) The city of Billings with the approximate boundaries Dover Island (black box). D) Approximate boundaries for Dover Island (black box) (left).
Figure 2 and 3: Images of collection areas at Dover Memorial Park in Billings, Montana (below).



Methods

Sampling – Voucher samples were collected from Dover Memorial Park (Figs. 2-3), Billings, MT during 2022 and additional tissue was preserved in silica gel for DNA extraction.
DNA extraction – DNA extraction was performed using the DNeasy Plant Pro kit (Qiagen, Germantown, Maryland, USA) which was previously shown to provide better results (Kersh, 2022). The manufacturer's protocol was used with the following modifications: 50 µL of PS solution and final elution in 50 µL of EB. DNA concentrations were quantified (Nanodrop) and gel electrophoresis was used to verify yield and quality (Fig. 5A).
Enzymatic fragmentation – A time series was used to determine the optimal incubation time to yield fragments within the 6 to 10kb range. NEBNext dsDNA fragmentase (New England BioLabs Inc., Ipswich, Massachusetts, USA) was used for fragmentation and the manufacturer's protocol was followed with variations in incubation time (3, 4, and 5 minutes) (Gautreau, 2016). Gel electrophoresis (Fig. 5B) and ONT sequencing were used to assess fragmentation (Fig. 6).
Size selection – The BluePippin (Sage Science Inc., Beverly, Massachusetts, USA) was used to perform automated size selection of DNA fragments between 6 and 10kb.
NGS Sequencing – To verify DNA fragment size distribution and effects of enzymatic fragmentation the Oxford Nanopore Technologies NGS platform (New York, New York, USA) was used. Library preparation and sequencing was performed with Minlon Mk1C with the Flongle adapter and flow cells (version R9.4.1) following the manufacturer's protocols. Six runs were completed using samples that were high molecular weight, naturally fragmented and enzymatically fragmented (Fig. 6).

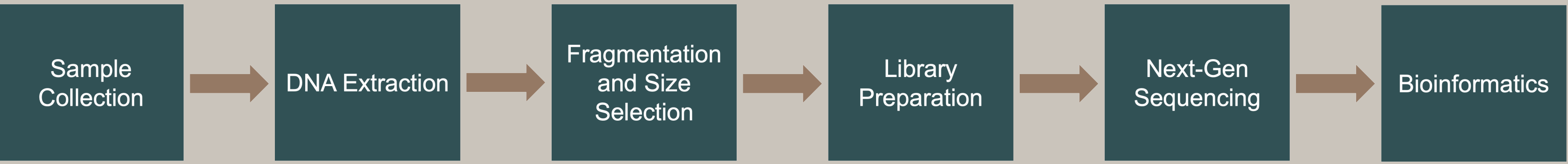
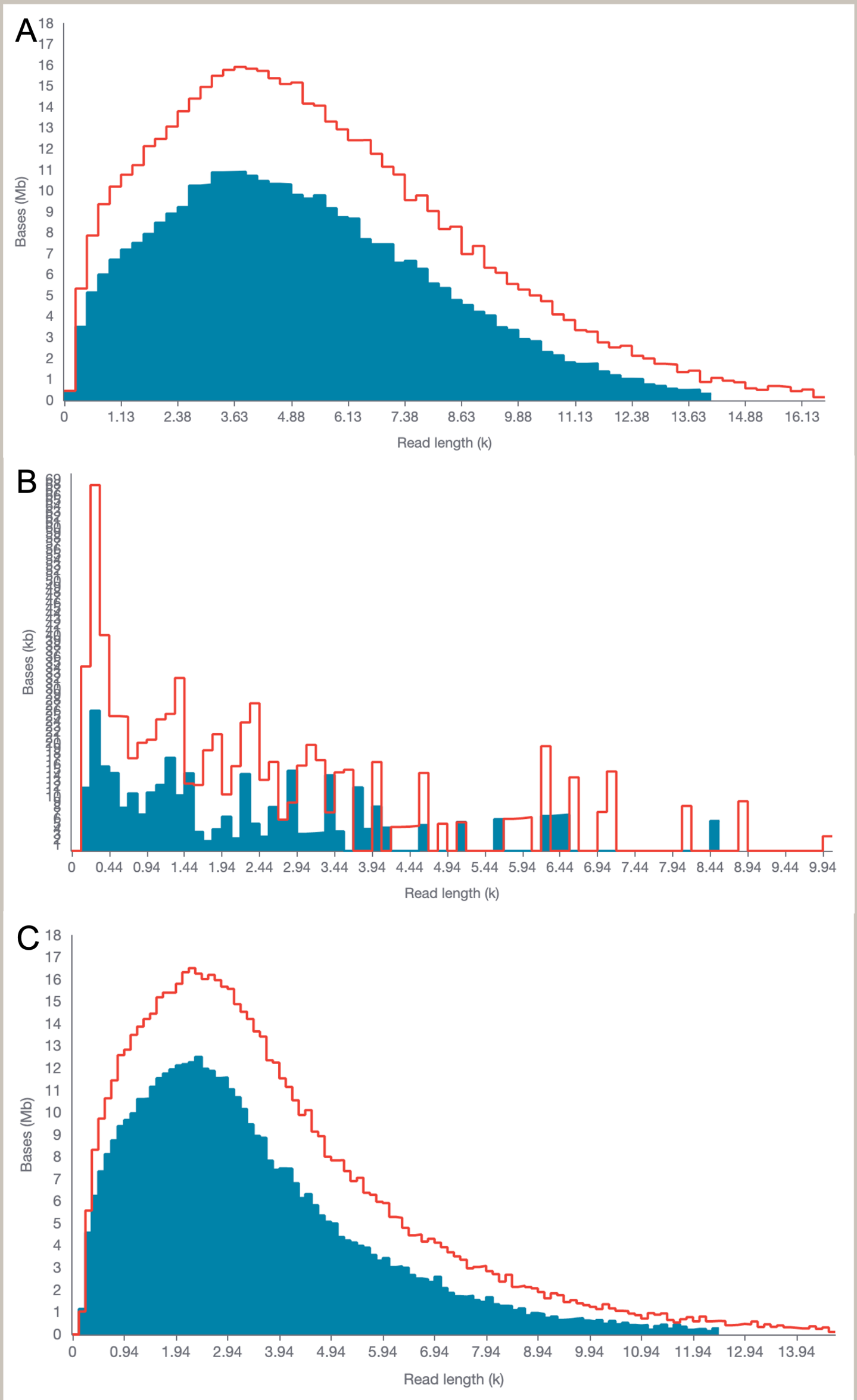
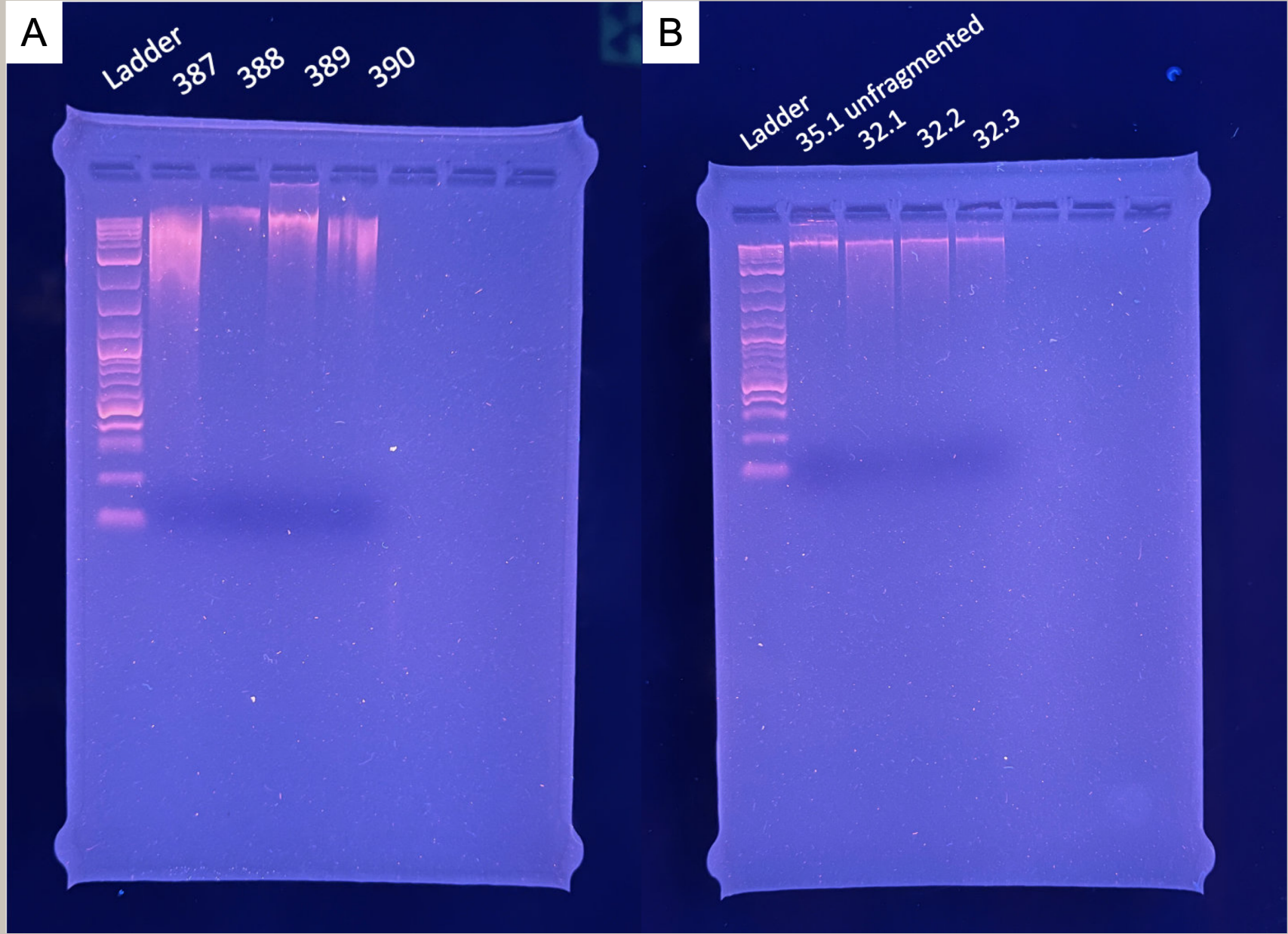


Figure 4: Workflow depicting the methods being optimized for long-term studies

Results

Table 1. This table provides an example of the ranges of DNA concentrations and purities from different species collected at Dover Memorial Park.			
Collection Number	ng/µL	260/230	260/280
387	265.10	1.85	1.96
388	33.30	1.90	0.85
390	60.20	1.82	0.85
391	27.20	2.09	0.64
392	30.30	2.16	1.02
393	76.00	1.81	1.41
394	69.50	1.82	1.29
408	9.50	2.13	1.40
409	3.10	2.66	0.55

Figure 5: Electrophoresis gels. A) Gel showing the variation in quality of DNA samples from Dover Park (below left). B) Gel showing results from the enzymatic fragmentation time series (below right).
Figure 6: A) Output from ONT sequencing depicting fragment size distribution for an intact sample using long fragment buffer. B) Output from ONT sequencing depicting fragment size distribution for an enzymatically fragmented sample. C) Output from ONT sequencing depicting fragment size distribution for a naturally fragmented sample (right).



Discussion

DNA extraction – Previous research (Kersh, 2022) determined that the DNeasy Plant Pro kit was the most suitable extraction kit, but further modifications were needed for taxonomically varied species collected from Dover Memorial Park. DNA extraction protocols were modified to increase DNA yield for certain species with high levels of secondary compounds or dense trichomes (PS solution and tissue selection). The optional PS solution was determined to be necessary, especially for samples with high phenolics (e.g., Cactaceae and Lamiaceae). DNA concentrations varied across the samples while using modified protocols, but overall were more effective in producing consistently high yields, average for all samples 53.51ng/µL (Table 1).
Enzymatic fragmentation and size selection – The enzymatic fragmentation time series showed that the best time for yielding fragments between 6 and 10 kb was 4 minutes (Fig. 5B). However, DNA fragmentation was deemed to be unnecessary as the DNA extraction process naturally shears the DNA (Fig 5A). There was adequate shearing (Figs. 6-8) that automated size selection was sufficient for isolating DNA fragments of the desired size range (6-10kb, data not shown).

Conclusion

This study provides a basis for future studies to effectively extract DNA and determine phylogenies in Dover Memorial Park. Optimization of DNA extraction and size selection protocols will continue. Future studies will leverage the methods described here to produce genomic data for Dover Memorial Park and other Montana plant communities. This genomic data will be combined with species inventories to assess plant biodiversity in Montana.

Acknowledgements

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number **P20GM103474**. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This material is based upon work supported by the National Science Foundation under Grant No. OIA-1757351. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Citations

Ammar, I., M. Ennouri, and H. Attia. 2015. Phenolic content and antioxidant activity of cactus (*Opuntia ficus-indica* L.) flowers are modified according to the extraction method. *Industrial Crops and Products* 64: 97–104.
Gautreau, I. 2016. Digestion with NEBNext dsDNA Fragmentase (M0348).
Hermans, S. M., H. L. Buckley, and G. Lear. 2018. Optimal extraction methods for the simultaneous analysis of DNA from diverse organisms and sample types. *Molecular Ecology Resources* 18: 557–569.
IDigBio Specimen Search. IDigBio Specimen Portal. [Online]. Available: <http://portal.idigbio.org/portal/search>. [Accessed: 24-Jul-2021].
Ignatov, K. B., K. A. Blagoderzhik, D. S. Shcherba, T. V. Kramarova, Y. A. Monakhova, and V. M. Kramarov. 2019. Fragmentation Through Polymerization (FTP): A new method to fragment DNA for next-generation sequencing. *PLoS ONE* 14: e0210374.
Kämper, W., S. J. Trueman, J. Cooke, N. Kasinadhuni, A. J. Brunton, and S. M. Ogbourne. 2021. Single-nucleotide polymorphisms that uniquely identify cultivars of avocado (*Persea americana*). *Applications in Plant Sciences* 9: e11440.
Kersh, J. 2022. Preliminary phylogenomic assessment of *Convolvulus arvensis* L., a widespread noxious weed'. [Poster]. 2022 MSUB Research and Creativity Conference, April 22 2022, Montana State University Billings.
Kolmogorov, M., J. Yuan, Y. Lin, and P. A. Pevzner. 2019. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* 37: 540–546.
Martin, S., D. Heavens, Y. Lan, S. Horsfield, M. D. Clark, and R. M. Leggett. 2022. Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *Genome Biology* 23: 11.
McLay, T. G. B., J. L. Birch, B. F. Gunn, W. Ning, J. A. Tate, L. Nauheimer, E. M. Joyce, et al. 2021. New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Applications in Plant Sciences* 9.
miniPCR. 2016. Nanopore sequencing – DNAdots by miniPCR. Website <https://dnadots.minipcr.com/dnadots/nanopore-sequencing> [accessed 31 October 2022].
Payne, A., N. Holmes, T. Clarke, R. Munro, B. Debebe, and M. Loose. 2021. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nature Biotechnology* 39: 442–450.
QIAAGEN. 2019. DNeasy Plant Pro Kit Handbook. Website <https://www.qiagen.com/us/resources/resourceDetail?id=71236f8b-84ba-4926-a7d8-913c91ab4a30&lang=en> [accessed 9 April 2023].
Snirc, A., T. Silberfeld, J. Bonnet, A. Tiller, S. Tuffel, and Jian-Sheng Sun. 2010. Optimization of Dna Extraction from Brown Algae (phaeophyceae) Based on a Commercial Kit. *Journal of Phycolgy* 46: 616–621.
Syar, O., M. Zveak, K. Konate, and M. Brestic. 2022. Phenolic Acid Patterns in Different Plant Species of Families Asteraceae and Lamiaceae: Possible Phylogenetic Relationships and Potential Molecular Markers. *Journal of Chemistry* 2022: e9632979.
Tretlakov, M. Yu., V. K. Tokhtar, A. E. Ileva, J. N. Kurkina, I. V. Lyashenko. 2022. Optimizing the Dry Plant Sample Preparation Method For Efficient DNA Extraction in *Syringa vulgaris*. *International Journal of Ecosystems and Ecology Science (IJEES)* 12: 201–206.