



HAL
open science

Extraction and selection of high-molecular-weight DNA for long-read sequencing from *Chlamydomonas reinhardtii*

Frédéric Chaux, Nicolas Agier, Stephan Eberhard, Zhou Xu

► **To cite this version:**

Frédéric Chaux, Nicolas Agier, Stephan Eberhard, Zhou Xu. Extraction and selection of high-molecular-weight DNA for long-read sequencing from *Chlamydomonas reinhardtii*. PLoS ONE, 2024, 19 (2), pp.e0297014. 10.1371/journal.pone.0297014 . hal-04765776

HAL Id: hal-04765776

<https://hal.science/hal-04765776v1>

Submitted on 4 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LAB PROTOCOL

Extraction and selection of high-molecular-weight DNA for long-read sequencing from *Chlamydomonas reinhardtii*Frédéric Chaux^{1*}, Nicolas Agier¹, Stephan Eberhard², Zhou Xu^{1*}

1 CNRS, UMR7238, Institut de Biologie Paris-Seine, Laboratory of Computational and Quantitative Biology, Sorbonne Université, Paris, France, **2** CNRS, UMR7141, Institut de Biologie Physico-Chimique, Laboratory of Chloroplast Biology and Light-Sensing in Microalgae, Sorbonne Université, Paris, France

* zhou.xu@sorbonne-universite.fr (ZX); fchauj@irb.hr (FC)

OPEN ACCESS

Citation: Chaux F, Agier N, Eberhard S, Xu Z (2024) Extraction and selection of high-molecular-weight DNA for long-read sequencing from *Chlamydomonas reinhardtii*. PLoS ONE 19(2): e0297014. <https://doi.org/10.1371/journal.pone.0297014>

Editor: Ramachandran Srinivasan, Sathyabama Institute of Science and Technology, INDIA

Received: August 25, 2023

Accepted: December 26, 2023

Published: February 8, 2024

Copyright: © 2024 Chaux et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All sequencing data, which include raw Nanopore FAST5 files and read FASTQ files, and genome assemblies (as FASTA files) have been submitted to the European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/ena/browser/home>) under the project accession number PRJEB59713.

Funding: Research in ZX's laboratory was supported by ANR grant "AlgaTelo" (ANR-17-CE20-0002-01; <https://anr.fr/>) and by Ville de Paris (Programme Émergence(s); <https://www.paris.fr/>)

Abstract

Recent advances in long-read sequencing technologies have enabled the complete assembly of eukaryotic genomes from telomere to telomere by allowing repeated regions to be fully sequenced and assembled, thus filling the gaps left by previous short-read sequencing methods. Furthermore, long-read sequencing can also help characterizing structural variants, with applications in the fields of genome evolution or cancer genomics. For many organisms, the main bottleneck to sequence long reads remains the lack of robust methods to obtain high-molecular-weight (HMW) DNA. For this purpose, we developed an optimized protocol to extract DNA suitable for long-read sequencing from the unicellular green alga *Chlamydomonas reinhardtii*, based on CTAB/phenol extraction followed by a size selection step for long DNA molecules. We provide validation results for the extraction protocol, as well as statistics obtained with Oxford Nanopore Technologies sequencing.

Introduction

In recent years, long-read sequencing technologies, such as the ones developed by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (Nanopore), have emerged as a solution to the pitfalls of short-read technologies in the detection of structural variants and in assembling repeated sequences and other complex regions [1]. Additionally, because native DNA is used, long-read technologies can directly detect a variety of modified bases, including the most commonly studied methylated cytosines [2, 3]. For their applications in genome assembly and structural variant detection, these technologies typically sequence DNA molecules ranging in size from kilobases to hundreds of kilobases as a continuous read. Reads traversing repeated sequences are necessary to correctly assemble neighboring regions, with longer reads enabling more contiguous genome assemblies. Today, the major bottleneck to sequence long reads comes from the ability to extract high-quality DNA devoid of polyphenol and polysaccharide contaminants with sizes compatible with this purpose. This is especially true for most plant tissues and algae cells, because polyphenols and polysaccharides are often co-extracted with DNA and can inhibit downstream applications such as sequencing [4, 5].

appels-a-projets). SE was supported by the “Initiative d’Excellence” program of the French State (‘DYNAMO’, ANR-11-LABX-0011-01; <https://anr.fr/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Chlamydomonas reinhardtii is a unicellular green alga that is widely used as a model organism to study photosynthesis and cellular motility [6], and is an organism of choice for biotechnological application, with many synthetic biology tools being currently developed [7, 8]. In *C. reinhardtii*, as for other plants and algae, contending with phenolic and polysaccharide contaminants while preserving HMW DNA is a major challenge and requires an optimized protocol. PacBio and Nanopore sequencing have been performed on this organism, contributing to important advances in our understanding of its genome structure and content, base modifications and evolution [9–16]. However, it appears that a size selection step can substantially enrich for longer molecules, as noted in [14, 15] and as we demonstrate in this work. An efficient and well documented protocol is therefore needed for sequencing projects that require long DNA molecules.

Here, we present a detailed protocol dedicated to efficiently extract and select HMW DNA from *C. reinhardtii* cells. The protocol minimizes DNA-shearing manipulations [17] and comprises an additional step to enrich for HMW DNA. We validated the method by pulse-field gel electrophoresis (PFGE) and measurement of read length from Nanopore sequencing.

Materials and methods

The protocol described in this peer-reviewed article is published on protocols.io, [dx.doi.org/10.17504/protocols.io.8epv59j9jg1b/v2](https://doi.org/10.17504/protocols.io.8epv59j9jg1b/v2) and is included for printing purposes as [S1 File](#).

Nanopore sequencing

Sequencing libraries were prepared as per manufacturer’s recommendations, using NEBNext companion module (E7180S, NEB) and Ligation Sequencing Kit SQK LSK-109 (Nanoporetech), except for the ligation time, which we increased to 30 min. For each run, 500 ng were loaded on MinION flow cells (R9.4.1, Nanoporetech) and sequenced for 6h to 16h, depending on flow-cell kinetics. Libraries were loaded at least twice, with 1h wash using the manufacturer’s washing buffer (EXP-WSH004) between runs. Basecalling was performed using Guppy (version 4.3.4) with parameters set to “high accuracy”.

Results

We extracted genomic DNA following the presented protocol ([S1 File](#)) and applied size selection using the Short Read Eliminator (SRE) kit (Circulomics), an easy-to-use method that does not require dedicated devices which is based on a length-dependent precipitation of nucleic acids driven by polyvinylpyrrolidone crowding. Large amounts of small DNA fragments can be detrimental for long-read Nanopore sequencing [18], not only because the subsequent reads are short, but also because these molecules can outcompete the longer ones, both for adapter ligation and pore usage, thus yielding suboptimal results.

The size distribution of the extracted DNA was assessed by PFGE and Nanopore sequencing, with and without size-selection for HMW DNA. Samples were migrated in a pulse field, stained by ethidium bromide and imaged with UV light ([Fig 1A](#)). The DNA molecules extracted without size selection migrated as a large smear spread between approximately 1.5 and 150 kb. After size selection with the SRE kit, the upper part of the distribution remained unchanged while the low-molecular-weight fragments (< 10 kb) were visibly reduced. We made a similar observation after electrophoresis and staining of the samples in a 0.3% agarose gel ([Fig 1B](#)).

Size-selection of DNA fragments before preparation of libraries for Nanopore sequencing led to a substantially decreased number of shorter molecules and an enrichment of longer ones ([Fig 2A and 2B](#)), without negatively affecting read quality ([Fig 2D](#)) and with no effect on

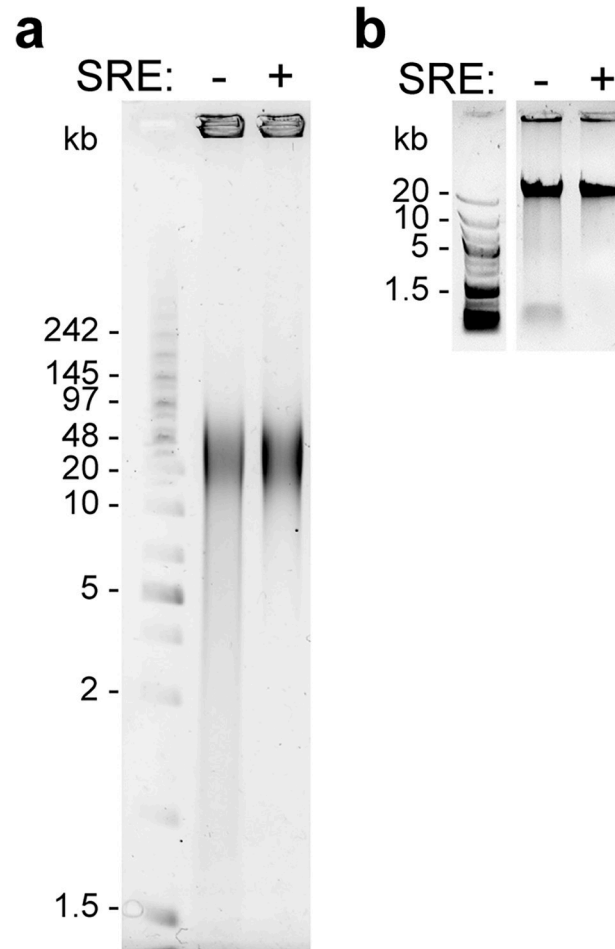


Fig 1. Visualization of extracted genomic DNA size distributions. (a) PFGE using 0.5 μ g of DNA prepared with (+) or without (-) SRE size-selection, embedded in 30 μ l of 0.5% low-melting agarose plugs, migrated in a 1% SeaKem GTG agarose (Lonza) gel. The ladder is a mix of PFG mid-range (N0342S, NEB) and GeneRuler 1 kb Plus (SM1331, ThermoFischer). Electrophoresis conditions: 0.5X TBE (Tris Borate EDTA) buffer, 6 $V \cdot cm^{-1}$, 120° angle, for 11h, switching time ramp from 1 to 60 seconds. Gel stained in ethidium bromide and imaged with UV. (b) Standard gel electrophoresis (0.3% agarose) of the indicated samples. GeneRuler 1 kb Plus (SM1331, ThermoFischer) is used as the ladder. See S3 Fig for the uncropped images.

<https://doi.org/10.1371/journal.pone.0297014.g001>

genome-wide sequencing depth (S1 Fig). Size-selection doubled the mean read length, increased the N50 from 12 kb to 17 kb, with reads in the top decile being longer than 21 kb (S1 Table). The length distribution after size-selection was robust across different experiments using two other independent biological samples, and reached an N50 of up to 20 kb and a top decile length of up to 27 kb (Fig 2C and S1 Table). The longest molecules we sequenced were over 100 kb, which are instrumental for genome assemblies. Indeed, we recently assembled the genome of *C. reinhardtii* based on these reads and found a genome size between 114 and 117.7 Mb [15], depending on the assembler, which is consistent with the 114 Mb of the recently released version 6 of the reference genome [16]. Overall, this protocol and the resulting quality and length of the DNA molecules are suitable for reaching highly contiguous genome assemblies.

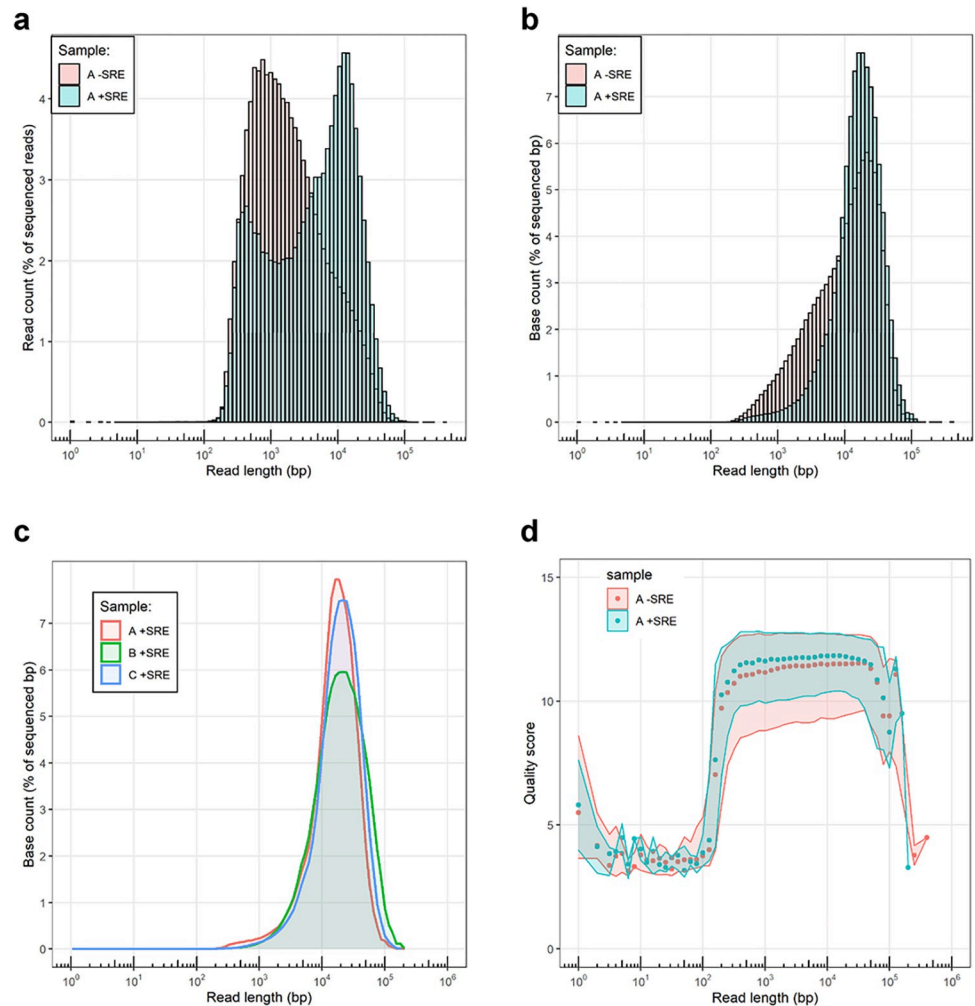


Fig 2. Distributions of read length in Nanopore-sequenced datasets. (a, b) Count percentage of (a) reads and of (b) bases as a function of read length obtained from genomic DNA of *C. reinhardtii* (experiment “A”, see S1 Table) with or without size selection (+SRE and -SRE). (c) Count of bases after size-selection (+SRE) as a function of read length obtained from three different biological samples (see S1 Table and S2 Fig). (d) Quality score for individual reads, grouped into bins of 0.1 log unit for samples “A-SRE” and “A+SRE”. The shaded areas represent the values between the 1st and 3rd quartiles.

<https://doi.org/10.1371/journal.pone.0297014.g002>

Supporting information

S1 File. Step-by-step protocol, also available on protocols.io: [dx.doi.org/10.17504/protocols.io.8epv59j9jg1b/v2](https://doi.org/10.17504/protocols.io.8epv59j9jg1b/v2).

(PDF)

S1 Table. Summary statistics for 6 DNA preparations and sequencing experiments. Major limiting outputs are shown in red. ^a <https://www.chlamylibrary.org> and reference [19]. ^b with quality > 7. ^c as per manufacturer’s protocol (Monarch® HMW DNA Extraction Kit for Tissue Cat. no. T3060L, New England Biolabs). ^d cell lysis using DNeasy Maxi Plant (Cat. no. 68163, Qiagen) as in [20] and purification using Genomic-tip 100/G (Cat. no. 10243, Qiagen), then AMPure beads (Cat. no. A63880, Beckman Coulter).

(PDF)

S1 Fig. Genome-wide sequencing depth normalized to the median, for all chromosomes, using DNA obtained with (+) or without (-) SRE size selection.

(PDF)

S2 Fig. Count percentage of bases as a function of read length with alternative sample preparations without size selection (-SRE). See S1 Table for details. Sample C was sequenced in the presence of control DNA ("DNA CS" from Oxford Nanopore sequencing), which peaked at 3 kb.

(PDF)

S3 Fig. Raw images.

(PDF)

Acknowledgments

We thank Samuel O'Donnell for his help in the initial development of this protocol.

Author Contributions

Conceptualization: Frédéric Chauv, Zhou Xu.

Formal analysis: Frédéric Chauv.

Investigation: Frédéric Chauv, Nicolas Agier, Stephan Eberhard.

Methodology: Frédéric Chauv, Nicolas Agier, Stephan Eberhard.

Supervision: Zhou Xu.

Writing – original draft: Frédéric Chauv, Zhou Xu.

Writing – review & editing: Frédéric Chauv, Nicolas Agier, Stephan Eberhard, Zhou Xu.

References

1. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet.* 2020; 21(10):597–614. <https://doi.org/10.1038/s41576-020-0236-x> PMID: 32504078
2. Rand AC, Jain M, Eizenga JM, Musselman-Brown A, Olsen HE, Akeson M, et al. Mapping DNA methylation with high-throughput nanopore sequencing. *Nat Methods.* 2017; 14(4):411–3. <https://doi.org/10.1038/nmeth.4189> PMID: 28218897
3. Feng Z, Fang G, Korlach J, Clark T, Luong K, Zhang X, et al. Detecting DNA modifications from SMRT sequencing data by modeling sequence context dependence of polymerase kinetic. *PLoS Comput Biol.* 2013; 9(3):e1002935. <https://doi.org/10.1371/journal.pcbi.1002935> PMID: 23516341
4. Porebski S, Bailey LG, Baum BR. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant molecular biology reporter.* 1997; 15(1):8–15.
5. Healey A, Furtado A, Cooper T, Henry RJ. Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods.* 2014; 10:21. <https://doi.org/10.1186/1746-4811-10-21> PMID: 25053969
6. Harris EH. *The Chlamydomonas Sourcebook*: Elsevier/Academic Press; 2009.
7. Scaife MA, Nguyen G, Rico J, Lambert D, Helliwell KE, Smith AG. Establishing *Chlamydomonas reinhardtii* as an industrial biotechnology host. *Plant J.* 2015; 82(3):532–46. <https://doi.org/10.1111/tpj.12781> PMID: 25641561
8. Crozet P, Navarro FJ, Willmund F, Mehrshahi P, Bakowski K, Lauersen KJ, et al. Birth of a Photosynthetic Chassis: A MoClo Toolkit Enabling Synthetic Biology in the Microalga *Chlamydomonas reinhardtii*. *ACS Synth Biol.* 2018; 7(9):2074–86. <https://doi.org/10.1021/acssynbio.8b00251> PMID: 30165733
9. O'Donnell S, Chauv F, Fischer G. Highly Contiguous Nanopore Genome Assembly of *Chlamydomonas reinhardtii* CC-1690. *Microbiol Resour Announc.* 2020; 9(37). <https://doi.org/10.1128/MRA.00726-20> PMID: 32912911

10. Liu Q, Fang L, Yu G, Wang D, Xiao CL, Wang K. Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. *Nat Commun.* 2019; 10(1):2449. <https://doi.org/10.1038/s41467-019-10168-2> PMID: 31164644
11. Chaux-Jukic F O'Donnell S, Craig RJ, Eberhard S, Vallon O, Xu Z. Architecture and evolution of subtelomeres in the unicellular green alga *Chlamydomonas reinhardtii*. *Nucleic Acids Res.* 2021; 49(13):7571–87. <https://doi.org/10.1093/nar/gkab534> PMID: 34165564
12. Craig RJ, Hasan AR, Ness RW, Keightley PD. Comparative genomics of *Chlamydomonas*. *Plant Cell.* 2021. <https://doi.org/10.1093/plcell/koab026> PMID: 33793842
13. Lopez-Cortegano E, Craig RJ, Chebib J, Balogun EJ, Keightley PD. Rates and spectra of de novo structural mutations in *Chlamydomonas reinhardtii*. *Genome Res.* 2023; 33(1):45–60. <https://doi.org/10.1101/gr.276957.122> PMID: 36617667
14. Payne ZL, Penny GM, Turner TN, Dutcher SK. A gap-free genome assembly of *Chlamydomonas reinhardtii* and detection of translocations induced by CRISPR-mediated mutagenesis. *Plant Commun.* 2023; 4(2):100493. <https://doi.org/10.1016/j.xplc.2022.100493> PMID: 36397679
15. Chaux F, Agier N, Garrido C, Fischer G, Eberhard S, Xu Z. Telomerase-independent survival leads to a mosaic of complex subtelomere rearrangements in *Chlamydomonas reinhardtii*. *Genome Res.* 2023; 33(9):1582–98. <https://doi.org/10.1101/gr.278043.123> PMID: 37580131
16. Craig RJ, Gallaher SD, Shu S, Salome PA, Jenkins JW, Blaby-Haas CE, et al. The *Chlamydomonas* Genome Project, version 6: Reference assemblies for mating-type plus and minus strains reveal extensive structural mutation in the laboratory. *Plant Cell.* 2023; 35(2):644–72. <https://doi.org/10.1093/plcell/koac347> PMID: 36562730
17. Kovacic RT, Comai L, Bendich AJ. Protection of megabase DNA from shearing. *Nucleic Acids Res.* 1995; 23(19):3999–4000. <https://doi.org/10.1093/nar/23.19.3999> PMID: 7479050
18. Delahaye C, Nicolas J. Sequencing DNA with nanopores: Troubles and biases. *PLoS One.* 2021; 16(10):e0257521. <https://doi.org/10.1371/journal.pone.0257521> PMID: 34597327
19. Li X, Zhang R, Patena W, Gang SS, Blum SR, Ivanova N, et al. An Indexed, Mapped Mutant Library Enables Reverse Genetics Studies of Biological Processes in *Chlamydomonas reinhardtii*. *Plant Cell.* 2016; 28(2):367–87. <https://doi.org/10.1105/tpc.15.00465> PMID: 26764374
20. Eberhard S, Valuchova S, Ravat J, Fulnecek J, Jolivet P, Bujaldon S, et al. Molecular characterization of *Chlamydomonas reinhardtii* telomeres and telomerase mutants. *Life science alliance.* 2019; 2(3). <https://doi.org/10.26508/lsa.201900315> PMID: 31160377