Benchmarks

Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules

Baptiste Mayjonade[1], Jérôme Gouzy[1], Cécile Donnadieu[2], Nicolas Pouilly[1], William Marande[3], Caroline Callot[4], Nicolas Langlade[1], and Stéphane Muños[1]

[1]LIPM, Université de Toulouse, INRA, CNRS, Castanet-Tolosan, France, [2]Get-PLAGE, Université de Toulouse, INRA, CNRS, Castanet Tolosan, France, [3]CNRGV, Université de Toulouse, INRA, CNRS, Castanet Tolosan, France, and [4]CRCT, INSERM, Université de Toulouse, CNRS, Toulouse, France

Supplementary material for this article is available at www.BioTechniques.com/article/114460.

De novo sequencing of complex genomes is one of the main challenges for researchers seeking high-quality reference sequences. Many de novo assemblies are based on short reads, producing fragmented genome sequences. Third-generation sequencing, with read lengths >10 kb, will improve the assembly of complex genomes, but these techniques require high-molecular-weight genomic DNA (gDNA), and gDNA extraction protocols used for obtaining smaller fragments for short-read sequencing are not suitable for this purpose. Methods of preparing gDNA for bacterial artificial chromosome (BAC) libraries could be adapted, but these approaches are time-consuming, and commercial kits for these methods are expensive. Here, we present a protocol for rapid, inexpensive extraction of high-molecular-weight gDNA from bacteria, plants, and animals. Our technique was validated using sunflower leaf samples, producing a mean read length of 12.6 kb and a maximum read length of 80 kb.

Next-generation sequencing (NGS) is commonly used in genomics and, until recently, only produced sequences hundreds of nucleotides long. The genomes of higher eukaryotes, especially plants, contain a large number of repeated sequences that can reach several tens of kilobases in length. De novo assemblies of large complex genomes based on short reads result in fragmented genome sequences containing thousands to millions of contigs or scaffolds (1). Third-generation sequencers based on the sequencing of a single DNA molecule produce sequences of several tens of kilobases, which can improve genome assemblies because the longer reads can often span repeated regions.

High-quality and high-molecular-weight DNA (>50 kb) is needed to fully benefit from the potential of third-generation sequencers. The polymerase used for sequencing (DNA/Polymerase Binding Kit P6 v2; Pacific Biosciences, Menlo Park, CA) is highly sensitive to contaminants bound to DNA, which could result in shorter sequences.

While preparation of megabase-sized DNA in an agarose gel matrix is possible (2), the process takes 3 days, and it is difficult to extract the DNA from the agarose plug. Commercial kits for the extraction of high-molecular-weight DNA are also available, but they are expensive.

Here we describe a protocol (see Supplementary Material) to extract micrograms of high-molecular-weight genomic DNA (gDNA) up to 150 kb length in a single microtube in under 60 min at a cost of <$0.20 per sample (Figure 1). We fine-tuned this protocol for sunflower leaf samples and subsequently validated it for bacteria and human cells. Sunflowers are known to contain polyphenols and other contaminants, such as polysaccharides and tannins (3); therefore, it is challenging to extract very pure DNA from this plant species.

We harvested leaves from 3-week-old plantlets. The samples were immediately frozen in liquid nitrogen and stored at -80°C until grinding to prevent DNA degradation.

The first step in the gDNA extraction consisted of mechanically and chemically disrupting cell membranes and (for plants) cells walls to release nuclei and DNA molecules. We ground the leaves into a fine powder in liquid nitrogen. For the lysis step, we used an SDS-based buffer to dissolve the cell membranes (4,5). The detergent SDS neutralizes and inactivates proteins, especially endonucleases, that degrade DNA. During this step, we also included RNase A in the lysis buffer to digest RNA. During the lysis step, it is crucial to immediately homogenize the frozen ground tissue in the lysis buffer to quickly inactivate nucleases. The degree of homogenization of the lysate is also correlated with the yield

METHOD SUMMARY

We present a protocol for the rapid (~1 hr), inexpensive, and efficient isolation of high-molecular-weight genomic DNA (gDNA) that can be used to prepare libraries for long-read sequencing on the PacBio RSII instrument.

Table 1. Quality and quantity of genomic DNA (gDNA) samples.

Protocol	Sample	Agarose gel lane (Figure 2)	Repetition	Ethanol washes	Magnetic bead resuspension in ethanol	Concentration determined by Qubit (ng/µl)	Concentration determined by Nanodrop (ng/µl)	$A_{260/280}$	$A_{260/230}$	DNA quantity: Qubit-based (ng)	Sample amount
Genomic Tips	sunflower leaves	2	1	NA	NA	98	106.7	1.9	2.34	4900	46.4 mg FW
Genomic Tips	sunflower leaves	3	2	NA	NA	109	123.4	1.87	2.34	5450	42.4 mg FW
Genomic Tips	sunflower leaves	4	3	NA	NA	18.6	20	1.84	1.74	930	42.6 mg FW
This article	sunflower leaves	7	1	2	no	151	131	1.83	1.81	9060	89.8 mg FW
This article	sunflower leaves	8	2	2	no	148	118	1.83	1.73	8880	90.5 mg FW
This article	sunflower leaves	9	3	2	no	137	116	1.87	1.67	8220	91.7 mg FW
This article	sunflower leaves	10	1	2	yes	130	111	1.84	2.13	7800	94.4 mg FW
This article	sunflower leaves	11	2	2	yes	123	112	1.83	2.13	7380	94.8 mg FW
This article	sunflower leaves	12	3	2	yes	113	102	1.83	2.11	6780	86 mg FW
This article	human cells	15	1	2	yes	206	182.2	1.84	2.26	41,200	confluent cells
This article	human cells	16	2	2	yes	228	190.7	1.84	2.21	45,600	confluent cells
This article	human cells	17	3	2	yes	90.2	93.3	1.82	2.12	18,040	confluent cells
This article	E. coli	19	1	2	yes	51	39.3	1.92	2.16	3060	1 mL of culture OD = 4.5
This article	E. coli	20	2	2	yes	60	39.4	1.96	2.16	3600	1 mL of culture OD = 4.5
This article	E. coli	21	3	2	yes	52	39.7	1.97	2.10	3120	1 mL of culture OD = 4.5

The quality and concentration of the gDNA were measured by spectrometry (Nanodrop 2000; ThermoFisher, Waltham, MA) and fluorometry (Qubit 3.0 Fluorometer; ThermoFisher). OD: optical density. FW: fresh weight.

of the extraction process and the purity of the DNA. In some cases, heating the lysis buffer can help to homogenize the solution. However, heating the sample above 65°C for more than 1 h can lead to DNA damage.

At the end of the lysis step, the addition of potassium acetate to the lysis solution leads to the formation of potassium dodecyl sulfate, which is insoluble. Proteins and polysaccharides then form a complex with the potassium dodecyl sulfate precipitate, which is then removed by centrifugation (6).
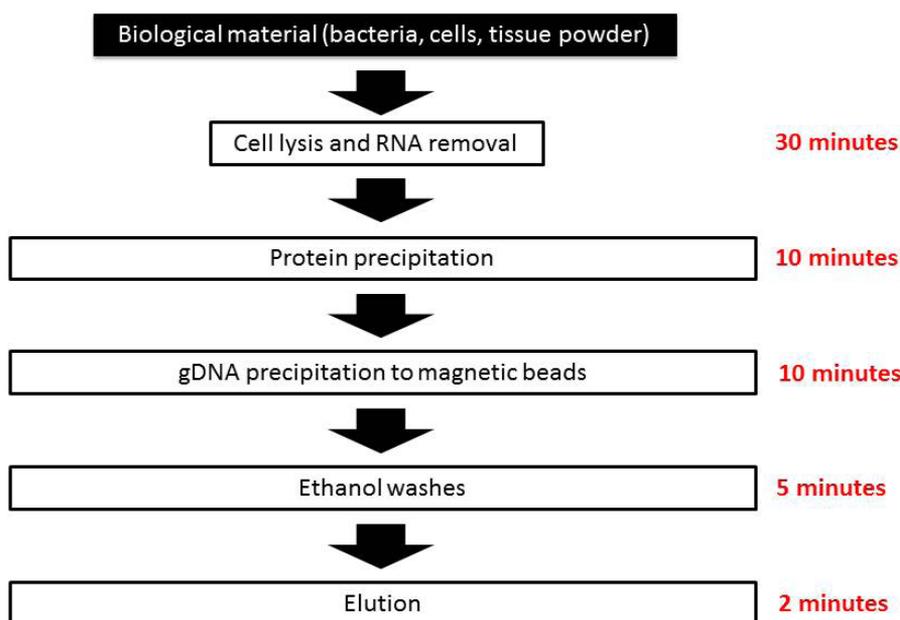
The second step is purification of DNA from the lysis solution. Carboxylated magnetic beads were used because they are cost-effective (7) and cause minimal DNA shearing. Polyethylene glycol 8000 (PEG8000) was used to bind DNA to the beads, as previously described (7,8). PEG8000 is safe and precipitates fewer contaminants compared with isopropanol or ethanol (9).

For the final step, gDNA was washed in 70% ethanol to remove contaminants that were poorly bound to the magnetic beads as well as the remaining reagents used in the previous steps. DNA was eluted from the beads by resuspension in a buffered solution. We used a spectrophotometric method to evaluate the purity of our DNA solutions. The $A_{260/230}$ ratio is affected by the presence

Table 2. Characteristics of PacBio sequences from sunflower leaf libraries with different size cutoffs.

g-TUBE fragmentation parameters on Eppendorf centrifuge #5424	BluePippin size-selection cutoff (kb)	Movie length (min)	Total bases (Mb/SMRT cell)	Polymerase read length (bp)	N50 subread (bp)	Max subread (bp)
6000 g/1 min	12	240	1106.01	13,940	12,866	50,298
5000 g/1 min	15	240	1109.57	14,740	15,100	49,331
4000 g/2 min	18	360	1057.60	15,836	19,206	80,974
3800 g/3 min	20	360	1026.39	17,126	20,507	73,379

N50 is the value for which 50% of all bases come from subreads longer than this value. g-TUBEs were purchased from Covaris (Woburn, MA). The BluePippin instrument was purchased from Sage Science (Beverly, MA).
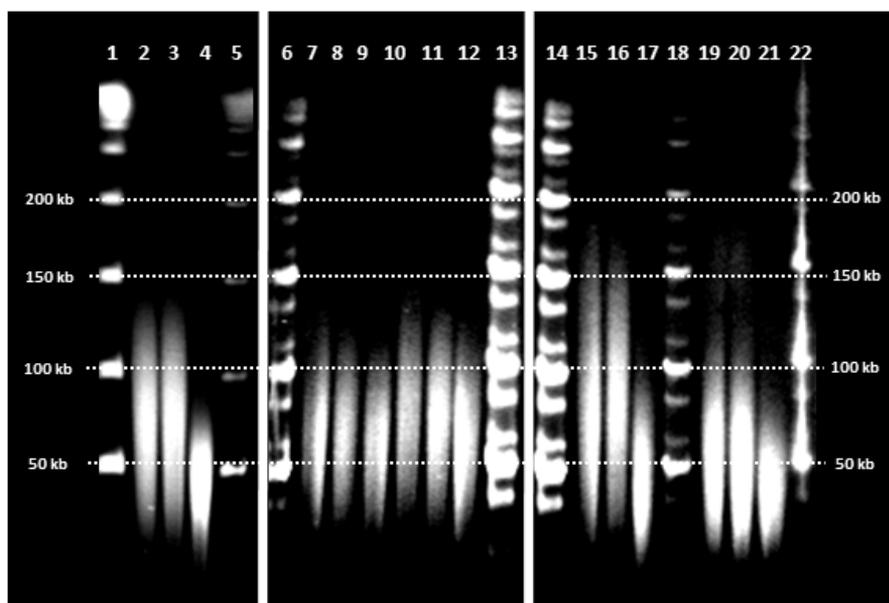

Figure 1. Genomic DNA (gDNA) extraction workflow.

Figure 2. Pulsed field agarose gel electrophoresis of genomic DNA (gDNA). The integrity of gDNA, extracted by different protocols was evaluated on field electrophoresis agarose gels. Lanes 1 and 5: lambda PFG ladder; Lanes 6,13,14,18, and 22: MidRange I PFG marker; Lanes 2–4: sunflower gDNA extracted with QIAGEN Genomic Tips; Lanes 7–9: sunflower gDNA extracted with our protocol (two ethanol washes without bead resuspension); Lanes 10–12: sunflower gDNA extracted with our protocol (two ethanol washes with bead resuspension); Lanes 15–17: HEK293 human cell gDNA extracted with our protocol; Lanes 19–21: NEB 5-alpha electrocompetent *E. Coli* gDNA extracted with our protocol.

of certain contaminants. We found that the resuspension of the beads in 70% ethanol was crucial for complete removal of contaminants that absorb light at 230 nm (Table 1). Two additional washes could be performed without decreasing the integrity of the DNA (data not shown). We obtained an $A_{260/280}$ ratio >1.8, indicating that our DNA solution was protein-free. The concentrations quantified by either spectrophotometry or fluorescent intercalating dye–based methods (Table 1) were similar, indicating that the DNA solution was pure. Our extractions yielded 6.7–7.8 µg of highly pure DNA from 90 mg of fresh sunflower leaves.

We evaluated the integrity of the extracted gDNA by performing pulsed-field electrophoresis, which showed that the DNA molecules ranged in size from ~20 kb to ~130 kb (Figure 2).

Comparing DNA solutions extracted using our protocol to DNA extracted with the Genomic Tips 20/G commercial kit (QIAGEN, Hilden, Germany), it can be seen that the purity and integrity of the gDNA were similar for both protocols (Table 1 and Figure 2). However, the commercial kit is 35 times more expensive ($7 per sample) and requires at least 3 h to complete the extraction, compared with 1 h for our protocol.

We next evaluated the suitability of our protocol for gDNA extraction from bacteria and animal cells (*E. coli* and human cell culture, respectively). High-molecular-weight gDNA, with similar characteristics to the sunflower gDNA we obtained, was successfully extracted (Table 1 and Figure 2).

Finally, we validated the quality of sunflower gDNA extracted with our protocol by sequencing with the PacBio RS II (Pacific Biosciences, Menlo Park, California) using 4 different PacBio libraries with 4 DNA size cutoffs ranging 12–20 kb (Table 2). No inhibition of the sequencing reaction was observed because the sequence reads were consistent with PacBio specifications (10–15 kb for 240-min-long movies and up to 17,126 bp for 360-min-long movies). These sequencing results confirmed that the gDNA was pure and of high molecular weight. According to Pacific Biosciences specifications and the activity of the P6 polymerase, the maximum read length that can be obtained is limited to 46,000 bp with a 240-min-long movie. With a 240-min-long movie, we obtained sub-reads (i.e., inserts of gDNA) of up to 50,298 bp, confirming the quality of the gDNA extracted with our protocol. The maximal length of the sub-reads increased to 80,974 bp with a movie length of 360 min.

Author contributions

Acknowledgments

Competing interests

The authors declare no competing interests.

References

1. Treangen, T.J. and S.L. Salzberg. 2011. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat. Rev. Genet. *13*:36-46.

2. Zhang, M., Y. Zhang, C.F. Scheuring, C.C. Wu, J.J. Dong, and H.B. Zhang. 2012. Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. Nat. Protoc. *7*:467-478.

3. Horne, E.C., S.P. Kumpatla, K.A. Patterson, M. Gupta, and S.A. Thompson. 2004. Improved high-throughput sunflower and cotton genomic DNA extraction and PCR fidelity. Plant Mol Biol Report. *22*:83-84.

4. Dellaporta, S.L., J. Wood, and J.B. Hicks. 1983. A plant DNA minipreparation: version II. Plant Mol Biol Report. *1*:19-21.

5. Tai, T. H. and S.D. Tanksley. 1990. A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue. Plant Mol Biol Report. *8*:297-303.

6. Bryant, J.L., P.M. Dey, and J.B. Harborne. 1996. Methods in Plant Biochemistry, Volume 10B: Molecular Biology. Molecular Biology. Academic Press, Cambridge, MA.

7. Rohland, N. and D. Reich. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. *22*:939-946.

8. Hawkins, T.L., T. O'Connor-Morin, A. Roy, and C. Santillan. 1994. DNA purification and isolation using a solid-phase. Nucleic Acids Res. *22*:4543-4544.

9. Shan, G., W. Jin, E.K. Lam, and X. Xing. 2008. Purification of total DNA extracted from activated sludge. J. Environ. Sci. (China) *20*:80-87.

To purchase reprints of this article, contact:
biotechniques@fosterprinting.com