NEB Monarch® HMW DNA Extraction Kit improves sample preparation for Oxford Nanopore Technologies sequencing of malaria parasites

Mitchell J. Godin and Scott E. Lindner, Ph.D., Department of Biochemistry and Molecular Biology, The Huck Center for Malaria Research, The Center for Eukaryotic Gene Regulation, Pennsylvania State University

INTRODUCTION

Advancements in long-read DNA sequencing by Pacific Biosciences® (PacBio®) and Oxford Nanopore Technologies® have led to major breakthroughs in *de novo* assembly. An advantage of both PacBio and Oxford Nanopore sequencing is that they provide exceptionally long sequence reads, but in order to achieve those, the use of high quality, high molecular weight (HMW) DNA fragments as an input material is required. With many existing genomic DNA (gDNA) extraction kits, this prerequisite for optimal Oxford Nanopore sequencing is not met, resulting in less than ideal read and contig lengths and a greater number of contigs overall, contributing to suboptimal de novo assemblies. These issues have led to a return to classic molecular biology methods (e.g., phenol/chloroform extractions, gel plug extractions (1,2)), and have increased demand for new, streamlined methods of isolating and purifying HMW gDNA.

Recently, New England Biolabs® released the Monarch® HMW DNA Extraction Kit for Cells & Blood, and we sought to compare this new method with our standard approach (Qiagen® QIAamp® DNA Blood Mini Kit) to produce HMW gDNA of high quality from two species of malaria parasites. In this work, we used both *Plasmodium falciparum* NF54, a common laboratory strain of a human-infectious malaria parasite, and *Plasmodium yoelii* 17XNL, a non-lethal laboratory strain of a rodent-infectious malaria parasite. For both of these eukaryotic pathogens of red blood cells, we have observed substantial improvements in the extraction of HMW gDNA using the NEB kit when compared with other methods. Phenol/chloroform extractions have yielded impure samples that require extensive and sometimes iterative purification steps. Extractions using the Qiagen kit result in lower molecular weight gDNA samples that cannot fully leverage the power of Nanopore sequencing. In contrast, the NEB Monarch HMW DNA Extraction Kit for Cells & Blood has allowed for the rapid production of high quality, high molecular weight gDNA that is ready for Nanopore sequencing and other applications. We anticipate that this approach will make the whole genome sequencing and assembly of genomes from a variety of organisms (including *Plasmodium* field isolates and transgenic laboratory strains) routine and reproducible.

METHODS

Culturing of *P. falciparum* NF54 Parasites

Plasmodium falciparum (NF54 strain) parasites were cultured in standard conditions at 37°C, 5% CO_2 / 5% O_2 / 90% N, 4% hematocrit (Type O blood) using filter-sterilized RPMI 1640 media supplemented with 0.25% w/v Albumax II, 25mM HEPES pH 7.4 at RT, 23.8 mM sodium bicarbonate, 0.1mM hypoxanthine, and 0.05 mg/ml gentamycin. Collected parasites were then subjected to the same protocols as described below for *P. yoelii* samples.

MATERIALS

- Saponin (0.1% w/v)
- 1X PBS
- Heparin
- Cellulose Column (Sigma #C6288)
- Monarch HMW DNA Extraction Kit for Cells & Blood (NEB #T3050)
- Qiagen QIAamp Blood Mini Kit (Cat #51104)
- Genomic DNA ScreenTapes (Cat #5067-5366 and 5067-5365)
- Qubit dsDNA BR Assay Kit (Cat #Q32853)
- Oxford Nanopore Technologies Rapid Sequencing Kit (SQK-RAD004)
- Oxford Nanopore Technologies Flow Cell Wash Kit (EXP-WSH004)
- Oxford Nanopore Technologies Flow Cell R9.4.1 (FLO-MIN106D)

White Blood Cell Removal and Optional Saponin Lysis

A 1 ml sample of mouse blood infected with *P. yoelii* (17XNL strain) was treated with heparin and mixed with 5 ml of 1X PBS. The blood sample was split in half to provide identical, matched samples for comparison of the NEB Monarch HMW DNA Extraction Kit for Cells &

Blood (NEB #T3050) and the Qiagen QIAamp DNA Blood Mini Kit (Cat #51104). The blood and PBS mixture was passed through a cellulose column (Sigma #C6288) to remove mouse leukocytes and washed with 1X PBS. In the case of the Qiagen QIAamp DNA Blood Mini Kit protocol, blood was lysed in 0.1% w/v saponin in 1X PBS with subsequent washing in 1X PBS to release the parasites from the erythrocyte membrane.

DNA Extraction with Qiagen QIAamp Blood Mini Kit

DNA was extracted according to the manufacturer's protocol with slight modifications to maximize yield. Briefly, a saponin-lysed parasite pellet was washed in 1X PBS, and then resuspended in 200 µl 1X PBS. Next, 10 µl Proteinase K and 200 µl Buffer AL (supplied in the kit) were added, and samples were incubated at 56ºC for 10 minutes. Addition of 200 µl 100% reagent grade ethanol to the sample was included to promote gDNA binding to the column. Samples were bound to the column, washed twice with 500 µl Buffer AW1 and once with 500 µl Buffer AW2. Samples were centrifuged at 6000 x *g* for 1 min between all steps. A final spin at maximum speed for 1 minute was carried out to remove residual wash buffer. The column was transferred to a fresh microfuge tube, and gDNA was eluted with 50 µl Elution Buffer AE by centrifugation at 6000 x *g* for 1 min. To assess DNA purity and quantity, Thermo Fisher Scientific® Nanodrop® 2000 and Qubit® instruments with the Qubit dsDNA BR Assay Kit (Cat #Q32853) were used, respectively. DNA Integrity Number (DIN) and a primary measurement of DNA fragment length were carried out using an Agilent Technologies® TapeStation® 4200 system with Genomic DNA ScreenTapes (Cat #5067-5366 and 5067-5365).

DNA Extraction with NEB Monarch HMW DNA Extraction Kit: Blood Protocol

Genomic DNA was extracted from leukocyte-depleted blood samples using NEB Monarch HMW DNA Extraction Kit for Cells & Blood samples following the manufacturer's protocol for fresh blood samples, with slight modifications. The NEB workflow includes a red blood cell lysis buffer for erythrocyte lysis which enables the isolation of a parasite pellet, and thus does not require prior saponin lysis. In brief, 500 µl of infected blood was washed with 15 ml 1X PBS, pelleted and resuspended in 500 µl 1X PBS, and *Plasmodium* parasites were then released from erythrocytes by addition of three volumes of the RBC Lysis Buffer with mixing by inversion. The sample was pelleted, washed in 1X PBS and then treated with the 150 µl Nuclei Prep and 150 µl Nuclei Lysis solutions as directed. Parasite lysis was conducted at 56ºC for 10 minutes in an Eppendorf® ThermoMixer® at 300 rpm to produce the highest molecular weight fragments for the downstream Oxford Nanopore Technologies Rapid Sequencing Kit workflow. For ligation-based nanopore sequencing, agitation at the maximum of 2000 rpm is recommended by NEB. The lysed material was then mixed with Precipitation Enhancer, two DNA Capture Beads, and isopropanol. The gDNA was captured on the beads by slow end-over-end rotation. The liquid was removed from the mixed solution by pipetting, taking care not to disturb the DNA wrapped around beads. The gDNA Wash Buffer was then added to the sample, and was mixed slowly by inversion. This wash was repeated once more to efficienctly wash the sample. Elution was carried out with the Bead Retainer, as directed, using 100 µl Elution Buffer II at 56ºC for 5 minutes in the ThermoMixer at 300 rpm. Eluted gDNA was separated from the beads by centrifugation, and wide bore pipette tips were used to transfer the gDNA to a final microfuge tube for storage and to disperse any aggregates. DNA purity, concentration, and fragment lengths were assessed as described above.

Preparation of gDNA for Oxford Nanopore Technologies Sequencing

A Rapid Sequencing Kit (SQK-RAD004) was utilized using *P. yoelii* DNA isolated from either the NEB or Qiagen kit with 61.4 ng (NEB) or 28.8 ng (Qiagen) purified genomic DNA used for library construction for the tagmentation step. Nanopore sequencing was completed with no modifications to the manufacturer's protocol with an R9.4.1 flow cell. The Qiagen sample was run first for 20 hours, then the flow cell was washed using the Oxford Nanopore Technologies Flow Cell Wash kit (EXP-WSH004) and associated protocol. Thereafter, the NEB sample was run for 20 hours.

Data Analysis of Rapid Sequencing Reads

For both the Qiagen and the NEB sample types, the same data analysis pipeline was used to permit a direct comparison of sequencing results. During the sequencing run, the most recent version of the Guppy base caller was used. The resulting fastQ files were analyzed in NanoPlot (3) to determine the initial sequencing run statistics. Nanopore adapters and low quality reads were trimmed and filtered, with the raw reads analyzed by the program fastp with default parameters (4). The results were then run through the NanoPlot program a second time to determine the success of post-processing by fastp. A *de novo* assembly was then performed using the Flye program with default parameters (5). To compile this information, a basic fasta summary statistics tool from the Bandage program was used to determine number of contigs, assembly N50, and total assembly length (6).

RESULTS

The amount of *Plasmodium* DNA obtained from the two *Plasmodium* species tested was around 1 µg for both extraction kits used (Table 1). Most samples showed a high A260/A280 ratio, indicating the presence of RNA. The Qubit-based concentration values were significantly lower than the absorbance-based values, which also indicates the presence of RNA, with the Qiagen samples displaying a higher level of RNA contamination (Table 1). Fragment length quantification by TapeStation demonstrated that the majority of gDNA fragments from the NEB samples were longer than the 60,000 bp upper limit of detection of the assay and had maximal DIN values (9.9 for Py17XNL and 9.7 for PfNF54) (Figure 1, Table 1). DNA fragments obtained with the Qiagen kit were significantly more sheared, with peak sizes of around 30 kb and 50 kb and DINs of 7 for Py17XNL and 7.8 for PfNF54 (Figure 2, Table 1).

The mean read length of trimmed and quality filtered reads from NEB-produced HMW gDNA was more than double that of the Qiagen gDNA, which resulted in triple the estimated N50 contig value, which is a measure of the median contig length used in the assembly (Table 2). Even with the significantly increased read length, the average quality of the reads produced from gDNA from the NEB kit remained higher than those from the Qiagen kit. The mean quality remained unchanged through fastp processing steps, indicating higher quality raw reads were obtained with the NEB sample than what was obtained with the Qiagen sample (Table 2).

Without any consensus polishing or short read error correction, the fastp processed reads from the NEB gDNA sample could be used to assemble a genome with far fewer gaps compared to the Qiagen gDNA sample (Table 2). Additionally, the NEB assembly length was significantly closer to the anticipated genome size of *P. yoelii* (~23 Mb), resulting in a remarkably better assembly overall (Table 2).

TABLE 1: **Quality control values from Nanodrop, Qubit and TapeStation assessments of gDNA**

NEB- and Qiagen-extracted DNA samples from two *Plasmodium* species were analyzed on for yield (Qubit), purity (Nanodrop), peak size and DINs (TapeStation).

		QUANTITY MEASUREMENTS (QUBIT)		PURITY (NANODROP)		TAPESTATION ANALYSIS	
SAMPLE	DNA EXTRACTION METHOD	CONCENTRATION (ng/µl)	YIELD (µg)	A260/A280	A260/A230	PEAK SIZE (bp)	DIN
Py17XNL	Qiagen	24.0	1.2	2.38	1.97	29,903	7.0
Py17XNL	NEB	6.1	0.61	1.73	1.05	>60,000	9.9
PfNF54	Qiagen	7.1	0.36	2.09	2.86	49,816	7.8
PfNF54	NEB	12.6	1.26	2.02	2.67	>60,000	9.7

FIGURE 1: **An example data analysis pipeline for Nanopore DNA sequencing of gDNA**

FIGURE 2: **TapeStation analysis of extracted DNA**

Comparative TapeStation electrophoresis of gDNA produced from Qiagen Blood Amp and NEB HMW kits from A) *P. yoelii* 17XNL or B) *P. falciparum* NF54.

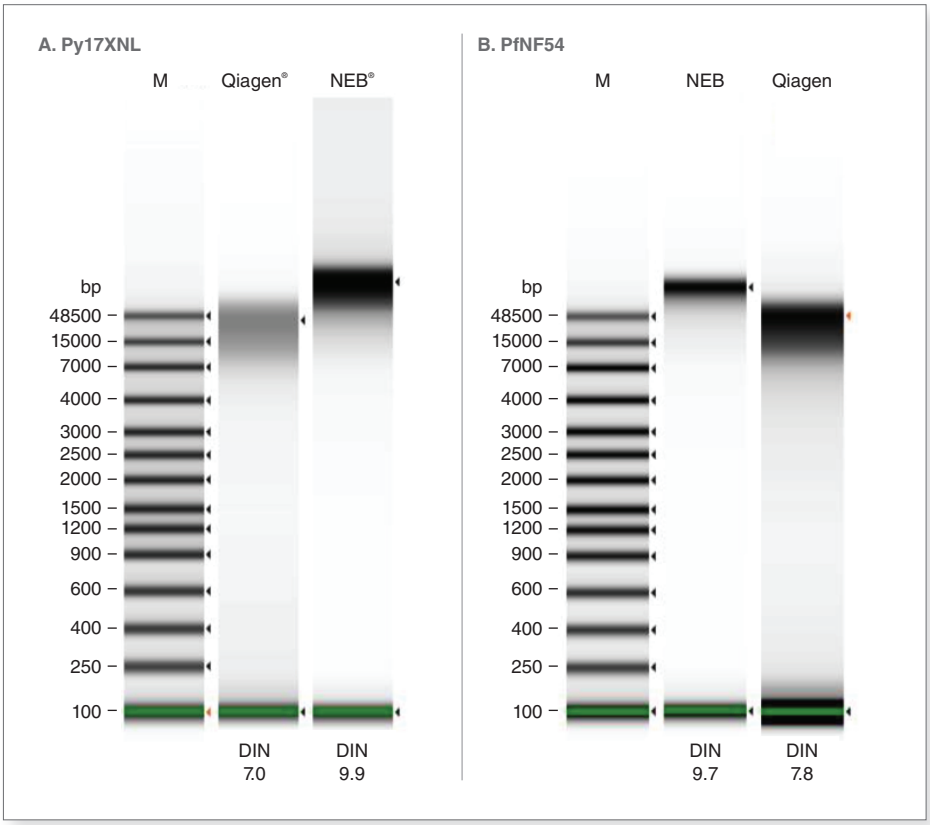

TABLE 2: **Nanopore rapid sequencing statistics before and after quality filtration and adapter trimming by fastp**

Assembly done using Flye program with default parameters. For *P. yoelii*, the minimum number of contigs possible is 16 (14 chromosomes and 2 organellar sequences) and expected total assembly length is ~23 million bps.

	QIAGEN		NEB	
	PRE-PROCESSING	POST-PROCESSING	PRE-PROCESSING	POST-PROCESSING
Total Reads (bp)	1,248,893	733,050	245,837	189,244
Total Bases (bp)	2,203,146,627	1,447,884,969	1,267,987,926	952,647,904
Mean Read Length (bp)	1,764.10	1,975.20	1,963.00	5,034.00
Mean Read Quality (Qscore)	10.5	11.5	12	12
Read Length N50	3,620.00	4,056.00	13,594.00	13,494.00
Reads >Q10 (%)	62.90	97.20	98.50	98.50
Reads >Q12 (%)	17.20	29.20	48.70	49.30
Longest Read (bps)	76,294	76,294	163,006	163,006
Flye Assembly N50	N/A	1,411,486	N/A	1,647,171
Flye Assembly Length (bp)	N/A	20,476,473	N/A	22,545,676
Flye Assembly # of Contigs	N/A	77	N/A	54

DISCUSSION AND CONCLUSION

The NEB kit workflow led to better purity (less RNA) as well as significantly larger fragment length and DIN than the routinely-employed *Plasmodium* DNA purification workflow based on Qiagen's QIAamp Kit, as demonstrated by TapeStation. Although purity assessment based of the purity ratios is unreliable with samples <25 ng/µl, Qubit readings confirmed the likely presence RNA contamination. The presence of significant RNA amounts in the samples can be explained by the absence of an RNase digestion in the Qiagen workflow and the placement of RNase treatment step before lysis of the *Plasmodium* cells in the NEB protocol. In both cases, RNA content in the eluates can likely be reduced by adding an RNase digestion step at the end of the Proteinase K digestion.

For the Nanopore rapid sequencing workflow, the raw reads from the NEB sample proved to be longer and of consistently higher quality. We anticipate that further improvements in Nanopore sequencing results will be produced by using Oxford Nanopore Technologies Ligation Sequencing Kits, and by using corresponding faster shaking speeds with the NEB kit. Upon processing these reads through adapter trimming and quality filtration, the assembly N50 contig (a metric the median length of contiguous sequences used in the assembly), mean read length, and Q Scores for the NEB gDNA are significantly greater than that obtained from Qiagen purified gDNA. When using the NEB sample for *de novo* assembly with the Flye assembler, the NEB fastp processed reads generated an assembled consensus genome with the fewest gaps and greatest fraction of the genome sequences even without any consensus polishing or error correction.

With the ease of the sample prep using the NEB Monarch HMW DNA Extraction kit including an erythrocyte lysis step and the high quality of the isolated HMW DNA, we anticipate that this workflow for Nanopore DNA sequencing will make whole-genome sequencing of eukaryotic *Plasmodium* parasites a streamlined and routine process in dedicated laboratories.

References

1. Sambrook, J. & Russell, D.W. & Cold Spring Harbor Laboratory (2001) *Molecular cloning : a laboratory manual*. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory.
2. Hicks, L., et al. (2018) *J. Biol. Methods.* 5, e86.
3. De Coster, W., et al. (2018) *Bioinformatics.* 34, 2666–2669.
4. Chen, S., et al. (2018) *Bioinformatics.* 34, i884–i890.
5. Kolmogorov, M., et al. (2019) *Nat. Biotechnol.* 37, 540–546.
6. Wick, R.R., et al. (2015). *Bioinformatics*, 31, 3350-3352.

nqa ISO 9001 Registered Quality Management
nqa ISO 14001 Registered Environmental Management
nqa ISO 13485 Registered Medical Devices

www.neb.com

NEW ENGLAND BioLabs® Inc.

be INSPIRED
drive DISCOVERY
stay GENUINE

New England Biolabs, Inc., 240 County Road, Ipswich, MA 01938-2723 Telephone: (978) 927-5054 Toll Free: (USA Orders) 1-800-632-5227 (USA Tech) 1-800-632-7799 Fax: (978) 921-1350 e-mail: info@neb.com