Extraction of High Molecular Weight (HMW) Genomic DNA for Long-Read NGS Applications

Andy Zegers, Don Smith, Douglas Horejsh

Promega Corporation, 2800 Woods Hollow Rd, Madison, WI 53711

1. Abstract

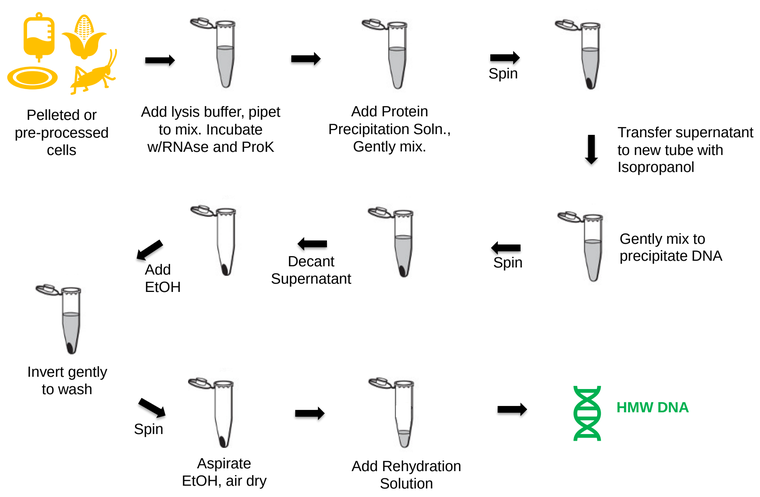
Genetics and disease research has often been limited by the technologies and tools available to researchers. Tools such as DNA sequencing have been the primary means researchers employ to attempt to gain a better understanding not only of the form and content of the human genome, but also of the genetic differences that can be found in cases of unusual phenotypes and disease.

The field of DNA sequencing has evolved, growing to include technologies that can provide enormous amounts of sequence data. Certain technologies have struggled to provide data over large or complex regions in particular, and to do so with high accuracy.

The past several years have shown tremendous progress in both read length and accuracy. Platforms such as the PacBio® Sequel II System and the Oxford Nanopore Technologies ® MinIon® offer high single-base accuracy as well as read lengths in excess of 2Mb.

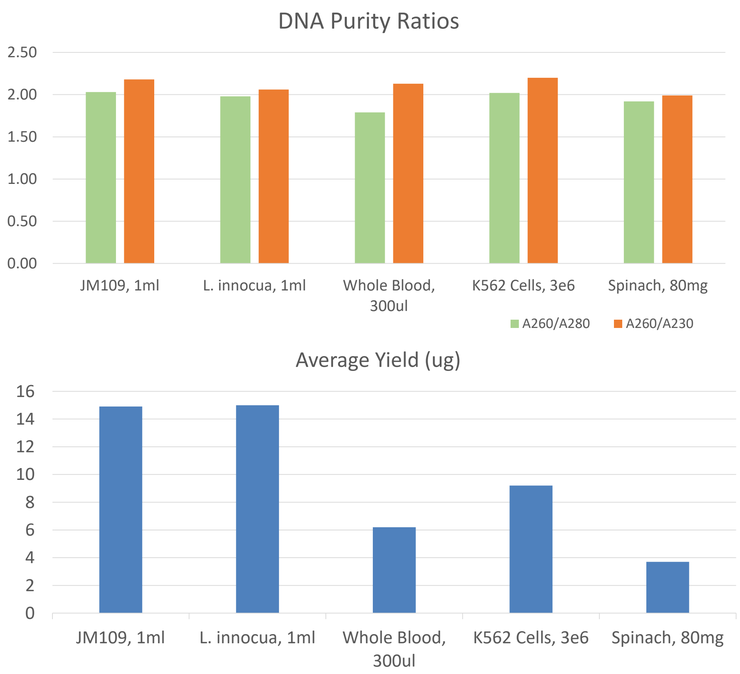
The success of these platforms depends greatly on the ability to isolate and preserve large DNA fragments, commonly referred to as High Molecular Weight (HMW) DNA. Common methods are extremely resource-intensive, costly, or provide DNA that does not perform well in sequencing. In this study we demonstrate a new method that offers improvements in the ability to extract DNA that yields high performance in Long Read Sequencing (LRS) without unnecessary burdens in workflow or reagent cost.

2. Simple, Gentle Workflow to Preserve HMW DNA



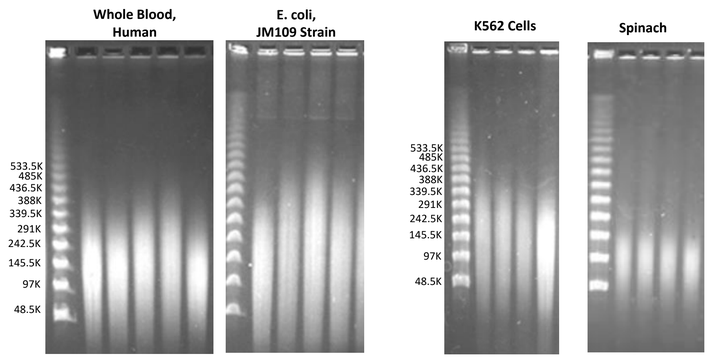
The Promega Wizard® HMW DNA extraction workflow employs a precipitation-based approach with enhancements and proprietary reagents that contribute to maximizing the size of DNA that is obtained. Sample compatibility includes whole blood, cultured mammalian cells, Gram-positive bacteria, Gram-negative bacteria, and plant leaf tissue. Certain minor adaptations are made in early processing and lysis steps, after which a gentle purification procedure yields high-mass DNA that is highly pure and performed well in long-read sequencing.

3. High Yields and Purities



A spectrum of samples (n=8) were extracted and analyzed by spectrophotometry for purity, and by fluorometry (Promega QuantiFluor® dsDNA System) for concentration and yield. Purities were consistently very high across sample types, with yields varying according to input sample type and amount.

4. HMW DNA Size Analysis by PFGE

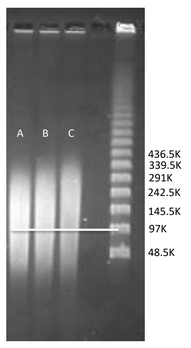


Following extraction, DNA fragment size was evaluated using Pulsed-Field Gel Electrophoresis (PFGE). PFGE gels are is useful for approximating the fragment size range of HMW DNA samples, which do not properly migrate or separate using traditional gel electrophoresis methods.

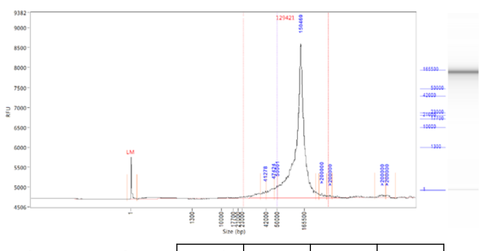
Using the Promega HMW extraction method, large fragments were isolated from a variety of samples. DNA fragment size varies by sample type, with fragments in excess of 500kb obtained in some cases. From fresh whole blood, overnight *E. coli* cultures, and freshly cultured mammalian cells, DNA in excess of 250kb was routinely obtained. The results with spinach are representative of extractions from plant samples, which require freezing with liquid nitrogen and subsequent processing with mortar and pestle. This yields somewhat smaller but still very large DNAs.

5. Large fragment size analysis by CE

Differences in extraction method performance can be seen not only on PFGE gels, but also on more precise instruments such as the Agilent® Femto Pulse System. The Wizard® HMW method is compared to a competing method in this respect below.



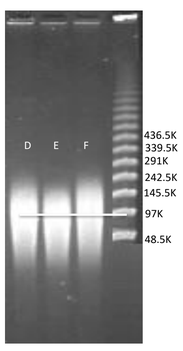
Wizard HMW

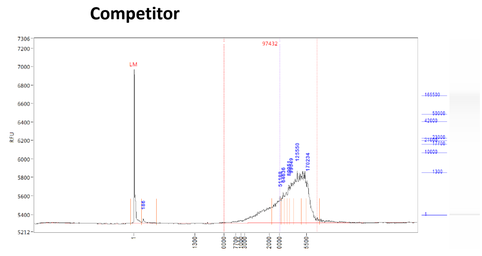


Example Femto Pulse trace for Samples A-C (above) and calculated GQN figures (right)

Sample	GQN 50	GQN 97	GQN 125
A	8.7	7.4	6.1
B	8.7	7.6	6.3
C	8.4	7.4	6.4

Competitor





Example Femto Pulse trace for Samples D-F (above) and calculated GQN figures (right)

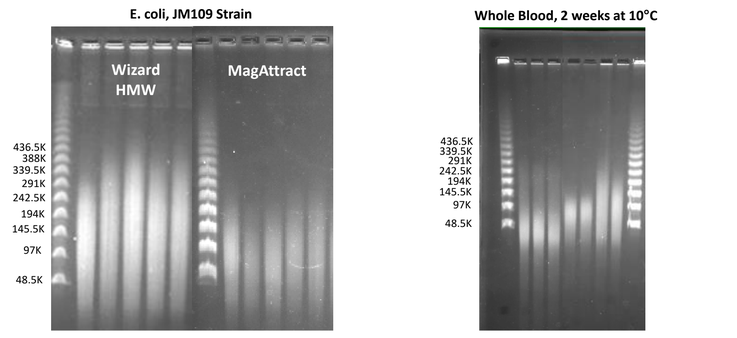
Sample	GQN 50	GQN 97	GQN 125
D	8.1	6.4	4.7
E	8.3	6.8	5.2
F	8.5	7.2	5.9

PFGE gel images are accompanied by example Femto Pulse electropherograms and the calculated GQNs (Genomic Quality Number) for each sample. The GQN is a proprietary metric used to assist in quantitating the distribution of fragment size. For instance, a GQN-**97** of 6.4 indicates that 64% of the DNA present in the sample is **97**kb or larger.

The Wizard® HMW DNA method was used for samples A-C, whereas a competing method was used for samples D-F. Gel images offer a useful general perspective of the fragment size distribution, but it remains difficult to assess and quantitate how much of the sample is of sufficient size to proceed to Long-Read Sequencing. GQN measurements at several molecular sizes show that while the two extraction methods may appear to produce similar proportions of HMW DNA (>50kb) when looking at a gel only, the HMW method results in a significantly higher proportion of DNA that is over 125kb.

Predicting success in downstream applications, such as Long-Read Sequencing, becomes a more reliable and quantitative process when using the capillary-based Femto Pulse system. While this method does not directly predict any sequencing performance metric specifically, it offers a tool to more precisely assess the quality of DNA prior to costly sequencing or library preparation.

6. Extraction Technology Comparison



Extraction Kit	Yield, ug	260/230	260/280
Wiz HMW	17.70	2.11	1.91
MagAttract	2.95	1.89	1.90

Lanes	1	2-4	5, 6	7,8	9
Sample	Marker	MagAttract	Control	HMW	Marker

Commercial HMW DNA extraction kits currently available are frequently used by researchers, though their relative ease and low cost compared to the gold-standard gel plug method may come with excessive trade-offs. In comparing the QIAGEN® MagAttract HMW DNA Kit with the Promega Wizard ® HMW DNA method, there are clear differences in not only yield and purity, but also in the recovered fragment size.

The QIAGEN kit employs magnetic beads for separation and may therefore be more prone to HMW DNA shearing. The Promega method, being based on precipitation and avoiding damaging handling steps, yields larger fragments across multiple sample types.

7. Long-Read Sequencing

Physical analysis methods such as PFGE, CE, and spectroscopy can serve as useful in-process checks to assess the quality of an extraction or a sample's potential suitability for molecular analysis. However, these methods may also fail to detect DNA damage, impurities, or inhibitors that could reduce performance in sequencing.

DNA extracted from samples of fresh, whole human blood and K562 cells using the Wizard HMW Kit was prepped for sequencing using the standard ligation-based protocol, and subsequently sequenced on Oxford Nanopore instruments.

	Whole Blood	K562 Cells
Mean Read Length	15,159	15,063
N50	54,882	47,278
Mean read quality	11.3	13.4
Longest read	434,461	422,150
Reads greater than 100kb	2.3%	1.3%
Flow cell yield (Gb called)	4.4 Gb*	8.11 Gb

*Blood sample was one of three samples run on this flow cell, whose total yield was ~15 Gb. The K562 sample was run on its own flow cell

In addition to obtaining good sequencing yields and quality scores, both sample types produced very long maximum reads and high N50 scores. This indicates that the DNA isolated with the Wizard® HMW kit was not only large in size, but also highly functional and pure.

8. Summary and looking forward

The advances in long-read sequencing, including longer read lengths, higher throughputs, and lower costs, are driving the field of genomics to embrace this technology area as a tool to solve new and ongoing genetic puzzles. This will have lasting and meaningful impact on a variety of fields including genetic diseases, oncology, microbiology, and agriculture.

The new Promega Wizard® HMW extraction kit chemistry and method enables researchers to obtain high-mass DNA from a variety of samples relevant to these areas that will give strong performance in long-read sequencing. In addition, as a cost-effective method, it will broaden the reach of long-read sequencing to a wider audience, including those with limited resources or less experience in extracting HMW DNA.

In subsequent studies, potential sequencing optimizations and sample types will be investigated. Of particular interest is the use of the rapid library prep method, which may better preserve large fragments that were successfully extracted than the traditional ligation-based method which employs bead-based clean-ups that are likely shearing DNA.

Corresponding author: andy.zegers@promega.com