Chapter 7

DNA Isolation Long-Read Genomic Sequencing in Ctenophores

David Moraga Amador, Andrea B. Kohn, Yelena Bobkova, Nedka G. Panayotova, and Leonid L. Moroz

Abstract

Long-read sequencing has proven the necessity for high-quality genomic assemblies of reference species, including enigmatic ctenophores. Obtaining high-molecular-weight genomic DNA is pivotal to this process and has proven highly problematic for many species. Here, we discuss different methodologies for gDNA isolation and present a protocol for isolating gDNA for several members of the phylum Ctenophora. Specifically, we describe a Pacific Biosciences library construction method used in conjunction with gDNA isolation methods that have proven successful in obtaining high-quality genomic assemblies in ctenophores.

Key words Genome, Ctenophora, *Pleurobrachia*, *Mnemiopsis*, *Beroe*, *Bolinopsis*, Evolution, Basal metazoans, PacBio sequencing, Long-read sequencing

1 Introduction

Long-read DNA sequencing technologies for long-range mapping [1–5] were introduced about a decade ago. Initially, these technologies were vastly more expensive than their short-read counterparts, and their use was mainly restricted to small genomes. However, recent advances in throughput and data quality have expanded their use, thus enabling the analysis of large, complex genomes at a resolution that had not been possible before by other sequencing methods. These long-range technologies are rapidly advancing the field with improved reference genomes, more comprehensive variant identification, and more complete views of transcriptomes and epigenomes [1, 3, 6]. Long-read sequencing improves mappability for resequencing and simplifies de novo assembly at the chromosome-scale level, including ctenophore

Authors David Moraga Amador and Andrea B. Kohn have equally contributed to this chapter.

Leonid L. Moroz (ed.), *Ctenophores: Methods and Protocols*, Methods in Molecular Biology, vol. 2757,
https://doi.org/10.1007/978-1-0716-3642-8_7,

species [7–9], which convincingly confirms the phylogenetic position of comb jellies as sister to the rest of animals [7].

The most developed of the long-read technologies, Pacific Biosciences (PacBio) sequencing, is also referred to as single-molecule, real-time (SMRT) sequencing [10]. It does not rely on an amplification step for either library construction or sequencing. The product of the library construction process is adapter-ligated DNA in its native form, leading to more uniform coverage across the target genome. Palindromes and low-diversity regions of the genome can be effectively resolved. However, these advantages come at the cost of stringent sample requirements. To take full advantage of the benefits of SMRT sequencing, samples must be free of impurities that may potentially inhibit library construction or sequencing reactions, retain high integrity throughout the isolation process (i.e., undamaged high-molecular-weight DNA), and be supplied in sufficient amounts (typically 10–20 micrograms, but now these requirements are reduced to about ng range) to ensure appropriate library construction yield. The most current PacBio platform (https://www.pacb.com/products-and-services/sequel-system/) in its current state of development (beginning of 2019, with version 3 chemistry and SMRT Link 6.1 software) produces some of the longest average polymerase read lengths available in the industry (average > 30,000 bp at that time and signigicantly longer now), with more than half the data in reads >45 kilobases (Kb).

This chapter discusses and describes DNA isolation and library construction protocols that are suitable for genomic analysis of ctenophores through PacBio long-read sequencing.

2 Materials

2.1 Overview of Genomic DNA Isolation Methods

The numerous genomic DNA (gDNA) isolation methods described [6, 11–14] reflect that no universal protocol will work for every possible sample and all probable experiments. One's choice depends largely on two main factors: the source (i.e., nature) of the sample and the requirements of the downstream application. Relatively straightforward procedures can attain efficient isolation of high-quality and pure gDNA from mammalian cell cultures. The same is true for plasmid DNA extraction from recombinant DNA constructs. However, other sample sources (e.g., plant, soil, and marine organisms) can be more challenging. Steps involving chemical, enzymatic, or physical treatments often require much optimization to obtain gDNA of the necessary quality, purity, and yield that subsequent procedures may use. Many marine organisms have mucous membranes and gelatinous bodies, making obtaining pure, high-quality gDNA difficult. Besides, many marine organisms are not easily separated from contaminants or do not have cell-dense tissues readily available.

There are several critical steps in the process of gDNA isolation: effective disruption of cells or tissue, separation of cellular debris from the gDNA, denaturation of nucleoprotein complexes, inactivation of nucleases including DNases, purification of gDNA, and quality control (purity, quantity, integrity assessment) of the resultant product.

Early procedures for DNA extraction were developed from CsCl gradient centrifugation in which DNA is separated and isolated on a density gradient [15, 16]. Today, most protocols have been developed into commercial kits that perform the DNA or RNA extraction processes faster, cheaper, and easier while generating adequate quality products.

Many specialized gDNA (and/or RNA) isolation methods have been developed with attention to the specific nature of the sample and the desire for automation. Generally, they are divided into solution-based, column-based, and magnetic-particle-based protocols. Due to the fragile nature of ctenophores, the common, solution-based acid guanidinium thiocyanate-phenol-chloroform extraction protocol such as the Invitrogen™ TRIzol™ (Cat # 15596026) reagent did not generate high-quality gDNA. The TRIzol™ reagent can simultaneously isolate RNA, DNA, and protein from diverse biological sources, but it requires a lot of starting material, uses harsh and toxic reagents, and is more time-consuming than other kits.

The column-based protocols of gDNA isolation may be the most popular and convenient today. The column-based methods are divided into two categories, anion-exchange chromatography and silica-membrane technology. One of the most popular kits for reasons of brevity, low price, and convenience is the DNeasy family of products from QIAGEN (Cat # 69504). The DNeasy membrane is a silica-based membrane column that can undergo centrifugation in a plastic microcentrifuge tube. DNA is selectively bound to the membrane while contaminants pass through. The DNA-containing column is then washed with an ethanolic solution (i.e., 70–80% ethanol, 20–30% water), and the DNA is eluted off the column with 100% water or low-salt (10 mM Tris, pH 8.0) solution. The major disadvantage of this method is the centrifugation step inevitably compromises the gDNA size that can be isolated. Another disadvantage is the sample size; the column is placed in a 1.5 mL microcentrifuge tube, and whole ctenophores sometimes are too large to fit into the tube. However, DNeasy is ideal for isolating <20 Kb gDNA.

More recently, a variety of magnetic-particle-based kits have become available. One example of this type of kits is the QIAGEN MagAttract® H.M.W. kit (Cat # 67563). This product is as convenient to use as the DNeasy kit, with the added advantage of being suitable for automation. Also, the QIAGEN MagAttract® protocol seems better suited for generating high-molecular-weight gDNA.

However, the yield was ten times lower than the QIAGEN Genomic-tip kits. Of note, we tested other kits such as the Jet-Flex™ genomic DNA purification kit (Cat # A30701, ThermoFisher Scientific) and OMNIPREP™ (Cat # 786-136, G-BIOSCIENCES), but none of these protocols produced high-quality gDNA from ctenophores.

3 Methods

3.1 High-Molecular-Weight gDNA Isolation Protocol for Ctenophores

Here we present a protocol for DNA isolation from ctenophores that resulted in high-molecular-weight gDNA. This document uses the terms extraction, isolation, and purification interchangeably. However, isolation encompasses the process of extraction and purification of gDNA.

In our experience, the most successful protocol to generate high-molecular-weight ctenophore gDNA was with the QIAGEN Genomic-tip kit (Cat # 10262) [17]. Gel electrophoresis and Agilent TapeStation analysis showed gDNA fragments in the >50 Kb range, an integrity level sufficient for demanding downstream applications, including constructing large-insert sized gDNA libraries for PacBio sequencing. The QIAGEN Genomic-tip uses a solid-phase anion-exchange resin, which yields high-quality high-molecular-weight gDNA through a gravity-fed column. The protocol was followed according to the manufacturer's recommendations with a few modifications (where noted).

1. Evaluating necessary steps to ensure the isolation of high-quality gDNA is critical; *see* **Notes 1**–**3**.
2. Before DNA extractions for genomic analysis, all animals were microscopically examined for potential ectoparasites such as copepods and then washed three times for at least 30 min in filtered seawater (FSW) that the animals are native to. Use fresh animals when possible to proceed further (*see* **Note 4**).
3. The QIAGEN Genomic-tip comes in various sizes, but the 500/G is preferred because of the large volume capacity. Animals were placed in a lysis buffer containing 200 μg/mL of RNAse A as well as both protease and proteinase K at concentrations of 1 mg/mL. Samples are incubated at 50 °C with gentle shaking as recommended by the manufacturer's protocol. However, the incubation time is shortened to the point when the animal has just dissolved. Longer incubations (e.g., 2 h as recommended by the manufacturer) resulted in degraded gDNA.
4. The lysate is loaded onto an equilibrated gravity-fed column. The column is washed with a medium-salt buffer to remove all contaminants like traces of RNA and protein.

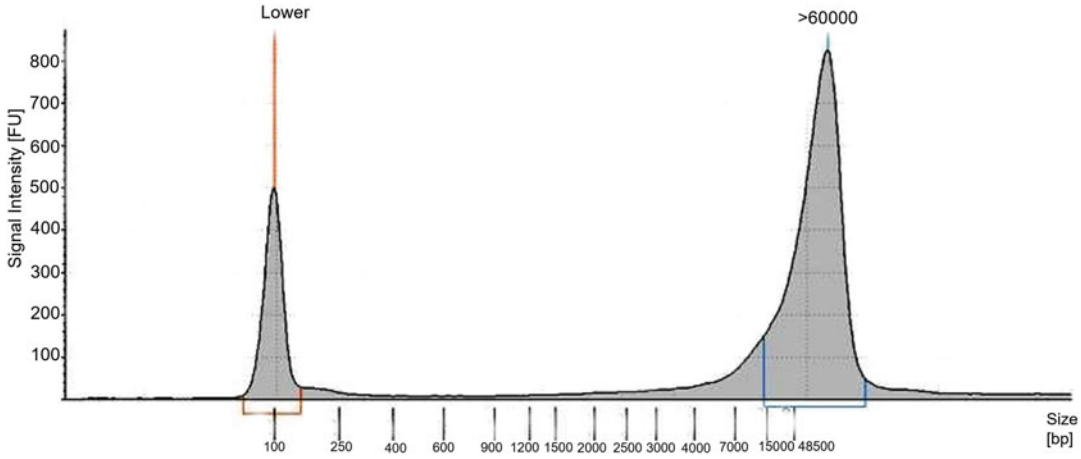

Fig. 1 *Beroe abyssicola* gDNA, as seen on the Agilent 2200 TapeStation (genomic tape) system. The peak at 100 bp, labeled "lower" is a size marker. The average gDNA size is estimated as >60,000 Kb

5. The gDNA is then eluted from the QIAGEN Genomic-tip with a high-salt buffer.
6. Finally, the eluted gDNA is desalted and concentrated by isopropanol precipitation.
7. The purified gDNA is then brought up in Dnase-/RNase-free water (Cat # AM9938, ThermoFisher Scientific). Interestingly, the use of T.E. or Tris buffers resulted in degradation of ctenophore gDNA.
8. Quality and quantity of genomic DNA are analyzed on a Qubit® 2.0 Fluorometer (ThermoFisher Scientific) and genomic DNA analysis ScreenTape (Cat # 5067-5365) on an Agilent 2200 TapeStation system (Cat # G2964AA). See section below on "Evaluation of Genomic DNA." Using this protocol, we routinely isolated DNA from ctenophores that showed a >60 Kb peak in the Agilent TapeStation; *see* Fig. 1.
9. Proceed to the construction of the sequencing library as soon as DNA isolation is completed and quality control is acceptable. For additional considerations, *see* **Notes 5–8**.

3.2 Evaluation of Genomic DNA

3.2.1 Quantitative Assessment

Absorbance-based methods (e.g., NanoDrop or equivalent) have traditionally been a favorite in most molecular biology labs. However, these methods are not adequate as they lack specificity and almost always overestimate the gDNA concentration. U.V. absorbance measurements are not selective and cannot distinguish DNA, RNA, or protein. Absorbance values are easily affected by other contaminants (e.g., free nucleotides, salts, and organic compounds) and variations in base composition; *see* Fig. 2. In addition, the sensitivity of spectrophotometric methods is often inadequate, prohibiting the quantitation of DNA and RNA at low

DNA samples [QC]

By Nanodrop		By QUBIT	
Sample	ng/μL	Sample	ng/μL
A	32.3	A	32.3
B	148.9	B	167.2
C	46.1	C	19.2
D	70.6	D	35.7
E	141.3	E	5.3

Fig. 2 Comparison of Qubit and NanoDrop methods for gDNA quantification. Samples A and B appeared to be very pure preps as the NanoDrop and Qubit values were in good agreement. These samples behaved well during library construction and sequencing. Samples D and E were very problematic and required additional purification. Initial attempts for library construction failed. Yields after extensive purification were low, but samples were eventually sequenced successfully

concentrations. The most sensitive and reliable gDNA concentration measurements are obtained using double-stranded gDNA (dsDNA)-specific reagents, such as those in the PicoGreen or Qubit assay [18–21]. Although the PicoGreen assay is more accurate, the much faster Qubit dsDNA assay is both cheaper and adequate for most purposes. It employs a fluorescent, DNA-binding dye that enables reliable, sensitive (down to 0.1 ng/μL) and specific quantitation of small amounts of dsDNA. The dye shows a minimal binding to single-stranded gDNA (ssDNA) and RNA For more recommendations when performing Qubit assays; *see* **Notes 9–12**.

3.2.2 Purity Assessment

A full assessment of the purity of a gDNA isolation prep is not a trivial matter. However, there are a few practical ways for doing this using the NanoDrop (or equivalent) by looking at three main parameters: OD ratios at 260/280 and 260/230 and the scanning pattern over the 220–350 nm range. These are a few things to consider:

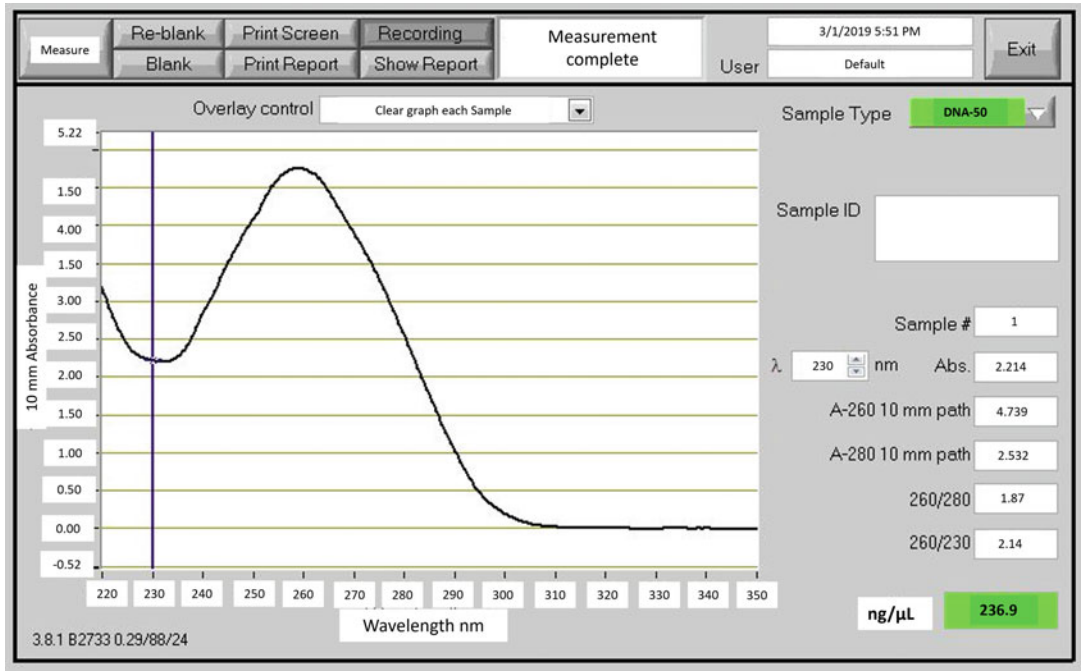

Fig. 3 Typical scan pattern and metrics for a pure gDNA preparation when seen on the NanoDrop

1. OD 260/A280 ratio of ~1.8 is generally accepted as "pure" for DNA. Some preps may have OD 260/A280 that may be as high as 2.2 (this should not be of concern).
2. OD 260/A230 ratio of 2.0–2.5 is generally accepted as "pure" for nucleic acid.
3. A low OD 260/A280 ratio may be the result of a contaminant such as protein or a reagent such as phenol. Although peptide bonds have an absorbance maximum at 280 nm, the presence of proteins in the sample can contribute significantly to the absorbance value at 260 nm.
4. A low A260/A230 ratio may be the result of a contaminant absorbing at 230 nm or less. Such contaminants include carbohydrates, residual phenol, residual guanidine, and/or glycogen. On the other hand, a high A260/A230 ratio may be the result of a dirty pedestal or using a blank solution that is not of similar ionic strength as the sample solution.
5. The 220–350 nm scan of a pure DNA prep will typically have a valley at 230 nm and a peak at 260 nm that extends down to baseline absorption at ~300 nm; *see* Fig. 3.

3.2.3 Qualitative (Size) Assessment

gDNA quality is of crucial importance for long-read sequencing. An isolation protocol should ideally avoid causing any DNA handling that can result in depurination, the formation of interstrand cross-links, nicks, etc. There are no quick and straightforward ways of thoroughly assessing all these parameters.

Nevertheless, evaluating the average size of the material is a good starting point that can be measured through routine gel electrophoresis and densitometry procedures, commonly available in molecular biology labs. Precise measurements of high-molecular-weight preps can be attained using the Bio-Rad CHEF Mapper X.A. Pulsed Field Electrophoresis system. Additionally, other commercially available systems are capable of resolving DNA fragments and smears up to ~50 Kb. These include Sage Science's Pippin Pulse Electrophoresis Power Supply and the Advanced Analytical Technologies, Inc. FEMTO Pulse (a fast and automated pulsed-field capillary electrophoresis instrument). However, we find that all these methods either require large DNA amounts (CHEF and Pippin Pulse) or are too tedious to use in routine workflows.

We have found that the Agilent TapeStation provides adequate sizing data for 20–30 Kb PacBio library construction. The TapeStation (genomic "tapes") requires very little material (few nanograms) and takes but a few minutes to run. The resolution above 20 Kb is unreliable because of compression in the high M.W. range. However, samples with peaks at >40–50 Kb can be processed without fragmentation and typically produce good libraries in the 20–30 Kb range. Very high-quality preps usually show a peak at >60 Kb (*see* Fig. 1).

3.3 Pacific Biosciences® Library Construction

3.3.1 PacBio Sample Requirements for Long-Insert Library Construction

Since long-insert (>20 Kb) library construction protocols for PacBio sequencing do not utilize any amplification, the quality of the input DNA will be directly reflected in the extent of sequencing success or failure. Any DNA damage (e.g., abasic sites, inter-strand cross-links, nicks, etc.) or contaminants in the DNA preparation (e.g., single-stranded DNA, RNA, proteins, polysaccharides, dyes, salts, etc.) will negatively affect the library construction process. Pure, high-quality gDNA is imperative for obtaining long read lengths and overall optimal sequencing performance. For additional information on the handling of gDNA samples for library construction, *see* **Notes 13–23**.

3.3.2 Pacific Biosciences® Library Construction Background

Once high-quality DNA has been isolated; long-insert sequencing libraries must be appropriately constructed with attention to the PacBio sequencing chemistry requirements. Here, we describe a protocol to generate 20–30 Kb SMRTbell sequencing libraries for four different ctenophores: *Pleurobrachia bachei*, *Beroe abyssicola*, *Bolinopsis microptera*, and *Mnemiopsis leidyi*.

The gDNA isolation protocol Subheadings 3.1 and 3.2 resulted in relatively low concentration solutions. However, because DNA was eluted in water, samples could be concentrated (if desired) by simple volume reduction on a SpeedVac with no significant change in the salt concentration. Typically, DNA preparations of adequate quality with a concentration >10 ng/μL were directly submitted to the MoBio PowerClean step without volume reduction.

3.3.3 PacBio Library Construction Protocol

Large-insert, SMRTbell (PacBio) libraries were constructed according to the recommended protocol by PacBio (*see* Fig. 4 for workflow), with a few important modifications as follows:

1. gDNA preparations from ctenophores were evaluated as described in Subheadings 3.2.1, 3.2.2 and 3.2.3. Samples with a concentration of at least 10 ng/μL, size >30 Kb, OD 260/280 = 1.8–2.0, and OD 260/230 > 2.0 were submitted to a MoBio PowerClean purification step (QIAGEN, Cat # 12877-50). The final elution step was substituted by 0.6X AMPure (Beckman Coulter, Cat # A63880); *see* **Note 23**.
2. The MoBio cleaned gDNA was quantified (Qubit) and sized on the TapeStation (genomic tape). This QC step was necessary because the MoBio procedure often resulted in significant fragmentation and loss of material. The extent of gDNA loss was sample dependent, from 30% to 95% in extreme cases (e.g., *Mnemiopsis* gDNA). The MoBio procedure was performed on as many DNA preps as necessary to obtain ~5 micrograms of cleaned material.
3. G-tubes (Covaris Inc. Cat # 520079) were used to fragment DNA preps that were still >30 Kb after the MoBio step. If necessary, this step required dilution of DNA in Tris-EDTA buffer (up to 160 uL). The samples were processed without G-tube fragmentation if the average DNA size was <30 Kb.
4. Whenever possible, five micrograms of sheared and concentrated DNA (140 ng/μL) were used for the subsequent SMRT bell library construction steps. However, in some cases, libraries were constructed with as little as 2 micrograms of input DNA. The protocol details for the various types of libraries are described in PacBio documents (P/N 100-286-000 Version 10 January 2018), except that size selection was performed on the SageELF, rather than the BluePippin™ (*see* **step 5**, in Fig. 4). The library construction steps included: ExoVII treatment, DNA damage repair, end repair, blunt-end ligation of SMRT bell adaptors, and ExoIII/ExoVII treatment.
5. As outlined in Fig. 4, the library construction procedure typically resulted in 1.2–1.5 micrograms of SMRT bell library (i.e., 25–30% yield, except when tested on *Mnemiopsis*). The final library was size-selected in the SageELF™ instrument (Cat # ELD 7510), using 0.75% agarose gel cassettes and the 1–18 Kb v2 cassette definition program. The desired SageELF fractions were cleaned using AMPure magnetic beads (0.6X AMPure sample ratio) and eluted in 15 uL of 10 nM Tris HCl, pH 8.0. Library fragment size was estimated by the Agilent TapeStation (genomic DNA tapes), and these data were used for calculating molar concentrations. Typically, fractions in

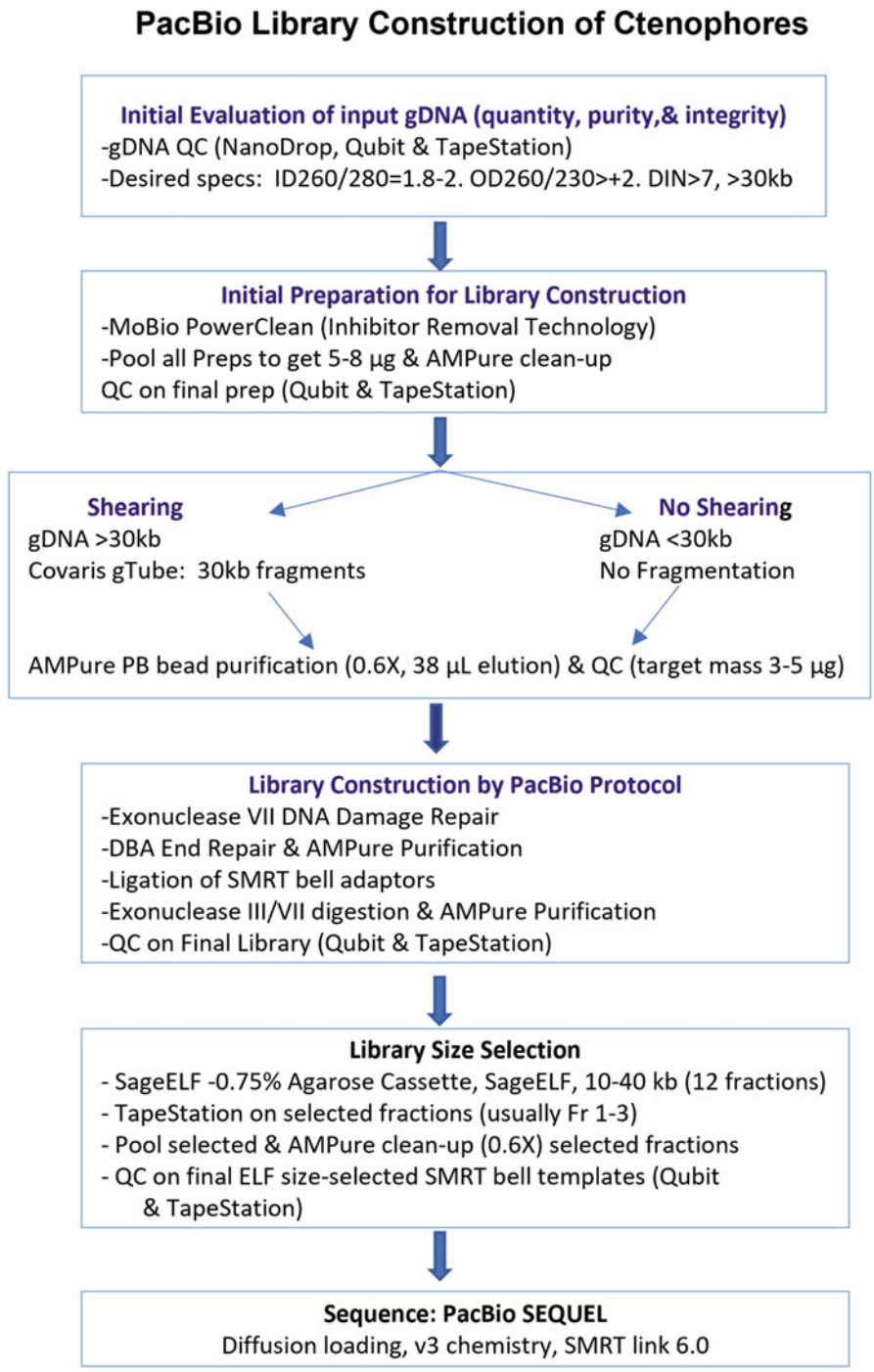

Fig. 4 Flow diagram of long-insert library construction steps for PacBio sequencing of ctenophores

wells 1, 2, and 3 contained library fragments in the 20–35 Kb range; *see* Fig. 5. Between 4 and 10 pM of the library was loaded onto the PacBio SEQUEL sample plate for sequencing.

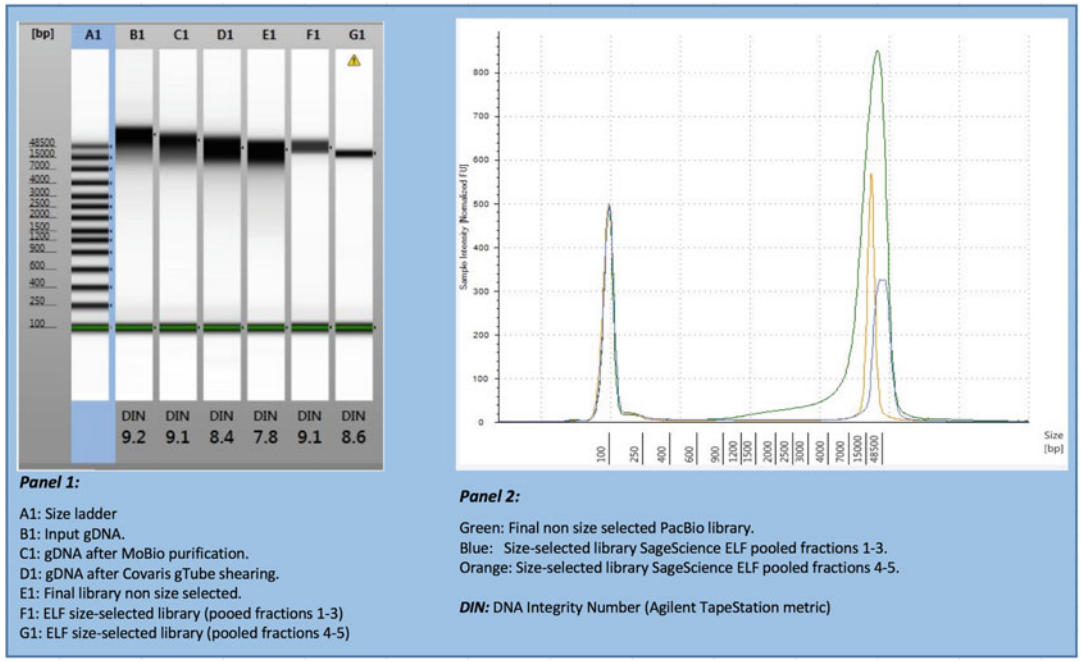

Fig. 5 Electrophoretic bands and peaks profiles for various PacBio library construction steps. Panel 1 shows the pseudogel graph of the indicated steps. Panel 2 displays the electrophoretic peak for the final non-size-selected library (largest, green peak) and two different size-selected library peaks that were created by pooling the indicated E.L.F. fractions

All other steps for sequencing were done according to the recommended protocol by the PacBio sequencing calculator.

3.4 Conclusions

Long-read sequencing promises to enable rapid advances in the study of ctenophore genomes. However, the fragile body plan of these organisms, combined with their unique mesoglea, offers many technical challenges when attempting to meet the stringent sample quality requirements for long-read sequencing technologies (e.g., PacBio). Among other difficulties to contend with, these organisms typically have a high mucilage content and lack cell-dense tissues. These factors make it hard to obtain high-quality DNA PacBio long-insert libraries. In particular, they require relatively large quantities of DNA, which must also be of high purity and integrity.

We have described a protocol that can be broadly used for high-quality gDNA isolation and PacBio library construction in ctenophores. This procedure was used with comparable success on several species (*Pleurobrachia bachei*, *Beroe abyssicola*, *Bolinopsis microptera*). The protocol was also tested on *Mnemiopsis leidyi*. However, this species yielded marginal library and sequencing results (i.e., lower library yields, shorter polymerase reads, and lower sequencing yields). MoBio cleaning of *Mnemiopsis* DNA

Table 1
Sequencing metrics for ctenophore: PacBio Sequel

Species	Approx. genome size	Gb per SMRT cell	Polymerase read length	Polymerase N50	Longest subread length	Longest subread N50	%P1 loading
Pleurobrachia bachei	0.2 Gb	3.4	6563	14,500	4361	7375	45
Beroe abyssicola	1 Gb	9.6	13,013	25,250	7981	12,536	72
Bolinopsis microptera	0.2 Gb	9.5	14,385	52,750	5841	9583	66
Mnemiopsis leidyi	0.16 Gb	1.6	3453	7648	3205	6839	46

resulted in a considerable loss (40–60%, depending on the DNA preparation). The DNA was resilient to binding to magnetic beads in the AMPure procedure even after cleaning. Most strikingly, the library size decreased to 3–5 Kb, and ~90% further material loss occurred after the Exonucleases III and VII step at the end of the library construction procedure. This behavior was consistent with damaged DNA upon isolation and/or during the library construction process. *Mnemiopsis* DNA library construction was attempted several times on freshly isolated DNA (never frozen) and on DNA preps that had been preserved at −80 °C. The results were the same for reasons that need further investigation. Interestingly, besides the library construction challenges with *Mnemiopsis*, the "surviving" library material produced average polymerase reads that were about one-fifth of the length of those generated by sequencing other ctenophores (e.g., *Beroe* and *Bolinopsis*).

Most of the gDNA isolation and library construction protocols steps were adapted from the manufacturer's manuals and procedures. However, several key steps in the workflow (Fig. 4) were modified, which resulted in relatively streamlined, robust, and efficient results. For DNA isolation, the QIAGEN Genomic-tip kit--based method was used with a few modifications to ensure sufficient purity and integrity of the final preparation. The most critical factors were gentle handling, optimized RNase/Proteinase K digestion conditions, proper washing of the lysate on the gravity-fed column, and prompt desalting-concentrating of the final DNA solution. The PacBio, long-insert library construction procedure was modified to include a stringent gDNA cleanup step (MoBio), and a library size selection by the SageELF rather than by the BluePippin™. Three out of four ctenophores species tested, yielded outstanding sequencing results that have enabled novel genomic analyses. Table 1 shows the sequencing run metrics

obtained in the PacBio SEQUEL for all four species tested, using v2.5 chemistry and SMRT link 5.1 software.

4 Notes

1. Only use wide-bore pipette tips when handling DNA and pipette very slowly to reduce shearing.
2. Minimize or eliminate any high-heating steps during isolation or preparation of DNA.
3. Minimize or eliminate high-speed vortexing; use gentle mixing techniques such as slow inversion.
4. Preferably, perform DNA extraction/isolation as soon as the tissue has been obtained. Otherwise, use tissue that has been flash-frozen with liquid nitrogen and stored at −80 °C. Alternatively, tissues may be adequately preserved when placed in a number of commercially available products [22]. However, these options should be carefully tested.
5. Include an extra cleanup step before library construction (e.g., MoBio PowerClean, same as QIAGEN DNeasy PowerClean Cleanup kit Cat # 12877-50).
6. Minimize or eliminate the number of freeze/thaw cycles with your sample to reduce DNA damage.
7. Allow sufficient thawing time for aliquots of DNA, as partially frozen DNA is prone to shearing.
8. Shipping DNA: Overnight shipping at 4 °C is preferred. However, if shipping overnight is not an option, flash-freeze the DNA sample with liquid nitrogen and ship frozen. Alternatively, some commercial products (e.g., DNAstable Plus) allow for the shipping of DNA at room temperature.
9. Assure that bubbles are not introduced into the sample at the reading as this can affect the results. Slight tapping on the tube wall or brief centrifugation will often help dissipate bubbles.
10. Samples should be diluted or concentrated as needed to remain within the quantitative range of the assay. If you get a concentration value as "too high" or "too low," it means that your sample is out of range.
11. The assay should be performed at room temperature, and the assay tubes must be at room temperature when the reading is taken. Do not hold assay tubes in your hand for too long while trying to read the samples.
12. The same Qubit protocol was used throughout the entire library construction process. The Qubit assay kits come in

quantification ranges: broad range and high sensitivity. The quantification results differ somewhat between the two kits.

13. The gDNA sample should be dissolved in 10 mM Tris, pH 7.5–8.0 at a minimum concentration of 30–50 ng/μL. However, for ctenophores, we used DNase/RNase-free water.
14. The gDNA sample needs to be double-stranded. Single-stranded DNA will not be ligated to the adaptor in the template preparation process and can interfere with quantitation and polymerase binding. For this reason, DNA must be quantified by fluorescence-based reagents such as PicoGreen or Qubit, which only detect double-stranded DNA.
15. A minimum of freeze-thaw cycles for your gDNA sample.
16. The gDNA samples do not need to be exposed to high temperatures (i.e., >65 °C for 1 h can cause a detectable decrease in sequence quality) or pH extremes (<6 or >9).
17. The gDNA sample should not be exposed to intercalating fluorescent dyes or ultraviolet radiation. If purified from a gel fragment, ethidium bromide, and UV must be avoided for staining and visualization. We recommend using SYBR safe with visualization on a blue lightbox (long wavelength).
18. An OD 260/280 ratio of approximately 1.8–2.0. OD 260/230 ratio higher than 2.0 is recommended for the gDNA sample.
19. The gDNA sample should be above 30 Kb in size if possible.
20. The gDNA sample should not contain insoluble material.
21. The gDNA sample should not contain RNA contamination.
22. The gDNA sample should not contain chelation agents (i.e., EDTA), divalent metal cations (i.e., Mg^{2+}), denaturants (guanidinium salts, phenol), or detergents (S.D.S., Triton-X100, CTAB).
23. The gDNA sample should not contain carryover contamination from the starting organism/tissue (heme, humic acid, polysaccharides, polyphenols, etc.) that may affect library construction and sequencing performance. For preparations containing a low level of contaminants, it may be sufficient to perform a 1X AMPure bead cleanup or a high-salt phenol-chloroform wash following gDNA extraction (http://www.pacb.com/wp-content/uploads/2015/09/Shared-Protocol-Guidelines-for-Using-a-Salt-Chloroform-Wash-to-Clean-Up-gDNA.pdf). Unfortunately, DNA preps from ctenophores contain significant levels of impurities. Preparations with OD 260/80 or 260/230 < 1.8, and with significantly different Qubit versus NanoDrop values, are good candidates for more stringent cleanup procedures with reagents such as MoBio

PowerClean (QIAGEN DNeasy PowerClean Cleanup kit Cat# 12877-50). In our experience, the use of the MoBio PowerClean DNA Clean Up Kit before library construction resulted in significantly improved sequencing results in the PacBio. In some cases, samples that had failed to sequence were "rescued" by a MoBio cleanup step. Other cleanup methods, namely, Zymo Research (Cat # D41010) and Clontech (Cat #740230.10) were tested. However, these methods resulted in significantly greater DNA fragmentation and loss than the MoBio kit.

Acknowledgments

This work was supported in part by the Human Frontiers Science Program (RGP0060/2017) and the National Science Foundation (IOS-1557923) grants to L.L.M. Research reported in this publication was also supported in part by the National Institute of Neurological Disorders and Stroke of the National Institutes of Health under award number R01NS114491 (to L.L.M). The content is solely the authors' responsibility and does not necessarily represent the official views of the National Institutes of Health.

References

1. Beaulaurier J, Schadt EE, Fang G (2019) Deciphering bacterial epigenomes using modern sequencing technologies. Nat Rev Genet 20(3):157–172
2. Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17(6):333–351
3. Jing Y et al (2019) Hybrid sequencing-based personal full-length transcriptomic analysis implicates proteostatic stress in metastatic ovarian cancer. Oncogene 38(16):3047–3060
4. Taylor TL et al (2019) Rapid, multiplexed, whole genome and plasmid sequencing of foodborne pathogens using long-read nanopore technology. bioRxiv:558718
5. Wenger AM et al (2019) Highly-accurate long-read sequencing improves variant detection and assembly of a human genome. bioRxiv:519025
6. Wagner Mackenzie B, Waite DW, Taylor MW (2015) Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences. Front Microbiol 6:130
7. Schultz DT et al (2023) Ancient gene linkages support ctenophores as sister to other animals. Nature 618(7963):110–117
8. Schultz DT et al (2021) A chromosome-scale genome assembly and karyotype of the ctenophore *Hormiphora californensis*. G3 (Bethesda) 11(11):jkab302
9. Hoencamp C et al (2021) 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. Science 372(6545):984–989
10. Vermeesch JR et al (2018) Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. Nucleic Acids Res 46(5):2159–2168
11. de Kok JB et al (1998) Use of real-time quantitative PCR to compare DNA isolation methods. Clin Chem 44(10):2201–2204
12. Nacheva E et al (2017) DNA isolation protocol effects on nuclear DNA analysis by microarrays, droplet digital PCR, and whole genome sequencing, and on mitochondrial DNA copy number estimation. PLoS One 12(7): e0180467
13. Psifidi A et al (2015) Comparison of eleven methods for genomic DNA extraction suitable for large-scale whole-genome genotyping and

long-term DNA banking using blood samples. PLoS One 10(1):e0115960

14. Varma A, Padh H, Shrivastava N (2007) Plant genomic DNA isolation: an art or a science. Biotechnol J 2(3):386–392
15. Garger SJ, Griffith OM, Grill LK (1983) Rapid purification of plasmid DNA by a single centrifugation in a two-step cesium chloride-ethidium bromide gradient. Biochem Biophys Res Commun 117(3):835–842
16. Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning, 1st edn. Cold Spring Harbor Laboratory
17. Moroz LL et al (2014) The ctenophore genome and the evolutionary origins of neural systems. Nature 510(7503):109–114
18. Ahn SJ, Costa J, Emanuel JR (1996) PicoGreen quantitation of DNA: effective evaluation of samples pre- or post-PCR. Nucleic Acids Res 24(13):2623–2625
19. Mardis E, McCombie WR (2017) Library quantification: fluorometric quantitation of double-stranded or single-stranded DNA samples using the qubit system. Cold Spring Harb Protoc 2017(6):pdb.prot094730
20. Nakayama Y et al (2016) Pitfalls of DNA quantification using DNA-binding fluorescent dyes and suggested solutions. PLoS One 11(3): e0150528
21. Rengarajan K et al (2002) Quantifying DNA concentrations using fluorometry: a comparison of fluorophores. Mol Vis 8:416–421
22. Gray MA, Pratte ZA, Kellogg CA (2013) Comparison of DNA preservation methods for environmental bacterial community samples. FEMS Microbiol Ecol 83(2):468–477