Ultra low input library preparation of single soil invertebrate specimens for Hifi PacBio sequencing

Leonie Schardt[1], Miklós Bálint[1]

[1]Senckenberg Gesellschaft für Naturforschung - LOEWE TBG

Senckenberg - FUG

Leonie Schardt
Senckenberg Gesellschaft für Naturforschung, LOEWE TBG

JAN 02, 2024

OPEN ACCESS

DOI:
dx.doi.org/10.17504/protocols.io.n92ldmqr7l5b/v1

Protocol Citation: Leonie Schardt, Miklós Bálint 2024. Ultra low input library preparation of single soil invertebrate specimens for Hifi PacBio sequencing. **protocols.io** https://dx.doi.org/10.17504/protocols.io.n92ldmqr7l5b/v1

Protocol status: In development
We are still developing and optimizing this protocol

DISCLAIMER

This protocol is not peer reviewed and so far not part of a peer reviewed publication - use at your own discretion.
Some safety precautions apply - includes steps with flammable chemicals (ethanol absolute).

ABSTRACT

This protocol is a modification of the Pacific Biosciences Ultra Low Input workflow for HiFi SMRTbell® Libraries. The modified workflow has been successfully used for the processing of samples well below the recommended quantity (down to 0.3 ng total DNA input) and quality thresholds (fragment distributions with means of around 10 kb length), producing libraries that are useable for Hifi sequencing on PacBio Sequel II platforms. The optimization of this process was done using single specimen soil invertebrates (Oribatida, Collembola, Protura) from newly ethanol stored as well older ethanol stored collections.

MATERIALS

Additional information

Version 11/2023

* Items in parentheses are recommended but can be interchanged

Created: Nov 08, 2023

Last Modified: Jan 02, 2024

PROTOCOL integer ID: 90635

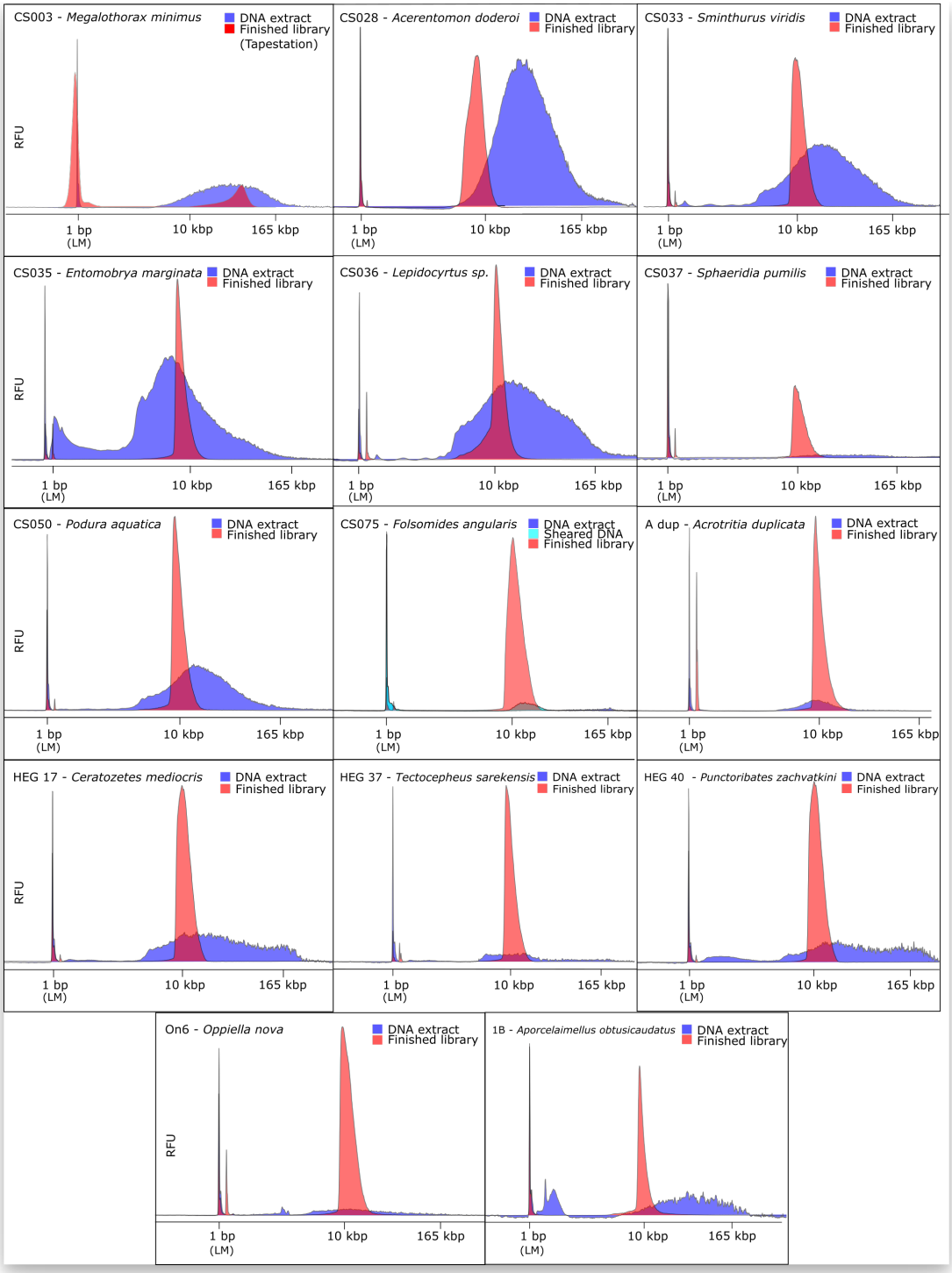
Keywords: PacBio, ultra low input, long read sequencing, soil invertebrate, single specimen, Pacific Biosciences, Hifi, oribatids, oribatida, collembola, collembolans, protura

Lab equipment

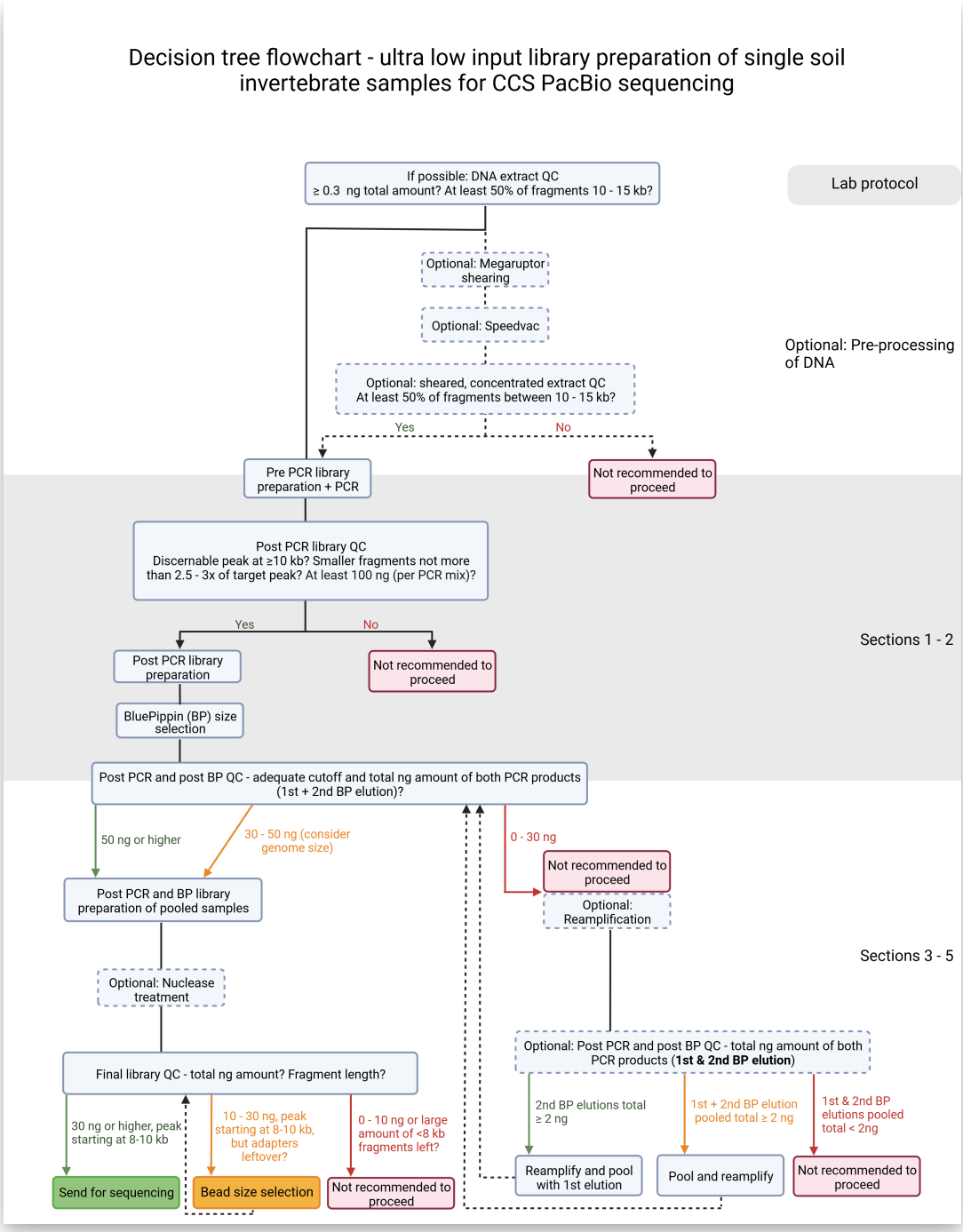
- Procedure & Checklist for Ultra Low Input libraries - PacBio, Version 02 (Nov. 2021)
- Procedure & Checklist for Low Input libraries - PacBio, Version 07 (Nov. 2021)
- (Megaruptor 2.0 - Diagenode)
- (Concentrator Plus - Eppendorf)
- (Quantus - Promega)
- Femto Pulse - Agilent
- BluePippin - Sage Science
- Magnetic rack for 0.2 ml PCR strips - for example Permagen
- Thermocycler with 0.2 ml tube block - for example Eppendorf
- Ice or cooling racks for Eppis and 0.2 ml tubes - for example Sarstedt

Consumable list

- SMRTbell® Express Template Prep Kit 2.0 - PacBio
- SMRTbell® gDNA Sample Amplification Kit - PacBio
- (Quantus dsDNA or dsDNA ONE reagents - Promega)
- (Ampure PB beads - PacBio)
- (Elution Buffer - PacBio)
- (Megaruptor Hydro Tubes and Hydropores long - Diagenode)
- (1.5 ml LoBind Microcentrifuge tubes for the final library - for example Eppendorf)
- Femto Pulse Genomic DNA 165 kb Kit - Agilent
- $MgCl_2$, 25mM stock solution - for example Genaxxon
- Ethanol absolute, 80% concentration needed for purification steps - for example Carl Roth
- PCR strips, 0.2 ml volume - for example Sarstedt & Sarstedt
- BluePippin reagents and Agarose Cassettes, BLF7510, 0.75% DF 3 - 10 kb (High Pass 4-20 kb), dye-free with external Marker S1 - SageScience, see also Biozym



Comparisons between extract inputs (blue) and final PacBio libraries (red) created with this protocol. Shown are overlaid Femto traces aligned by the lower marker (except sample CS003, showing a Tapestation electropherogram for the final library). Sample CS075 additonally shows the trace of the sheared fragments, since the original extract is barely visible.



Flowchart to help with sample processing decisions. Created with biorender.com.

Pre library steps: DNA quality control and lab setup

Extract quality check

The first parameter to be considered is the fragment length distribution. The availability of a Femto Pulse is crucial for accurate fragment length analysis,

especially when working with tiny single specimens, where the amount available for testing and quality checks (QC) is limited. This extends to later steps in the protocol. If there is no such device available, the protocol might still work, but it complicates the process considerably.

The samples shown above were all extracted using the MagAttract HMW DNA Kit from Qiagen, however, there are other suitable extractions like the salting out method that resulted in useable samples (not shown here). So far, only silica column based extractions failed to work with this library prep, due to them often resulting in a strong surplus of fragments below 10 kb.

This protocol was established for a diverse set of organisms, with varying fragment distributions found in the DNA extracts (see electropherograms above). The 50% threshold for large fragments given in our flowchart is a reference point, but not an ultimatum. It is also possible to use unsheared DNA that shows mid to low quality fragment distributions (see sample *Acrotitia duplicata* above). As long as there is some part/ fraction of the fragments in the target region of ≥10 kb, it is worth attempting the protocol. It is moreso the ratio of small to large fragments that is important. As long as fragments below 10 kb are not overrepresented, the size selection steps will usually result in sufficient fragment lengths in the final library without losing too much DNA in the process.

As with other NGS library preps, DNA concentration measurements via fluorescence are used as reference points here. Conversely, this protocol was established without purity ratio checks via absorbance measurement. DNA yields from single soil invertebrates are usually too low to consider additional purifications before library input, but there are fixed purification steps during the library prep that might mitigate potential issues.

The 5 ng threshold suggested by PacBio for ultra low input libraries is a good point of reference but not an ultimate limitation; single soil invertebrate samples often don't reach this yield. Due to the PCR step included in this protocol, the initial concentration of an extract does not matter much, but it complicates the initial quality control. So far, the lowest DNA amount successfully used as input for this protocol was 0.3 ng, but it is entirely possible lower amounts work as well.

Optional - shearing/fragmentation and volume reduction

Shearing of the samples here was done exclusively with the Megaruptor 2. We found that in some cases, the regular full washes between shearing runs was not sufficient to remove leftover DNA traces, so cross contamination is a possibility at this point. If available, users should consider using single-use shearing

methods, like g-Tubes from Covaris or the Megaruptor 3 device from Diagenode, which is using disposable syringe cassettes for sample shearing.

Adjust the elution volumes in the DNA extraction to 55 µl when shearing with a Megaruptor 2.0 device. This will result in a good working volume for the machine without diluting the sample too much. If the DNA concentration of a sample is still in an acceptable range after shearing, part of the sample should be used as is for the input. Evaporation or bead purification steps can skew the fragment distribution (evaporation) or lose DNA (beads) and should only be used if necessary. This protocol was established with the use of a Eppendorf Concentrator Plus, set to 45°C. Samples were run for 5 - 10 minute steps at a time, with volume checks in between. Run times should be approached in small steps of 5 - 10 minutes in the beginning and 2 - 5 minutes in the end, as to not expose the samples to this process more than necessary.

Laboratory pracitce

Due to the unspecific whole genome amplification step in the middle of the protocol, any contaminants introduced before that point will carry over into subsequent steps. Consider establishing a dedicated area with separate pipettes for these types of library preps, away from other sample types. In addition, any plasticware coming into contact with the samples should be irradiated with UV light before use. Regardless of the space available, the most important preventative measure is keeping samples covered as much as possible; samples should be covered with lids at all times when they are not being actively handled. PacBio kit reagents as well as samples should be kept on cooling blocks during the library preps. Make sure the reagents are properly thawed before use. Mix by flicking reagent tubes and short spin them down.

Pre PCR library preparation

QC is not strictly necessary after the PCR adapter ligation (pages 11 - 12 of Procedure & Checklist for Ultra Low Input libraries). A fragment analysis canbe done optionally to make sure leftover adapters and ultra small fragments were selected out. The concentration of the sample at this point is not indicative of PCR success.

Post PCR Quality control

QC is crucial at this point; the success of the PCR cascades into all subsequent protocol steps. There is no set concentration threshold, however, considering the

steps following after, we currently recommend to continue with samples only if they yield around 100 ng per PCR mix (total per sample close to 200 ng). Consider that the concentration after the PCR can only be evaluated hand in hand with the fragment distribution. A high concentration can also be the result of primarily short fragment amplicons that will be size selected out.

A Femto Pulse run will be the best measure of amplicon length and will indicate how to adjust for the following steps. The fragment distribution does not need to be ideal as shown in PacBio's protocol (page 15 of Procedure & Checklist for Ultra Low Input libraries), as there is a size selection step included here right afterwards. A second, high concentration peak at 1300 bp is common at this point, but should be of no concern if there are also enough target fragments present.

BluePippin automated gel size selection

General considerations

Cassettes and reagents need to warm up to RT before use. Cassettes that are too cold can have issues with the applied voltage which influences the way the samples run. After the run finishes, leaving the cassette in the device overnight is fine; the PCR products are usually stable at RT for some time. Some versions of the BP manual advise against this, but there was no quality loss observed in our trials.

The recovery success of the BP is not linear. Even if two samples have the same fragment distribution and are run with the same protocols and thresholds, the samples with lower amounts of input DNA will also have a lower recovery percentage. This might be due to DNA traveling through a gel matrix better in higher concentrations (https://pubmed.ncbi.nlm.nih.gov/6303150/). It is advisable to run PCR products on the BP first before pooling them, to be able to adjust to fragment length differences between PCR mixes 5A and 5B and reduce bias, but it is not strictly necessary. Pooling the two PCR mixes for one BP lane can improve the recovery percentage and save space on the cassette, but should only be done if the Femto electropherograms are nearly identical.

Choosing size selection cutoff values

The lower cutoff value used for the BluePippin needs to be adjusted based on the QC of the PCR products. The BP device used for establishing this protocol showed a slight delay in the set values versus the actual cutoffs (any automated

size selection device should be tested for this at the start). For samples that ran well in the PCR this delay was about 1.5 kb; i.e. 7 kb as the lower threshold setting resulted in a ~8.5 kb cutoff. (Note that this delay can be sample specific, as non soil invertebrate samples have shown different delays). The upper level used here was always 17 kb, an arbitrary value, as the improved recovery BP protocol collects all fragments after the lower cutoff value.
Factor in both peak shape and concentrations: if the fragment distribution is close to the target size and the concentration of the products is of no concern, you should push the limits and go for a 7 - 7.5 kb lower limit (see page 15 of Procedure & Checklist for Ultra Low Input libraries for ideal PCR product fragment lengths).
For PCR product peaks that are entirely shifted to the left, either 1.) adjust the lower cutoff value to 6.5 kb or lower to reduce the amount of DNA lost, or 2.) have a high cutoff, thus risking losing a lot of DNA and having to reamplify the samples afterwards. If the common secondary peak at 1300 bp is overrepresented and way more concentrated than the target peak, it can happen that not all of the small fragments can be selected out. BP runs can be repeated if necessary, but will always result in DNA loss. Elution from the BP cassette can be done twice. Yields are very different depending on the organism. For some samples it will not be needed, but it is advisable to keep the second elution as a backup anyway.

Sample loss after the BluePippin step

You should expect to lose an additional 30% - 40% of the library DNA amount in the steps **after** the BluePippin run (this does not include the optional nuclease digest). This means, right after the size selection, any amount above 100 ng in total is of no concern. Yields between 50 - 100 total ng are still manageable for most samples and genome sizes. Post BP libraries between 30 - 50 total ng can be difficult to manage, but this is highly dependent on the genome size. For soil organisms such as Collembola and Oribatida, that generally have genomes of a few hundred megabases, this is more than sufficient. Post BP concentrations between 0 - 30 total ngs are candidates for reamplification, but it might be that sequencing still works with final library amounts of around 10 ng. This needs to be evaluated together with the sequencing provider.

Reamplifications of post size selection PCR products

For very small animals that yield DNA at the lower end of the threshold, it might be necessary to perform reamplifications of the long range PCR products. It is not advisable to simply extend the existing PCR protocol, as it is optimized to run at 13 cycles (+- 5 cycles reamplification, as per pages 22 - 23 of "Procedure &

Checklist for Ultra Low Input libraries"). Raising the cycle numbers of the initial PCRs to 20 or higher will lead to exhaustion of the reagents and patterns of interrupted elongations (staccato fragment peaks where extension of the product was not finished).

BP size selected PCR products perform exceptionally well in long range reamplifications. Products with a concentration of below 5 ng can be reamplified to desired levels well below 14 cycles (this needs to be tested and adjusted for new samples). It is advisable to stick with the cycle limit recommended in the original PacBio protocol, e.g. if the first PCR ran for 13 cycles, try to reamplify the products with a maximum of 5 cycles.

It is possible to control for overamplification biases to some extent. You can 1.) retain part of the original size selected PCR product and pool it with the reamplified product later, normalized to equal levels or 2.) use the second BP elution for the reamplification and pool it with the first BP elution in equal parts. Reamplified products can be run on the BP again.

Post PCR and post size selection library prep - final steps

For streamlined sample types that worked well before, a quick gel picture or Bioanalyzer run is sufficient after the BluePippin size selection, a Femto Pulse run is not strictly necessary at this point. However, for the final libraries, the average fragment length measured on the Femto Pulse will be very close to the mean subread length of the sequenced product, making it easier to choose between replicate libraries if needed. A small amount of adapter dimers can be disregarded (110 - 125 bp fragments, see also electropherograms above), they will not factor in in the sequencing if the final library peak is overly present in comparison. If the concentration of the final library allows it, another size selecting bead purification (40%, 2.2x) can be added at this point to reduce the amount of leftover adapters.

If necessary, another BP run can be added at this point as well. However, we observed that finished libraries perform considerably worse than the intermittent PCR products on the BluePippin in terms of recovery percentage, which might be due to the circular structure of the library inserts with sequencing adapters attached.

Final library amounts ≥ 50 ng suffice for most samples, though organisms with very large genomes might need libraries close to 100 ng to reduce the amount of SMRTcells needed. Final library amounts ranging from 20 - 50 ng sufficed for all soil organisms tested for this protocol (Oribatida, Collembola, Protura, all with ~100s of MBs genome sizes). Final library amounts between 10 ng and 30 ng

should be discussed with the sequencing provider first to ensure they can be processed for loading onto the SMRTcells. Libraries prepared with this protocol that resulted in less than 10 ng were not tested for sequencing.

This protocol was established without the nuclease treatment detailed on pages 13 and 14 Procedure & Checklist for Low Input libraries. For samples that have shown general issues with adapter ligations, or if leftover adaptors were an issue with other PacBio libraries before, this step should be considered after the sequencing adaptor ligation. Keep in mind that the digest and subsequent purification of the sample will result in additional DNA loss.

BEFORE START INSTRUCTIONS

Users should become familiar with the Ultra Low Input protocol from PacBio before considering this modifcation. The time needed to finish this protocol is approximately three work days, with two major breaks (PCR step and BluePippin size selection step). Experienced users might be able to condense the time needed to two work days.

Following is the wet lab protocol for immediate sample processing. It is deliberately kept short for an easier overview in the lab, for further information like equipment lists and recommendations, as well as a decision tree flowchart, see the **Materials** tab.

Optional: Pre-processing of DNA

1 For high molecular weight (HMW) DNA extracts, fragmentation of the input material might be required. Additionally, if the whole extract is to be used for library preparation, concentrating the

volume down might be necessary.

1.1 Optional: shearing of gDNA with Megaruptor 2

- Run a full wash cycle on the device before starting (~5 minutes).
- Pipette the sample into a diagenode Hydrotube (up to two tubes/ samples per run).
- Dilute to 55 µl with 1 x TE buffer or water.
- Spin down the samples on table centrifuge, if impurities are to be expected, centrifuge longer to collect them at the bottom of the Hydrotube.
- Set machine to either 10 kb, 15 kb, or 20 kb (for soil invertebrates - 15 kb, but it depends on the sample type) for the respective samples, adjust input volumes in the menu if necessary.
- Attach Hydropores (long type) to the machine as shown on the display and initiate pre-load (adds 45 µl TE buffer).
  Do not attach Hydrotubes to the Hydropores without the pre-load, this will pull your samples into the pore via capillary forces!
- Stick Hydrotubes to the pores, making sure the end of the pore is centered in the tube as much as possible.
- Start the run, one sample takes approx. 15 minutes, two samples take approx. 30 minutes.
- Unscrew Hydropores with tubes still attached.
- Centrifuge the tubes with the Hydropore still attached if possible to get as much of the extractback out as possible. Spin carefully as to not detach the Hydropore mid spin.

1.2 Optional: concentrating down DNA or sheared gDNA

Let the samples run on a Concentrator Plus system at 45°C (system has fixed rpm) with an open lid to concentrate them down. Try to approach the optimal input volume of 45.4 µl in small steps.

1. Pre-PCR library preparation

2 Follow the Procedure & Checklist for **Ultra Low** Input libraries pages 9 and 10 without changes.

3 At the adapter ligation step (page 11), reduce the amount of ligation mix to 20 µl per sample. Add $MgCl_2$ to a final concentration of 3 mM to each sample (for example 12 µl of a 25 mM $MgCl_2$ stock solution, this results in a total volume of 96.5 µl per sample). Add the Diluted Amplification Adapters last.

4 Incubate at 20°C for 1:40 h, afterwards the samples can be kept at 4°C overnight if necessary.

5 Instead of the purification on page 12, perform the size selection purification detailed on page 15 and 16 of Procedure & Checklist for **Low** Input libraries, to get rid of as many remaining small fragments and leftover adapters before the PCR step. Example calculation here:

For every 96.5 µl sample, prepare 96.5 µl x 2.2 bead mix, rounded up to 215 µl (40 % aka. 86 µl beads mixed with 60 % aka. 129 µl PacBio elution buffer). It is important to carefully remove the supernatant after the bead pellet has formed, to avoid carryover of small fragments. After the ethanol cleaning steps, elute with 97 µl PacBio elution buffer and incubate for 30 minutes at RT. Since the PCR afterwards is performed in two separate reactions, split the sample after elution and transfer 48 µl into separate PCR strips. The adapter ligated samples can be stored at 4°C for a few days if needed.

6 Continue on page 13 of Procedure & Checklist for **Ultra Low** Input libraries. Add PCR mixes and primers to the samples one by one (no pooled master mix). Increase cycle numbers to 14 for both PCR protocols.

2. Post-PCR library preparation

7

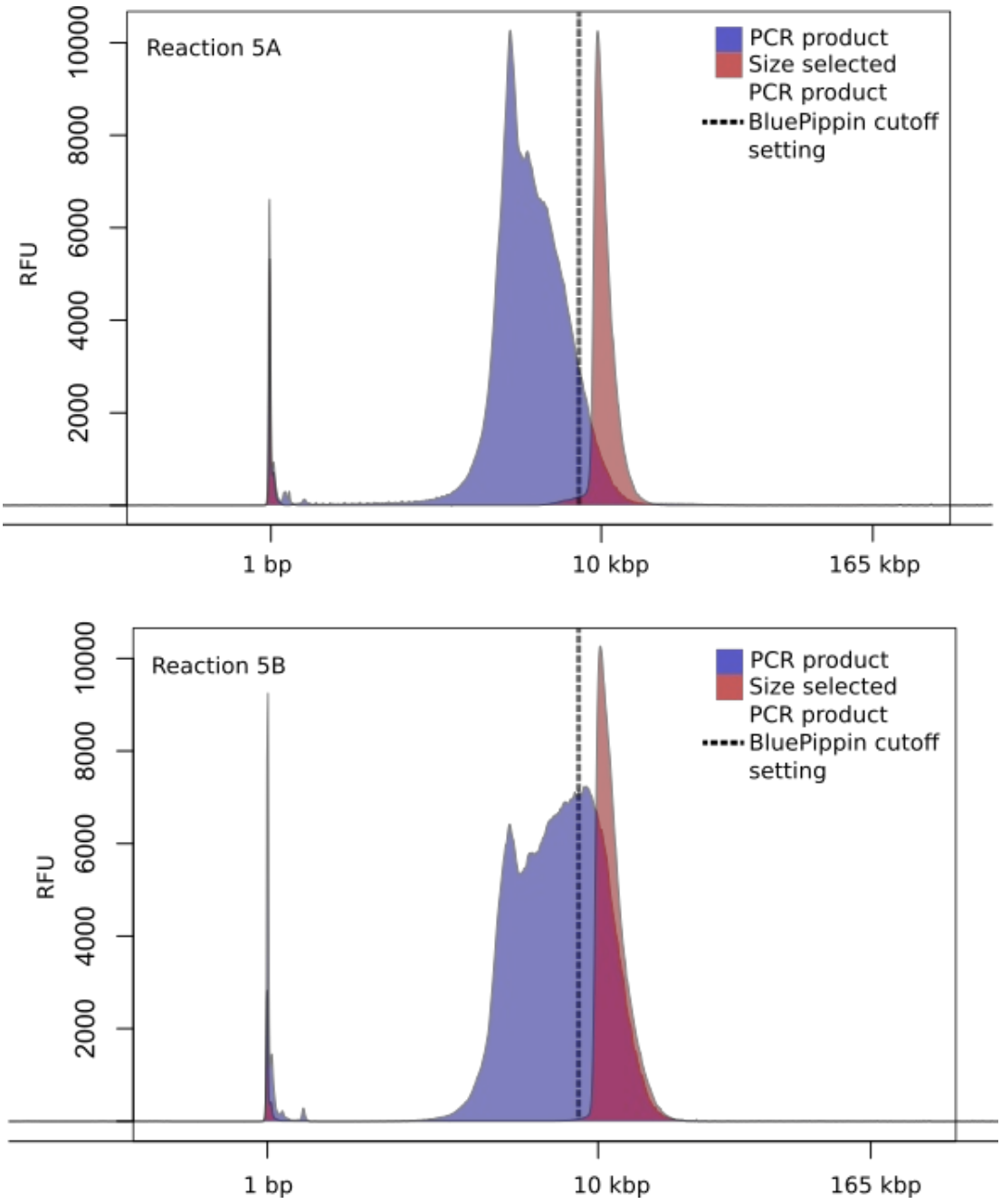
Note

This point marks the biggest shift from the traditional PacBio Ultra Low protocol. The BluePippin size selection is usually done at the very end of the libary preparation, we advise to do it with the unprocessed PCR products instead.

Perform the purification as shown in the Procedure & Checklist for **Ultra Low** Input libraries on page 14. Extend the bead binding and elution incubation times to at least 10 minutes and elute in ~32µl of elution buffer. 2 µl is kept for the QC, 30 µl is the volume needed for the subsequent size selection on the BluePippin. Purified PCR products are stable at 4°C for a few weeks or -20°C for a few months.

8

Measure the concentrations of the purified PCR products on a Quantus Fluorometer or comparable system and assess fragment lengths on a Femto Pulse system. For post PCR concentration and fragment length guidelines see Materials - section Post PCR Quality control.



Example for PCR product QC and the chosen cutoff values for the BluePippin size selection. Note that in our case, the real cutoff is slightly shifted from the BP setting, this needs to be adjusted based on the particular machine. The choice of where to put the lower threshhold is a tradeoff; in this case reaction 5A will retain less of the PCR product after size selection, due to the worse performance in the whole genome amplification (this is common). However, this means that the products can be mixed accordingly afterwards, securing a better representation of the 5A PCR product in the final library. In many cases it is wo

3. BluePippin (BP) size selection of PCR products

9 Let cassettes and reagents warm to RT, vortex and short spin down the loading solution, flick and short spin down the S1 marker.

10 Follow the general guideline for preparing samples and cassettes as described in the BluePippin User Guideline, add 42 µl of run buffer to the elution wells. Samples can be vortexed shortly to mix them properly with the loading solution.

11 Use the cassette definition “0.75% DF 3 – 10kb Marker S1 – Improved Recovery”.

12 Set the cutoff to any range from 3 - 8 kbp (the higher threshhold is arbitrary and can be set to 17 kbp for example), this depends on the fragment lengths after the PCR (see Materials - section BluePippin automated gel size selection). The separation and elution will run for 2.5 h to 3 h.

Example for BluePippin settings used in this protocol. Lane 1 - External reference marker S1, lanes 2 through 5 - PCR products 5A and 5B of two different samples.

13 After the program is finished, the cassette should be left in the device for at least 45 minutes to enhance elution efficiency; it can also be left in the device overnight. Gently pipette up and down a few times before removing the sample from the elution well. If a second BP elution should be done, immediately add another 40 µl of electrophoresis buffer to the elution wells and keep the cassette sealed at RT until needed.

14 Purify the samples by adding 40 µl magnetic beads (1 x purification), mix by slowly pipetting until the solution has a homogeneous color. Incubate for 10 - 15 minutes at RT. Pellet the beads on a magnetic rack until the solution is clear, discard the supernatant. Keep the tubes on the magnetic rack and wash twice with 200 µl of 80% ethanol. Quick spin the remaining ethanol to the bottom of the tube and return tubes to the magnetic rack. Pipette off any remaining ethanol from the bottom of the tubes. Immediately add 27 µl of elution buffer and incubate at RT for 10 - 15 minutes. Save 2 µl of the elution for the QC. Size selected and purified PCR products are stable at 4°C for a few weeks or -20°C for a few months.

15 Second BP elutions can be purified with the same protocol, but should be concentrated down as much as possible. When pooling first and second BP elutions, the total volume should stay below 74 µl, otherwise later volumes will not fit the 0.2 ml PCR tube.

If samples were checked with the Femto before the BP, it usually suffices to measure the concentration and proceed, alternatively do a quick Bioanalyzer run or gel. If sample concentrations are in an acceptable range (see Materials - section BluePippin automated gel size selection), pool both 5A and 5B PCR products (including second BP elutions if needed) and proceed with section "4. Post PCR and post BluePippin library preparation" below. When adding second BP elutions, try to stay below 74 µl, otherwise the later volumes will not fit the 0.2 ml PCR tube.

If sample amounts are lower than the recommended levels, proceed with "*Optional: reamplification of size selected PCR products"* below.

Optional: reamplification of size selected PCR products

16 Prepare the Master Mixes as detailed on page 13 of Procedure & Checklist for Ultra Low Input libraries. Adjust the volume of the sample to 48 µl with water or elution buffer (if the concentration post BP was not measurable, do not dilute the sample, just use the 25 - 27µl as is). Add PCR mixes and primers to the samples one by one (no pooled master mix), increase volumes of both PCR mixes to 52 µl. Reamplifications should run for a maximum of 5 cycles, for more information see Materials - section Reamplifications of post size selection PCR products.

17 Perform the purification as shown in the Procedure & Checklist for Ultra Low Input libraries on page 14. Extend incubation times to at least 10 minutes and elute in either ~15 µl (if you have original PCR product left to to pool with) or ~25µl of elution buffer (if you used all of the sample for

reamplification). 2 µl is kept for the QC.

4. Post PCR and post BluePippin library preparation

18 Proceed with Procedure & Checklist for Ultra Low Input libraries on pages 16 and 17. Reduce the amount of ligation mix to 20 µl. Add the Overhang Adapter v3 last. Incubate the mixture at 20°C for 1:40 h or alternatively overnight before storing them at 4°C.

19 Instead of the purification on page 18, perform the size selection purification detailed on page 15 and 16 of Procedure & Checklist for Low Input libraries, to get rid of leftover adapters and adapter dimers. Elute in 15 µl of PacBio EB.

If the Nuclease Treatment should be used, either elute in 97 µl of elution buffer (or adjust the volumes of the Nuclease Treatment Master Mix to a lower sample input).

5. Final library QC

20 Use 2 µl of the final library for the concentration measurement and fragment analysis. For notes on the concentration and fragment distribution of the final library see Materials - section Post PCR and post size selection library prep.

Final libraries should not be frozen anymore but rather stay at 4°C. We have successfully sequenced libraries that were kept at 4°C for two weeks.

Optional: Nuclease Treatment of SMRTbell Libraries

21 Follow the protocol as detailed on pages 13 - 14 of Procedure & Checklist for Low Input libraries. Adjust the elution volume of the bead purification to the volume needed for the subsequent sequencing preparation (usually 10-15 µl).