Sample preparation instructions for use on PacBio instruments

IMPORTANT INFORMATION

The PacBio® *RSII* and *Sequel* systems are based on single-molecule, real-time sequencing; it produces ultra-long reads beyond 50 kb. Most common library insert lengths are between 300 bp – 20 kb; sequencing of libraries over 30 kb is possible. Other advantages of PacBio® sequencing include library preparation without the amplification step, detection of base modifications, whole-length transcript sequencing etc.

Single molecule sequencing on the PacBio® requires stringent sample preparation procedures. Since amplification is not required, it is critical to maintain the highest sample quality in order to maximize both sequencing success and its output.

PacBio® technology has very specific requirements for both quality (molecule length and chemical purity) and quantity of the nucleic acids that are used for library preparation; these requirements are listed below. Please be warned that if any of these requirements are not met, NGI cannot guarantee that sequencing of your samples will be successful. In the case of poor sequencing performance due to template-related problems, we cannot reimburse you for either the library prep or the sequencing costs.

General requirements for templates, disregarding the sequencing application

- Samples must be stored at maximum -20°C
- Reduce number of freeze-thaw cycles to minimum
- Do not expose samples to temperatures above 65°C
- Do not expose samples to pH extremes (<6 or >9)
- OD260/280 between 1.8 and 2.0
- OD260/230 between 2-0 – 2.2
- There must not be any insoluble material in the prep
- There must not be RNA contamination
- Do not expose samples to UV or intercalating fluorsescent dyes
- Ensure that there are no chelating agents (e.g. EDTA), metal ions, denaturants, or detergents left in the sample (note: EDTA concentration in 0.1x TE buffer is OK)
- There must be no carry-over contamination that originates from the source organism or DNA-prep (phenols, salts, polysaccharides, lipids, pigments, etc)
- The majority of the DNA must be equal or greater than 40 kbp.

Always submit the following information

- Gel picture (0.8 % agarose for genomic, 1-1.5 % for amplicons; always add a size marker and inform NGI about marker's size values)
- Information about OD260/280 and OD260/230 ratios

- Estimation of concentration, name the measurement instrument
- Information about the extraction protocol
- Type of solvent that DNA is stored in (preferably, 0.1x TE or EB buffer)
- Filled in submission forms (obtained from the project coordinator at the sequencing facility, Olga V. Pettersson): as a print out together with the sample, as well as an electronic version sent within the Zendesk communication thread (see Appendix 2 to the Agreement document)

Samples submitted without this information will not be processed, nor any further notice will be given.

All new projects must be discussed with the project coordinator and/or respective bioinformaticians of the sequencing platform before sending the samples.

All PacBio sequencing requests must be done through the Genomics Portal: https://portal.scilifelab.se/genomics/.

Delivery method

- Always use dry ice to deliver the samples (also amplicons).
- Use an appropriate courier service.
- If DNA samples are shipped in tubes, these should be capped tightly to prevent accidental spillage or cross contamination. Sealing tops by wrapping Parafilm® wrap as an additional safeguard is recommended. Please ship DNA samples in a secondary container (e.g., a cardboard freezer box) with adequate padding.
- **Please DO NOT write any recipient names on the package; address it ONLY to the Uppsala Genome Center.**
- Include printed copies of the filled in submission forms in the package.

Specific applications

DNA: quantity and quality; sample-specific issues

We will require a variable amount of good quality DNA depending on the size of your insert for sequencing. Please consult the table below for estimates of the material needed. Please be aware that excess DNA is needed since initial fragmentation and AMPure bead concentration/clean up steps can be subject to unavoidable template loss. Often the loss of template is less than expected and so if unavoidable, lower input amounts can be accommodated. If the template is very limited/low, we can process these samples using our low-input or express library protocols. The low-input libraries are not without their disadvantages and are only recommended where absolutely needed. Upon library preparation we can then provide you with an appropriate answer as to how many SMRT cells can be loaded using your specific library. For sequencing of amplicons please consult the section named amplicon sequencing below. Sequencing of amplicons does not require as much input DNA since no fragmentation is need prior to library preparation.

INSERT SIZE	Amount needed (estimated by Qubit)
500bp	500ng
1kb	1ug
2kb	1.5ug
5kb	4ug
10kb	5-10ug
20kb	at least 10 ug
25+ kb	20 ug
30+ kb	at least 20 ug

For *de novo* projects with PacBio data alone, 75-100x sequencing is recommended (c:a 30-50x per allele). This is easily achieved for genomes no-larger than 500 Mbp. Larger genomes may also be accommodated, however specific plans must be made with the project coordinator. For population sequencing and studies of structural variants, 5-10x coverage of the genome is enough.

For re-sequencing and *de novo* sequencing in combination with short-read data (Illumina or Ion, as well as 10x Chromium), 10-30x sequencing depth is recommended depending on the genome size and type of available data.

Human and animal samples

DNA extraction from humans and animals is usually much easier than from any other type of material; use of appropriate QiaGen (e.g. MagAttract) or Zymogen kits preserving 20+ kb fraction, following elution with the supplied EB buffer is recommended.

Bacteria and yeasts

Grow cells to **midlog phase** and do not grow them to stationary phase. Grow enough volume of liquid growth to compensate for the lower OD_{600} growth point.
Use the QIAGEN Genomic-tip 500/G kit to extract DNA.

If a clear threadlike precipitate is observed upon addition of isopropanol during the precipitation step then **spooling the DNA and dip washing into 70% EtOH is recommended.**

Spool DNA eluted from Qiagen Genomic Tip columns.
Aim for a genomic DNA yield of 100 µg. If your eluted DNA is of high MW, white DNA strands will immediately come out of solution upon adding isopropanol. These strands will rapidly coalesce upon mixing the tube by inversion. Fish this spool of DNA out with a small pipet tip and transfer it to an eppendorf tube containing 70% EtOH to wash. After soaking for a few minutes in 70% EtOH, fish the spool out again and transfer it to an empty eppendorf tube. Air-dry until no visible droplets of ethanol wash are left.

This procedure is critical in order to get an accurate concentration measurement, which is particularly vital to the success of PacBio sequencing. To 100 µg of spooled DNA pellet, overlay 200 µl of 5 mM Tris pH 7.5 0.5 mM EDTA and incubate for 1 hr. Tap the tube to try to dislodge what will be an invisible pellet. Incubate overnight at room temperature. Repeat the tapping and overnight incubation step two more times.

If isolating DNA via ethanol precipitation, exposure to extended heat should be minimized. Incubation at 65° C for 1 hr may incur DNA damage resulting in impaired sequencing performance. Air-drying of pellets is preferred over heat- or vacuum drying.

For Gr+ bacteria avoid the bead-beating protocol. Additional post-extraction cleanup with QiaGen DNeasy PowerClean Pro Cleanup kit is recommended.

Plants and filamentous fungi

DNA extraction from plants and micromycetes possesses multiple challenges due to high levels of polysaccharides, phenolic compounds and a variety of secondary metabolites in the starting material. It is therefore recommended to use high-salt, 1% CTAB clean-up of such samples. You can either use CTAB-containing extraction buffer or use an additional CTAB clean-up of the DNA precipitate. CTAB removes the majority of the polysaccharides and some lipids, however the CTAB itself must be carefully removed from the prep, since it can inhibit the sequencing process.

If phenol-chloroform DNA extraction was performed, make sure phenol has been completely removed. Phenol removal can be achieved by running the DNA through a QiaGen column~~s~~ tailored for genomic DNA preparation (e.g. QiaQuick genomic DNA kit). Phenol has an absorption spectrum between 270-275 nm, which is very close to that of DNA. Phenol contamination can artificially inflate spectrophotometer readings and mimic both higher purity and concentration of the sample.

The QIAGEN Genomic-tip 500/G or MagAttract kits are highly recommended even for plant and fungal material, as well as organism-appropriate Zymogen kits. An additional CTAB clean-up is highly recommended as well. Make sure not to overload the columns with start material.

As an alternative, QiaGen DNeasy PowerClean Pro Cleanup kit is designed to make a post-extraction purification of DNA from troublesome material with a suspected high content of polysaccharides, phenolics and secondary metabolites. Beware: if DNA is heavily contaminated, the sample loss at this point might reach up to 90%.

RNA-seq (Iso-seq)

Iso-seq allows sequencing of full-length transcrips from the 5' end to the 3'end. This technique is proving to be invaluable for the annotation of *de novo* genomes. Furthermore, iso-seq is unequivocally contributing to the discovery and analysis of various isoforms both novel and rare.

The user must prepare their own cDNA containing full-length transcripts using the SMARTer PCR cDNA synthesis kit from Clontech (cat. 634925). Input RNA must be of a high quality and rRNA should be removed from the sample by either polyA-selection with a kit of choice, or by rRNA depletion of total RNA with an organism-appropriate RiboZero or RiboMinus kit (RiboZero is usually more efficient for bacterial and fungal samples). The Clonetech kit can accommodate very low input RNA amounts and so is suitable for applications where RNA is limited. A minimum 2ng total RNA or 1ng polyA RNA can be used. If material is not limited, we do recommend using 1ug total RNA or

0.5ug polyA RNA for cDNA synthesis (please see kit for details). For subsequent amplification steps, it is of upmost importance not to over amplify your template. (Please follow the recommendations as outlined the PacBio isoseq protocol at http://www.smrtcommunity.com/Share/Protocol?id=a1q70000000HqSvAAK&strRecordTypeName=Protocol). Please make sure to use the Phusion high fidelity DNA polymerase or Kapa HiFi from NEB instead of the Advantage 2 kit as outlined in the Clonetech kit insert.

The amount of SMRT cells to be run or coverage needed depends on your scientific question. Of course this is very sample dependent and specific plans can be made to run a pilot project of your samples before running extra SMRT cells.

Amplicon-seq

PacBio instruments can process amplicons from 250 bp to 13 kb in size. Barcoding and pooling of amplicons is also possible; please contact us for more details prior to designing your experiment. At least 500 ng of short amplicons (up to 1 kb) or 1-3 ug of long amplicons should be submitted. In some cases where concentration is an issue, smaller inputs can be accommodated through the use of our low input protocols. Low input protocols do have their disadvantages and use of these protocols must be discussed prior to sequencing.

Never send amplicons as a precipitate in iso-propanol. Always send amplicons on dry ice. Long amplicons must always be stored at -20°C.