Importance of Sample QC

Susana Wang, Kristi Spittle, Primo Baybayan
Pacific Biosciences, 1380 Willow Road, Menlo Park, CA 94025

Introduction

Single molecule sequencing places stringent sample requirements for sequencing on the PacBio® *RS*. Since amplification is not required, it is critical to maintain the highest sample quality in order to maximize sequencing. The introduction of the MagBead station with the PacBio *RS* addresses some sequencing challenges such as adapter dimer and short-insert contaminations, but it does not address problems associated with carry-over contaminants that could inhibit downstream sequencing. To increase success with your samples, we recommend several QC steps that you can perform prior to the construction of libraries. Here we present several cases on how to properly QC your samples upfront and provide additional sample purification steps which can help yield successful sequencing runs.

Methods

Before constructing SMRTbell™ libraries, it is important to characterize your samples as thoroughly as possible. The more information you have on your sample, the higher your success will be with PacBio *RS* sequencing.

Here are the questions we ask our collaborators and customers, when we receive a sample, and the QC procedures we take to assess for contaminants.

1. What is the source of the DNA (bacterial, tissues, blood, etc.)? Knowing this information may be critical in upfront-QC steps. For example, if the sample is from plants known to have high polysaccharides, they may be co-extracted and a clean-up procedure may be necessary to remove them from the sample.
2. What methods were used for DNA isolation? Does the isolation method properly remove salts or detergents, such as CTAB, that may not be easily removed otherwise? If not, performing an additional clean-up step here may save a SMRT® Cell further downstream.
3. Have you performed quantitation using both Qubit® and NanoDrop® systems? Both quantitation methods have their advantages and disadvantages. If your sample quantitation is very similar on both the Qubit and the NanoDrop systems, then this may provide greater confidence that the DNA is pure. If the readings are different, it may indicate a possible contaminant. Even if the contaminant is not known, it will be easier to determine whether the results of an unsuccessful instrument run are due to the quality of the sample or the instrument performance.
4. Have you run gels to assess quality of the gDNA? It is important to know whether you're working with RNA, DNA or degraded DNA.
5. Have you performed clean-up methods deemed necessary prior to shearing? It is best to clean up DNA prior to shearing. The cleaner the DNA is, the better your process will be.

As can be seen, the more information you have on a sample, the better equipped you are for a successful PacBio *RS* run.

Case Study I

Project:

A set of five samples arrived from one of our customers. Project goal was *de novo* sequencing coupled with methylation analysis of all five samples.

Information From Customer:

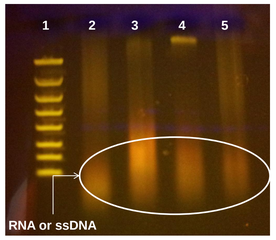
- Samples were lyophilized with a heated vacuum
- No gel pictures were provided
- Below is the sample information provided to us:

Sample ID	Amount (µg)	A260/A280	A260/A230
A	16	2.060	2.043
B	24	2.000	2.037
C	21	2.082	2.047
D	50	2.130	2.070

In-house QC:

Samples were rehydrated in 200 µL of EB Buffer (Qiagen®), we took our own readings as well as running a gel analysis.

		NanoDrop Reading				Qubit Reading	
Tube Label	Vol (µL)	[ng/µL]	260/280	260/230	Total ng	[ng/µL]	Total ng
A	200	332.6	2.16	1.52	66520	26	5200
B	200	295.1	2.15	1.65	59020	21.6	4320
C	200	205	2.18	2.08	41000	13.8	2760
D	200	178.5	2.06	0.21	35700	37.6	7520



1 = Lonza FlashGel® Ladder (0.1 to 4.0 kb)
2 = A
3 = B
4 = C
5 = D

Lonza FlashGel system has a unique feature in which degraded ssDNA or RNA looks diffuse and orange in color whereas dsDNA (or total RNA) looks bold and yellow in color.

Resulting Action:

The 260/280 readings from our customer, our own readings, plus the gel all pointed toward RNA contamination and not gDNA. We did not proceed with SMRTbell™ library preparation after the QC assay, but instead asked that the customer re-isolate their gDNA. The second set of DNA passed QC and sequenced well.

Case Study II

Project:

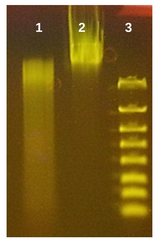
For this sequencing project, we wanted libraries that would produce the longest reads possible for hybrid assembly.

Information from customer:

- Sample was stored in a freezer for at least a year.
- The sample was phenol-chloroform extracted and no other QC data was given.
- This particular organism has an extremely poor reference due to its numerous and long repeats.

In-house QC:

		NanoDrop Reading				Qubit Reading	
Tube Label	Vol (µL)	[ng/µL]	260/280	260/230	Total ng	[ng/µL]	Total ng
E	500	131.2	2	1.74	65,600	36.8	18,400



1 = E
2 = F (not important to Case Study Two)
3 = Lonza FlashGel® Ladder (0.1 to 4.0 kb)

Lonza FlashGel system with the Qubit and NanoDrop readings indicates the presence of RNA and degraded DNA.

Resulting Action:

Since there is only a poor reference genome for this organism, we went ahead and turned the entire smear into a library for Circular Consensus Sequencing (CCS). Library yield was 36%. Project is ongoing.

Case Study III

Project:

De novo sequencing of a bacterial genome.

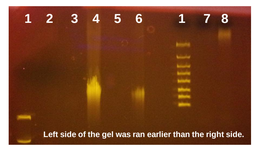
Information from customer:

- Bacterial genome of roughly 4 MB.
- Two lyophilized samples with different methods of isolation:
  1. Promega® Kit (G)
  2. Homebrew isolation method (H)

In-house QC:

Samples were rehydrated in 200 µL of EB Buffer.

		NanoDrop Reading				Qubit Reading	
Tube Label	Vol (µL)	[ng/µL]	260/280	260/230	Total ng	[ng/µL]	Total ng
G Kit	200	747.8	1.89	2.08	149,560	678	135,600
H	200	779.7	1.91	2.06	155,940	146	29,200



1 = Lonza FlashGel® Ladder (0.1 to 4.0 kb)
2 & 7 = I (Sample is not associated with this case study.)
3, 5 & 8 = G (Loaded at different concentrations)
4 & 6 = H (Loaded at different concentrations)

Lanes 7 & 8 contain sample that has been AMPure® system purified & concentrated.

- Qubit® and NanoDrop® system results show that the sample isolated from the Promega® Kit (G) is mostly DNA, and the sample isolated using the In-House method (H) is mostly degraded DNA or RNA.
- FlashGel® system data indicates H is actually completely degraded as no high molecular weight bands can be found on the gel.

Resulting Action:

- Sample H was not taken through library preparation.
- Sample G was made into 10 kb and 1 kb SMRTbell™ libraries for Continuous Long Read (CLR) and CCS respectively.
- A two SMRT® Cell sequencing run for the CLR library was completed in the PacBio *RS* to ensure sequencing yield prior to a larger run.

Importance of Sample QC

Susana Wang, Kristi Spittle, Primo Baybayan
Pacific Biosciences, 1380 Willow Road, Menlo Park, CA 94025

Case Study III continued

Sample ID	Yield
G	30%
0.9 K Lambda Ctrl	38%

Sample Name	Cell #	Reads
G 10K Lib	1	7,742
G 10K Lib	2	7,993
0.9 K Lambda Ctrl	1	40,836

The initial run shows that there is something in the sample inhibiting sequencing. Since we did not know the bacterial source of this sample and we did know that the sample was lyophilized, we tried a couple of purification methods to clean up the sample to improve sequencing.

Two additional clean up methods were performed:

1. Phenol/chloroform extraction followed by ethanol precipitation (PC)
2. Qiagen® DNeasy® Kit (Q)

Sample	Yield
G 10K PC	27%
G Qiagen 1K	26%
G 1K PC	28%
Lambda Ctrl 1K	30%

Sample Name	Cell #	Movie	Reads
G 10K PC	1	1	20859
G 10K PC	2	1	26183
G Qiagen 1K	1	1	28063
G Qiagen 1K		2	21637
G Qiagen 1K	2	1	26458
G Qiagen 1K		2	18799
G 1K PC	1	1	40139
G 1K PC		2	32725
G 1K PC	2	1	46875
G 1K PC		2	45916
Lambda Ctrl 1K	1	1	44043
Lambda Ctrl 1K		2	40223
Lambda Ctrl 1K	2	1	41886
Lambda Ctrl 1K		2	39696

The sequencing results from this study show that both purification methods increased the loading from the original 7 kb reads. However, the phenol/chloroform extraction improved the sequencing yield the most.

MO BIO® PowerClean® Kit

It is known that phenol/chloroform may cause problems with downstream sequencing when phenol is not completely removed. See the poster “Sample Quality-Effects of Contaminants on SMRTbell™ Library Preparation and Sequencing.” The MO BIO® PowerClean® Kit was also looked at as an alternative clean-up method. The PowerClean Kit is supposed to remove a variety of contaminants including polysaccharides, poly-phenols, and heme. These qualities add to the kit’s attractiveness in that it can remove other contaminants that phenol/chloroform is unable to remove.

For this project, the sample was cleaned with the MO BIO PowerClean (MBPC) or an AMPure kit (AMP) prior to shearing.

Sample	Yield
G 10K MBPC	43%
G 10K AMP	44%

Sample Name	Cell #	Reads
G 10K MBPC	1	39463
G 10K MBPC	2	38804
G 10K AMP	1	19622
G 10K AMP	2	21155

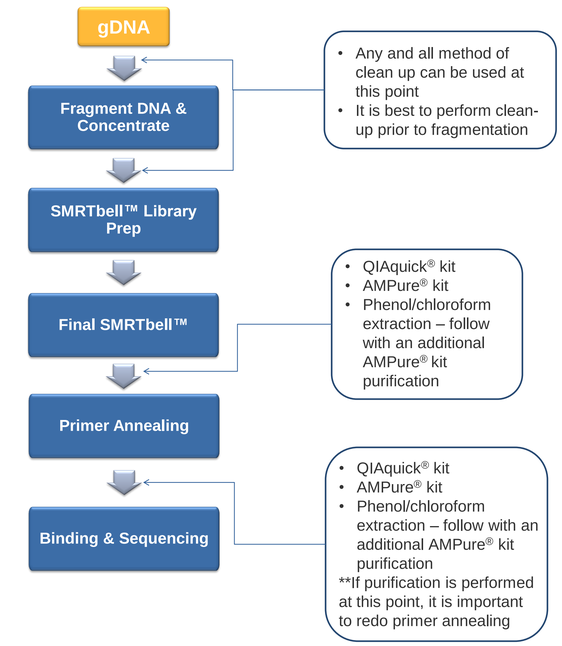
Recommendations / Discussion

In this poster, we demonstrated 3 different cases showing the importance of performing sample QC, and described different sample clean-up methods. Here are some of the methods used in the case studies, and the advantages and disadvantages of each of the methods.

Clean Up Method	Advantages	Disadvantages
Qiagen® Column (DNeasy® / QIAquick®)	• Quick, easy, convenient. • Can be used on final SMRTbell™ library	• Yields can vary greatly
Phenol / Chloroform Extraction and Ethanol Precipitation	• Yields are very consistent. Often >90% recovery of DNA • Common protocol and removes a variety of contaminants • Can be time-consuming, taking at least a few hours • Can be used on final SMRTbell™ library	• Needs a hood to perform phenol/chloroform extraction • If not properly cleaned, carry over phenol may negatively impact sequencing
AMPure® Beads	• Yields are very consistent. Often >90% recovery of DNA • No extra kit purchase required • Scales more easily for large amounts of DNA. • Can be used on final SMRTbell™ library	• Will not remove many contaminants • Requires an extra AMPure step
MO BIO® PowerClean® kit	• Readily available kit designed for re-purifying extracted gDNA • Clean up takes about 30-45 min • No hood required • Removes a variety of contaminants	• Does not scale for multiple samples or large amounts of DNA • Average yield tends to vary between 20-50% • Cannot be used on final SMRTbell™ library due to damage caused during the clean-up

Recommendations / Discussion

The general observation has been that the cleaner the gDNA, the better the quality of the library. Therefore, the earlier the purification is done, the better. In some instances, a simple AMPure® kit cleanup on the gDNA can improve even the fragmentation process. Below is a workflow to indicate at which step various clean up methods can be used.



As longer-insert libraries are sequenced on the PacBio RS, clean, high-quality DNA will become more and more critical. The best approach is to start with an effective DNA isolation method. It is then important to know the quality of the sample so that the appropriate clean up steps can be performed to remove contaminants.