



Review Article

Rapid quantification of DNA libraries for next-generation sequencing[☆]

Bernd Buehler^{a,*}, Holly H. Hogrefe^a, Graham Scott^b, Harini Ravi^b, Carlos Pabón-Peña^c, Scott O'Brien^a, Rachel Formosa^a, Scott Happe^b

^aAgilent Technologies, 11011 N. Torrey Pines Road, La Jolla, CA 92037, USA

^bAgilent Technologies, 1834 State Hwy 71 West, Cedar Creek, TX 78612, USA

^cAgilent Technologies, 5301 Stevens Creek Blvd., Santa Clara, CA 95051, USA

A B S T R A C T

The next-generation DNA sequencing workflows require an accurate quantification of the DNA molecules to be sequenced which assures optimal performance of the instrument. Here, we demonstrate the use of qPCR for quantification of DNA libraries used in next-generation sequencing. In addition, we find that qPCR quantification may allow improvements to current NGS workflows, including reducing the amount of library DNA required, increasing the accuracy in quantifying amplifiable DNA, and avoiding amplification bias by reducing or eliminating the need to amplify DNA before sequencing.

© 2010 Published by Elsevier Inc.

1. Introduction

Next-generation sequencing (NGS) platforms such as the Illumina Genome Analyzer, Roche 454-FLX, or ABI SOLID have revolutionized genomics by producing hundreds of megabases of sequence information in a single run. In order to optimize the amount of sequencing information, and thereby reducing the cost per base pair sequenced, it is vital to optimize the amount of a prepared DNA library for a sequencing run. For example, for the Illumina Genome Analyzer, the DNA library is immobilized by hybridization on a chip and amplified in situ in a process termed cluster generation. If the amount of DNA loaded is too high, the DNA clusters generated will overlap and thereby affect the quality of sequencing data. Loading a suboptimal amount of DNA results in a low cluster density, and reduces sequencing efficiency [1]. One current solution is to measure the amount of library material using the Agilent 2100 Bioanalyzer with the High Sensitivity DNA Kit for quantification of sequence-ready libraries down to the low pg/μl range. In addition, other DNA quantification methods such as UV spectrophotometry, or fluorescent nucleic acid stains are widely used. However, it is important to note that with these methods, DNA fragments lacking the necessary adapters for cluster generation will also be measured. If DNA fragments missing adapter sequences are present, this can result in a lower cluster density than expected when a standard concentration of DNA is loaded

onto the cluster generation station. Here, we highlight the use of qPCR for accurate determination of library quantity at high sensitivity of detection.

2. Library preparation

DNA library preparation followed the standard Illumina protocol for paired-end sequencing with a few modifications [2]. Briefly, 3 μg of genomic DNA (Coriell) was sheared using a Covaris E210 instrument to a median fragment size of 200–250 bp. The products were end-repaired, 3' non-template A's were added, and paired-end adapters were ligated. After size selection on a 4% Nusieve 3:1 agarose gel and QiaQuick gel extraction, 23 μl of the library was amplified by 6–8 cycles of PCR using the Illumina PE 1.0 and 2.0 containing sequences required for cluster generation on the flow cell.

3. SureSelect enrichment for targeted resequencing

We recently released the Agilent SureSelect Target Enrichment System that provides specific enrichment of user-defined subsets of a genome [3,4]. The method is based on hybridization of genomic DNA libraries to custom biotinylated 120mer RNA probes and subsequent immobilization on magnetic beads, followed by wash and elution steps. To verify the process we tested enrichment of several libraries with different RNA capture probe sets specific to the human X chromosome, all human exons, or regions on chromosome 4. After elution of the captured DNA fragments, the library was reamplified for 12–14 cycles of PCR with SureSelect Illumina-specific primers. Amplification enables accurate quantification using the Bioanalyzer High Sensitivity chip before sequencing.

[☆] This application note has been provided by Agilent Technologies as supplemental educational material to this thematic special issue. This application note was sponsored by Agilent Technologies and has not undergone a peer review process within Elsevier.

* Corresponding author.

E-mail address: bernd.buehler@agilent.com (B. Buehler).

4. Bioanalyzer quantification

Quantification of the enriched libraries was done using a High Sensitivity DNA kit on the Agilent 2100 Bioanalyzer according to the manufacturer's instructions [5]. Briefly, the reamplified material was diluted 1:50 and 1:100 and 1 μ l was run on a primed chip along with DNA markers for size determination and quantification. The concentration was determined on fragment sizes from 160 to 400 bp using the Bioanalyzer software. The data were corrected for dilution and averaged.

5. qPCR quantification

Measurement of DNA libraries by qPCR should give an accurate quantity of library material [4]. An added benefit of this method is that only DNA with adapters ligated to both ends can be measured, as only these fragments can be amplified to generate material for sequencing. This minimizes overestimation of the DNA concentration from fragments with no or only one adapter. Principally, two sources of a standard can be used for quantification. A dilution series of a previously quantified DNA library that was successfully

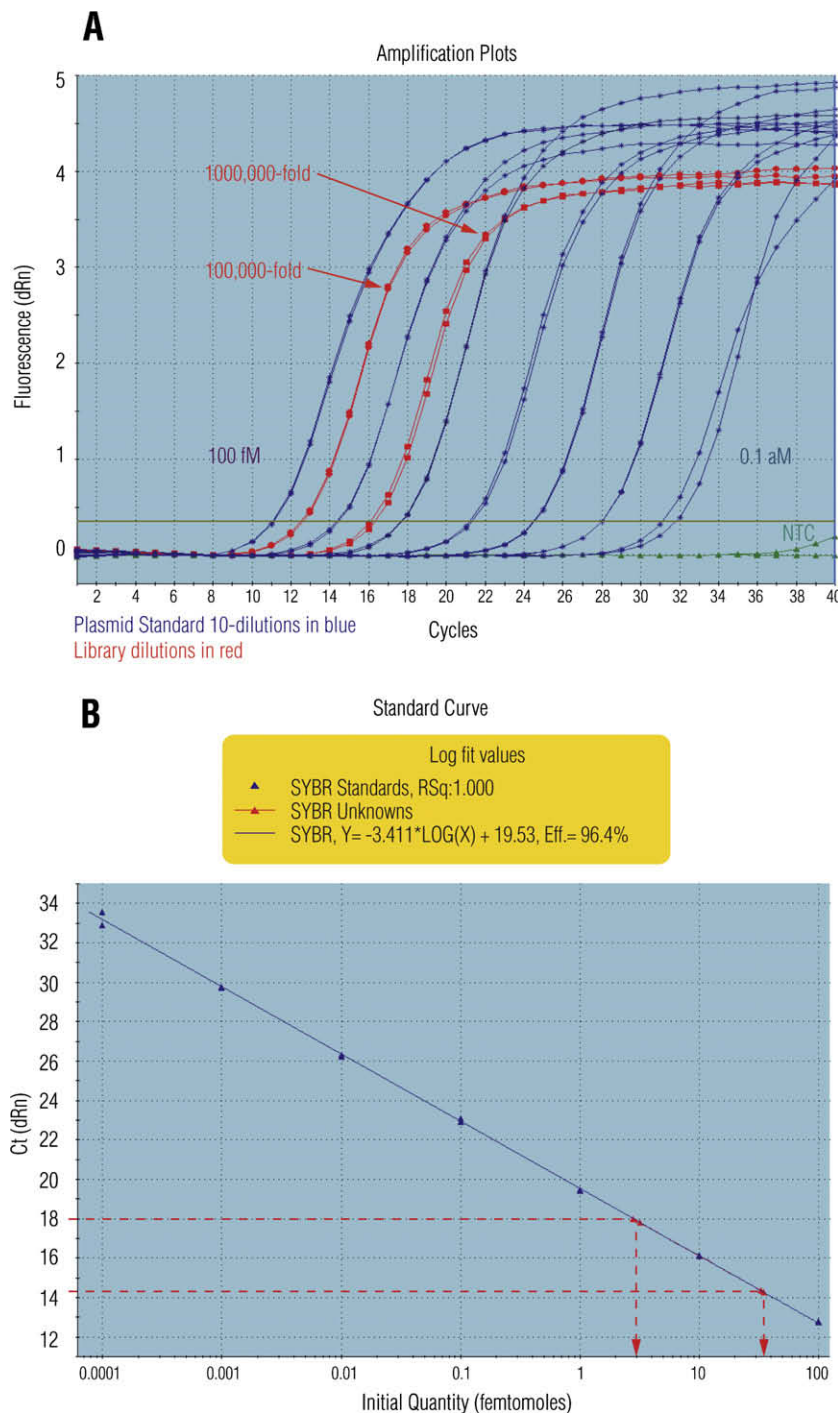


Fig. 1. Linearity and sensitivity of QPCR using a plasmid DNA standard: (A) Linearized plasmid standard with a 170 bp fragment between the Illumina paired-end adapters was diluted into 20 μ l of PCR mix to a final dilution between 100 femtomolar and 0.1 attomolar final concentration. As an example, sample 4 from Table 1A is shown at two different dilutions (red). MxPro software was used to calculate the concentrations of the sample dilutions and correct for the dilution factor to estimate starting concentration. (B) Standard curve from the same experiment and an illustration of how concentration can be read from the standard curve.

sequenced on the appropriate platform can be used as a standard. If a library of similar properties is chosen, differences in amplification of the standard and the test material are minimized and will result in accurate quantification. Alternatively, a plasmid standard can be used to measure the amount of DNA library. This approach has the advantage that the standard can be easily reproduced in the future and can be used in many labs to compare independent measurements of different DNA libraries. A concern, however, is the difference in complexity between a plasmid-based standard and the library sample. In addition, the library might contain elements that are difficult to amplify, such as inserts rich in G and C bases. If there is a negative bias toward GC-rich inserts, the quantification of a GC-rich library might be hampered.

We generated a plasmid standard by cloning a portion of a human genomic Illumina paired-end DNA library into StrataClone vector (pSC-B, Stratagene) and randomly selecting clones with different inserts. A clone with a 170 bp insert between the paired-end adapters and 48% GC content was chosen as a reference standard, because this clone represents a typical insert size and GC content. To investigate whether sequence and length of the insert influences amplification efficiency or linearity, we tested four additional clones with insert lengths of 70, 81, 193, and 283 bp and GC content of 24%, 43%, 47%, and 61%, respectively. All five clones tested showed a similar efficiency of approximately 100% in qPCR reactions under standard conditions with Brilliant II SYBR Green QPCR Master Mix (Agilent Technologies-Stratagene products, data not shown).

To prepare the plasmid standard for routine use, the DNA was quantified by spectroscopy and linearized by digestion with EcoRI. The standard curve was generated by diluting a 100 pM stock solution of the standard in 0.1% Tween 20 to 1 pM followed by 6 more 10-fold dilution steps. Two microlitres of each dilution was amplified with primers to the distal end of the Illumina adapters P5 (5'-AATGATACGGCGACCACCGA) and P7 (5'-CAAGCAGAAGACGGCA TACGA) at 400 nM each. We used Brilliant II SYBR Green QPCR Master Mix or the new Brilliant III Ultra-Fast SYBR Green Master Mix in a total volume of 20 μ l on the Agilent Mx3005P QPCR system. Cycling conditions for Brilliant II were 10 min activation at 95 °C, 40 cycles at 30 s 95 °C, 60 s 60 °C and a melt curve from 70 to 98 °C. Cycling conditions for Brilliant III Ultra-Fast were 3 min activation at 95/98 °C, 40 cycles at 10 s 95/98 °C, 20 s 60 °C, 20 s at 72 °C and a melt curve from 72 to 98 °C.

Fig. 1A shows a dilution series of the plasmid standard from a final concentration of 100 nM to 0.1 aM in duplicate (blue curves). As an example, two dilutions from a SureSelect enriched library were overlaid in red (Sample 4 in Table 1A). A signal from a non-

template control (NTC) is not observed and the standard curve is linear over six orders of magnitude (Fig. 1B). This assay is orders of magnitude more sensitive than necessary, as loading of the cluster generation station is typically done at 2–8 pM of denatured DNA library in a total volume of 120 μ l.

Table 1A shows a comparison of the quantification of paired-end libraries measured with the Bioanalyzer (column 4) and by qPCR as described above (column 5). Overall DNA concentrations are generally similar for both methods, and estimates obtained by qPCR quantification are typically slightly higher. For this experiment the libraries were loaded for cluster generation and sequencing on the Illumina GA IIx at 8 pM based on Bioanalyzer quantification after denaturation. The achieved cluster density was in the recommended range of 150,000–20,000 clusters/tile. The results indicate that, for typical library preparations, qPCR quantification is an acceptable orthogonal technique to Bioanalyzer analysis.

6. qPCR for GC-rich fragments

A concern with using a plasmid-based standard is that a DNA library with very different properties might not be accurately quantifiable. Specifically, libraries with very high GC content might result in an underestimation of the concentration because inserts with high GC content may not amplify efficiently in the qPCR assay. We previously demonstrated that amplification efficiency is lower in qPCRs employing targets of higher % GC content [6]. To address this issue, we generated two GC-rich fragments from human insulin-like growth factor binding protein 3 (IGFBP3) and attached Illumina paired-end adapter sequences by PCR using Herculase II Fusion DNA polymerase (Agilent Technologies- Stratagene Products) in the presence of 7% DMSO. These GC-rich PCR fragments were compared to two control fragments with qPCR under standard conditions for the Brilliant III Ultra-Fast SYBR Green Master Mix. As demonstrated in Table 1B, two control templates containing 42.7% and 61.3% GC content were amplified with nearly 100% efficiency (R-Squared close to 1.000). However, both GC-rich fragments failed to amplify. Simply increasing the denaturation temperature to 98 °C resulted in successful amplification of both fragments despite a GC content exceeding 78%. Amplification efficiency was slightly lower in qPCR reactions employing a 98 °C denaturation temperature, which may reflect loss of *Taq* DNA polymerase activity at higher temperature. Based on this observation, increasing the denaturation temperature for all qPCR quantifications is recommended.

7. Concluding remarks

Our results show that libraries can be readily quantified by qPCR using the Agilent Mx3005P QPCR system. One benefit of qPCR is the ability to only detect adapter-containing sequences since identical primer sequences are used for library amplification and qPCR quantification. Moreover, single-strand products resulting from linear amplification (e.g., misbalanced PCR primers) are

Table 1A

Comparison of quantification using Bioanalyzer High Sensitivity DNA chip and qPCR with Brilliant II SYBR Green Master Mix and P5/P7 paired-end adapter primers.

Sample	DNA library source	Bait	nM Bioanalyzer	nM QPCR	Ratio QPCR to BA	Cluster density (clusters/tile)
1	NA15510	X chromosome	56.39	79.66	1.41	156,252
2	NA10831	X chromosome	35.39	71.29	2.01	202,748
3	NA15510	Control	43.97	65.82	1.50	155,653
4	NA15510	1 MB Chr 4	21.61	29.02	1.34	128,251
5	NA18507	Human all exon	5.50	7.22	1.31	167,565
6	NA18507	Human all exon	5.51	8.50	1.54	179,478
7	NA10831	Human all exon	5.95	8.79	1.48	170,645
8	NA10831	Human all exon	4.89	7.55	1.54	178,033

Table 1B

qPCR of a 10-fold dilution series of PCR fragments with different inserts located between the Illumina paired-end adapters.

Insert length	% GC	% Efficiency		T_m		
		95 °C Denaturation	98 °C Denaturation			
80	42.7	0.997	108.1	0.999	97.3	84.2
295	61.3	0.996	105.8	0.984	89.3	89.4
168	79.8	0.251	>1000	0.999	85.5	95.3
237	78.1	0.58	>1000	0.99	95	95.6

detected by qPCR, which can lead to a more accurate estimate of amplifiable DNA.

A noteworthy benefit of qPCR quantification is its sensitivity, which allows researchers to monitor DNA library amount at each step of a next-generation sequencing workflow. For example, when using the SureSelect enrichment protocol, we can accurately measure the amount of DNA eluted from the capture beads and follow the fate of the library at all the steps in the protocol (not shown). This could potentially open up the possibility to sequence enriched libraries without an additional amplification step.

Notably, when using the current standard protocol, only a small fraction of the library preparation is loaded onto the sequencer. The high sensitivity of detection of qPCR provides quantification of libraries that have undergone only a few rounds of PCR amplification or no PCR amplification at all. A significant advantage of avoiding PCR amplification altogether is that this reduces the potential for bias introduced by PCR, virtually eliminates the number of duplicates, and simplifies mapping for de-novo sequencing [7]. An added benefit of these PCR-free libraries is a shorter and simplified workflow for next-generation sequencing. However, libraries produced without amplification will be of limited quantity and contain a significant amount of DNA fragments with no or only one sequencing adapter. Hence, the portrayed qPCR assay appears to be an ideal solution for quantification of PCR-free sequencing libraries.

Acknowledgments

The authors express gratitude you to their colleagues Ruediger Salowsky, Fred Ernani, and Knut Wintergerst for their comments, suggestions, and valuable input.

References

- [1] Michael A. Quail, Iwanka Kozarewa, Frances Smith, Aylwyn Scally, Philip J Stephens, Richard Durbin, Harold Swerdlow, Daniel J. Turner, *Nat. Methods* 5 (2008) 1005–1010.
- [2] Paired-End Sequencing Sample Preparation Guide. Available from: <https://icom.illumina.com/load.axd?type=Download&fid=2jMzHURgN0OWI3QIN5dyLQ&did=vCFO-bin9ES31x5cPVYFmA&name=Paired-End_SamplePrep_Guide_1005063_B>.
- [3] Andreas Gnirke, Alexandre Melnikov, Jared Maguire, Peter Rogov, Emily M LeProust, William Brockman1, Timothy Fennell, Georgia Giannoukos, Sheila Fisher, Carsten Russ, Stacey Gabriel, David B Jaffe, Eric S Lander, Chad Nusbaum, *Nat. Biotechnol.* 27 (2009) 182–189.
- [4] SureSelect User Manual. Available from: http://www.chem.agilent.com/Library/usermanuals/Public/G3360-90010_SureSelect_Protocol_v1.2.pdf.
- [5] Agilent High Sensitivity DNA kit guide. Available from: <http://www.chem.agilent.com/Library/usermanuals/Public/G2938-90321_HighSensitivityDNA_KG.pdf>.
- [6] Bahram Arezi, Weimei Xing, Joseph A. Sorge, Holly H. Hogrefe, *Anal. Biochem.* 312 (2003) 226–235.
- [7] Iwanka Kozarewa, Zemin Ning, Michael A. Quail, Mandy J. Sanders, Matthew Berriman, Daniel J. Turner, *Nat. Methods* 6 (2009) 291–295.