Lawrence Berkeley National Laboratory
LBL Publications

Title
Overcoming some of the challenges to single cell genomics

Permalink
https://escholarship.org/uc/item/3d18t54n

Authors
Lee, Janey
Tighe, Damon
Wang, Mei
et al.

Publication Date
2010-03-24

eScholarship.org Powered by the California Digital Library
University of California

Overcoming some of the challenges to single cell genomics

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

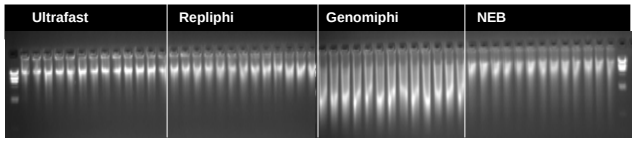
Janey Lee, Damon Tighe, Mei Wang, Stephanie Malfatti, Erika Lindquist, Feng Chen, Jan-Fang Cheng and Tanja Woyke
US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598 USA

Abstract

Single cell genomics, the amplification and sequencing of genomes from single cells, can provide a glimpse into the genetic make-up and thus life style of the vast majority of uncultured microbial cells, making it an immensely powerful and increasingly popular tool. This is accomplished by use of multiple displacement amplification (MDA), which can generate billions of copies of a single bacterial genome producing microgram-range DNA required for shotgun sequencing. Here, we would like to address several challenges inherent in such a sensitive method and propose solutions for the improved recovery of single cell genomes. While DNA-free reagents for the amplification of a single cell genome are a prerequisite for successful single cell sequencing and analysis, DNA contamination has been detected in various reagents, which poses a considerable challenge. Our study demonstrates the effect of UV radiation in efficient elimination of exogenous contaminant DNA found in MDA reagents, while maintaining Phi29 activity. Second, MDA is subject to amplification bias, resulting in uneven and sometimes insufficient sequence coverage across the genome. In a post-amplification method, we employed a normalization step within 454 Titanium library construction in which populations of highly abundant sequences were specifically targeted and degraded from the library via duplex-specific nuclease, resulting in decreased variability in genome coverage. While additional challenges in single cell genomics remain to be resolved, the two proposed methodologies are relatively quick and simple and we believe that their application will be of high value for future single cell sequencing projects.

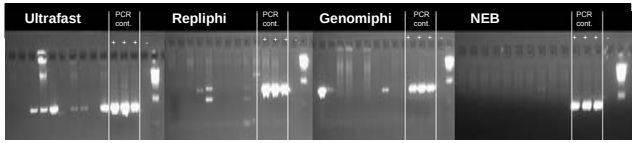
Contamination within commercial MDA kits

To screen commercial MDA kits for contamination, 20 µl MDA reactions were created from the REPLI-g® Ultrafast, RepliPHI™, GenomiPhi HY Kits and a homemade kit based upon NEB's phi29 enzyme. Reactions were incubated for 20 hours at 30°C.



MDA products from manufacture's recommended kit concentrations with no template DNA. 1 µl of product on a 1% agarose gel, EtBr staining, 120V, 45 minutes.

MDA products were screened for contamination via 16S/18S PCR. All the commercial kits were positive for 16S hits, but had none for 18S. Sequencing of the 16S contaminants identified them as *Delftia* in the Ultrafast, *E.coli* in the Repliphi, and a *Cytophage* in the Genomiphi. Larger data sets with the NEB phi29 has shown that it has a minimal amount of *E.coli* contamination



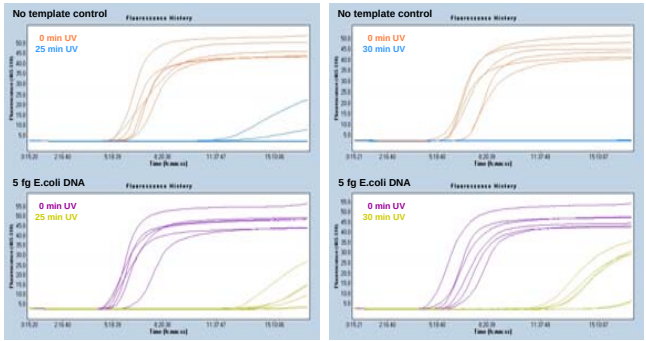
16S PCR products run on ~50ng of MDA DNA. Commercial kits show bacterial contamination. 1 µl of product on a 1% agarose gel, EtBr staining, 120V, 45 minutes

Despite the lowest level of contamination being found within the NEB kit, the Repliphi kit was chosen as the primary kit with which to continue our studies because of cost analysis. To suit our needs for a current project, further optimization was applied to 50 µl reactions instead of 20 µl.

Eliminating MDA reagent contamination using UV

Real time MDA was used to monitor the effect of UV treatment on the kinetics of the MDA reaction, in addition to cross linking of contamination. 50 µl MDA reactions were created from the Epicentre Repliphi kit according to the manufacturer's recommended concentrations, and were contaminated with 25 fg of *B. subtilis* DNA per reaction. These were UV treated in a Stratalinker 2400 (254 nm, ~4000 µwatts/sq cm) for up to 60 minutes at different time points. All UV exposures were performed in 96 well plate format in chilled, sterile MilliQ water in order to keep the temperature of the enzyme below 30C. 5 fg of *E.coli* DNA was added to each reaction as template and were run in real time using SYTO13 at 0.5 µM on a Roche LightCycler 480 for 17 hours at 30°C. Preliminary data (not pictured) suggested an optimal UV treatment in the 20-30 minute range. UV exposures of longer than 30 minutes showed a significantly reduced ability of the kit to amplify product.

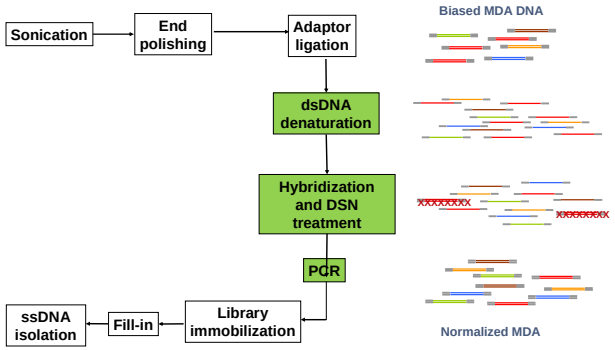
To show a comparison between a 25 and 30 minute UV exposure, 6X replicates of 50 µl MDA reactions were contaminated and UV treated at 0, 25 and 30 minute exposures. Template (5 fg E.coli DNA) and "no template" negative controls were run in parallel to gauge the efficacy of eliminating contamination within the kit, for a total of 48 reactions.



50 µl MDA reactions run in real time on a Roche LightCycler 480 with SYTO13 for 17 hours at 30C.

The 30 minute UV treatment is shown to be the most effective at eliminating contamination within the MDA reagents, as compared to the 25 minute treatment, which remained clean in only 4 out of 6 reactions (It is also possible this amplification could be a non-template product, i.e hexamer buildup. Further 16S analysis is required to determine if this is truly contamination). The rtMDA also demonstrated that the 30 minute UV treated reactions still had enough activity to amplify 5 fg of DNA, comparable to a single copy of a bacterial genome.

Normalized 454 Titanium Std library construction



The normalization steps, shown in green, were performed within the regular 454 Titanium library construction protocol. Because the process is also relatively quick and requires no specialized equipment, creation of normalized libraries is amenable for high throughput library construction, potentially in 96-well plate format.

Normalized 454 Assemblies

(A) Assembly comparisons

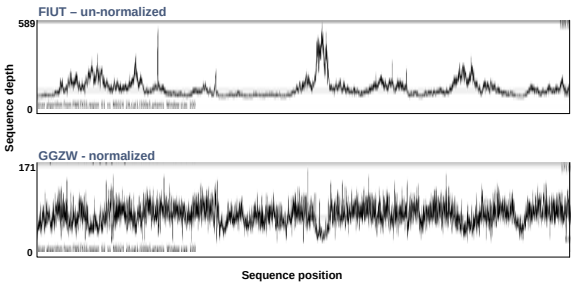
Assembly statistics of normalized and un-normalized 454 Titanium libraries

	un-normalized	*normalized*
Flavobacteria MS024-2A	**FIUT**	**GGZW**
Large contig statistics		
Number of contigs	223	104
Number of bases (Mbp)	1.95	2.07
Average contig size (bp)	8757	19912
N50 contig size (bp)	43951	61872
Largest contig (bp)	111995	146149
All contig statistics		
Number of contigs	348	135
Number of bases (Mbp)	1.99	2.08

A smaller number of large contigs were generated by the normalized library but, on average, they were over twice as long and an additional 0.12 Mbp was assembled, as compared to the original un-normalized library.

(B) Genome Coverage

Flavobacteria MS024-2A



MDA bias as evaluated by sequence depth distribution. Normalization results in significant improvements in the uniformity of coverage. The un-normalized library was sequenced to an average depth of ~45X with 454 Titanium, but the representation of specific regions of the genome ranged from 0 to >580X, with a mean coverage of 47X. After normalization, the difference in depth of coverage between the two libraries was reduced approximately 3.5 fold. Highly represented regions of the genome were drastically reduced, while under-represented regions saw an increase in depth of coverage.

Conclusions

- **The inherent contamination found within commercial MDA kits can be successfully eliminated with a 30 minute exposure of reagents to UV light.**
- **MDA produces bias in amplification of target genomes, leaving some regions of the genome highly overrepresented and others without sufficient coverage. Normalization of 454 Std Titanium libraries reduces this large bias and produces a more even depth of coverage across the genome, increasing sequencing efficiency and ease of genome assembly.**

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.