SAND2016-5702C

Parallel Uniform Sequencing of Microbial Aggregates by Barcode Droplet MDA

Yooli K. Light[1], Minsoung Rhee[3], Adam P. Arkin[2], Anup K. Singh[1], and Robert J.Meagher[1]
[1]Sandia National Laboratory, Livermore CA; [2]Lawrence Berkeley National Laboratory, Berkeley CA; [3]Illumina, Inc., San Diego, CA

ENIGMA
Ecosystems and Networks
Integrated with Genes and Molecular Assemblies

Sandia National Laboratories

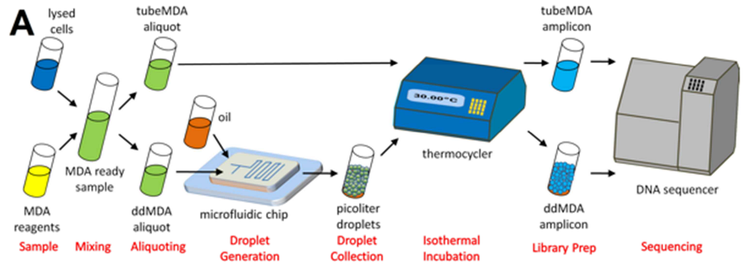
http://enigma.lbl.gov

Abstract

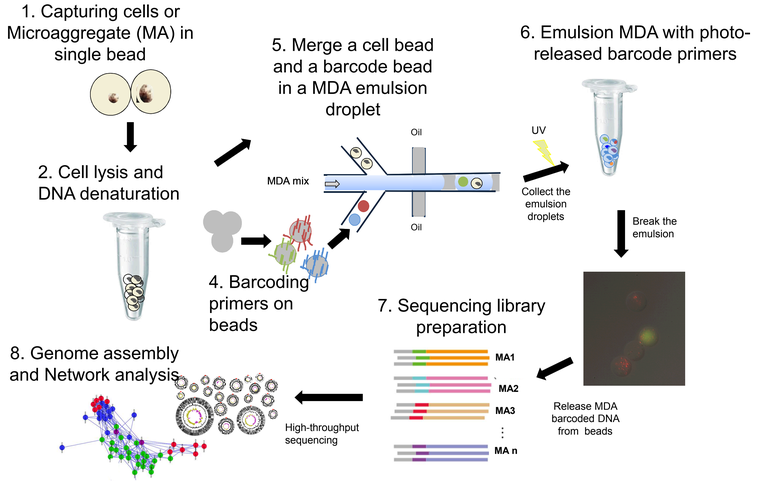
Microbial aggregates naturally occurring in an environmental niche are representative of direct microbial interactions, thus understanding this scale of interactions allows us to predict relationships between species and environmental traits with high resolution. To be able to study microbial aggregates at the micron scale, genomic amplification prior to sequencing is required, owing to limited amount of genomic content from microbial aggregates. Multiple displacement amplification (MDA) is robust but it has shown amplification bias that results in uneven sequence coverage. Here, we show droplet MDA achieves more uniform genomic amplification compared to conventional tube based method and a strategy of high-throughput sequencing library preparation to process large number of microbial aggregates in parallel using barcoded MDA droplets.

Method

1. Comparing droplet MDA to conventional tube MDA using *E.coli* genomic *DNA*

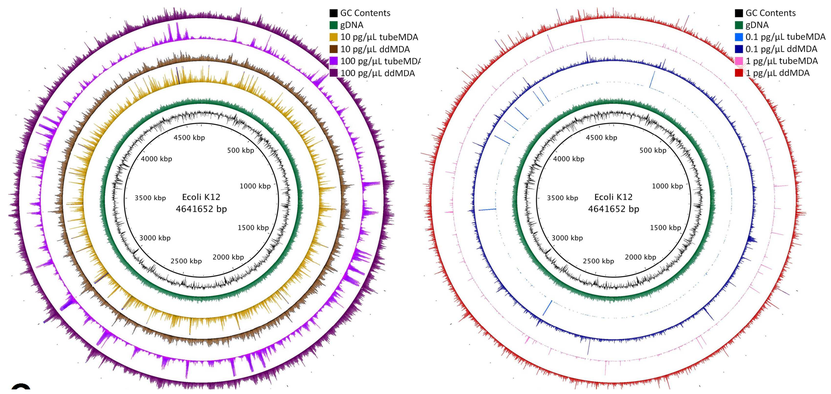


2. Developing a platform for parallel uniform sequencing for microbial aggregates by droplet MDA

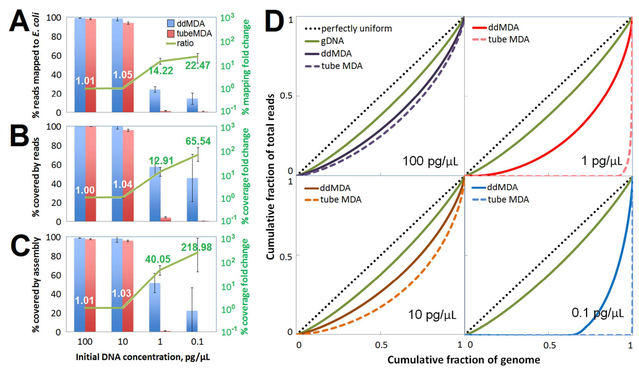


Results

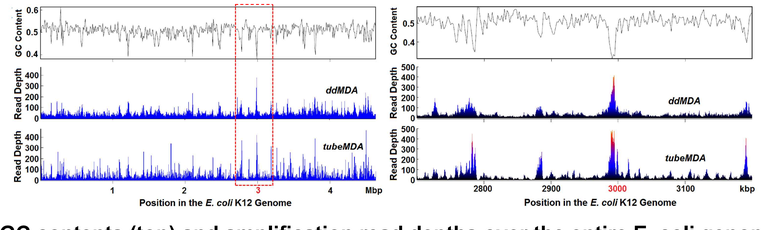
Improved genome coverage by droplet MDA



Genome coverage comparison of droplet MDA and tube MDA with various amounts of initial input DNA. Droplet MDA improved quality of amplification compared to tube MDA at the same concentrations, and more dramatically at low input DNA amounts. The sequencing output reads were filtered and trimmed and mapped to E.coli K-12 MG1655 reference genome using Bowtie2 with default settings in 'end-to-end' alignment mode. The mapping coverage was visualized by BRIG0.95



Sequence coverage comparison of droplet MDA and conventional tube MDA at different input DNA concentration, sequence reads mapped to *E. coli* K12 MG1655 reference genome (A) and genome coverage by raw reads (B) and *de novo* assembled contigs (C). The *de novo* assembly of sequences was performed using SPAdes 3.5 genome assembler with 'single-cell', 'paired-end libraries' and 'careful' options and was evaluated using QUAST. Gini indices were calculated based on the cumulative distribution of sequencing reads.



GC contents (top) and amplification read depths over the entire E. coli genome for ddMDA (middle) and tube MDA (bottom) at the DNA concentration of 10 pg/µL. The right panels show zoomed-in plots of the dotted-line box region of the genome for close-up visualization.

MDA Barcode primer synthesis and comparison of droplet MDA and tube MDA on mixed bacteria population

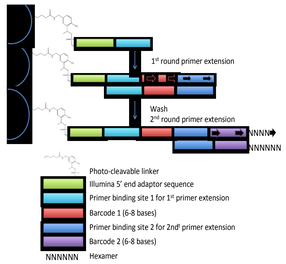
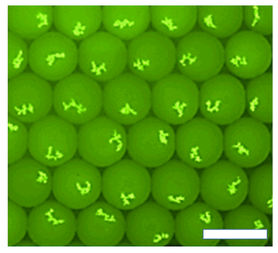


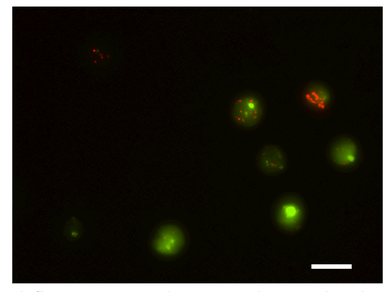
Diagram of barcode primer synthesis on microspheres by primer extension

	Target	*E.coli*	*D.vulgaris*	*C. actobutylicum*	mapped total
ddMDA	mixed	5.86	12.91	68.14	86.91
	E.C only	83.35	0.45	0.4	84.2
	D.V only	0.39	77.7	0.26	78.35
	C.A only	0.3	0.21	96.18	96.69
tMDA	mixed 2	13.8	0.63	44.17	58.6
	E.C only	85.86	0.43	0.43	86.72
	D.V only	0.49	79	0.39	79.88
	C.A only	0.35	0.24	79.06	79.65
posCont	E.C	97.3	0.37	0.31	97.3
	D.V	0.44	99.81	0.27	99.81
	C.A	0.54	0.43	99.89	99.89

Comparison of droplet and tube MDA on mixed bacteria population *E.coli (50%GC)*: *D.vugaris(62%GC)*: *C. actobutylicum (30%GC)* = 8: 1: 1. The sequencing output reads were filtered and trimmed and mapped to each reference genome using Bowtie2 with default settings in 'end-to-end' alignment mode



A fluorescence microscope image showing droplet MDA endpoint with the initial template DNA concentration of 100 pg/µL. The genomic DNA was heat denatured and mixed in with the MDA reaction solution and oil and incubated for18hrs at 30 ºC. scale bar =100 um



A fluorescence microscope image showing MDA endpoint with the cells captured in agarose gel beads. The cell beads were treated with alkaline lysis solution followed by neutralization, merged with MDA reaction mixture and incubated for 18hrs at 30 ºC. scale bar =100 um

Conclusion

- **The droplet MDA technique enabled significantly lower bias and non-specific amplification than conventional MDA thus achieving more uniform coverage of amplification over the entire genome**
- **This technique can be a powerful tool for genomic studies where DNA samples are limited such as single cells, microaggregates, and uncultured microbes from many different environments**
- **Indexing of each microbial aggregate together with unbiased amplification utilizing droplet MDA enables to sequence microbial aggregates with high-throughput and to provide improved sequence coverage of aggregate members.**

ACKNOWLEDGEMENTS
This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://enigma.lbl.gov), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231
Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

•Initially, we used *Ecoli* K-12 MG1655 to perform droplet MDA and compare the evenness of sequencing coverage to that of conventional tube MDA. *E.coli* genomic DNA was denatured and fragmented after heat lysis of cells and each fragmented DNA was partitioned in an emulsion droplet containing MDA reaction mixture and hexamer primer. MDA DNA was pulled together and sequenced. The coverage of genomic DNA was determined by mapping contigs to the reference sequence after *de novo* assembly. For microbial aggregate study, hexamer primer with specific barcode was assigned to each aggregate to handle many aggregates in parallel and whole genome assemblies of members of each microbial aggregate and network analysis of microbial aggregates were performed from the multiflexed sequence reads. For a sample containing 0.1 pg/μL of *E. coli* DNA (equivalent of ~3/1000 of an *E. coli* genome per droplet), droplet MDA achieved a 65-fold increase in coverage in de novo assembly compared to conventional tube MDA.

•

•Indexing of each microbial aggregate together with unbiased amplification utilizing droplet MDA enables us to sequence microbial aggregates with single

ACKNOWLEDGEMENTS

This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://enigma.lbl.gov), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231