HIGH THROUGHPUT MICROFLUIDIC SAMPLE PREPARATION FOR METAGENOMIC ANALYSIS

Feiqiao Yu*, Mark A. Horowitz, and Stephen R, Quake
Stanford University, USA

ABSTRACT

Metagenomics deal with analysis of genetic material from environmental samples. The ultimate goal is to reconstruct entire genomes of unknown microbial species. We present a sample preparation method using the Fluidigm© C1 microfluidic platform that integrates throughput of shotgun sequencing with bioinformatics simplicity of single cell microfluidics. Our method is tested with a benchmark sample and characterize its mapping quality and coverage uniformity. We conclude that our microfluidic based pipeline is suitable for metagenomic applications.

KEYWORDS: Metagenomics, C1 Microfluidics, MDA, Sequencing

INTRODUCTION

Advancement in sequencing technologies and bioinformatics methods have enabled the study of complex microbial samples. Through collective efforts, scientists have uncovered astonishing microbial diversity across communities [1]. Such biodiversity may change across temperature or nutrient gradients [2] and may even affect the health of the host [1]. Even the same species from different locations could display different genetic makeups due to mutation or horizontal gene transfer [3].

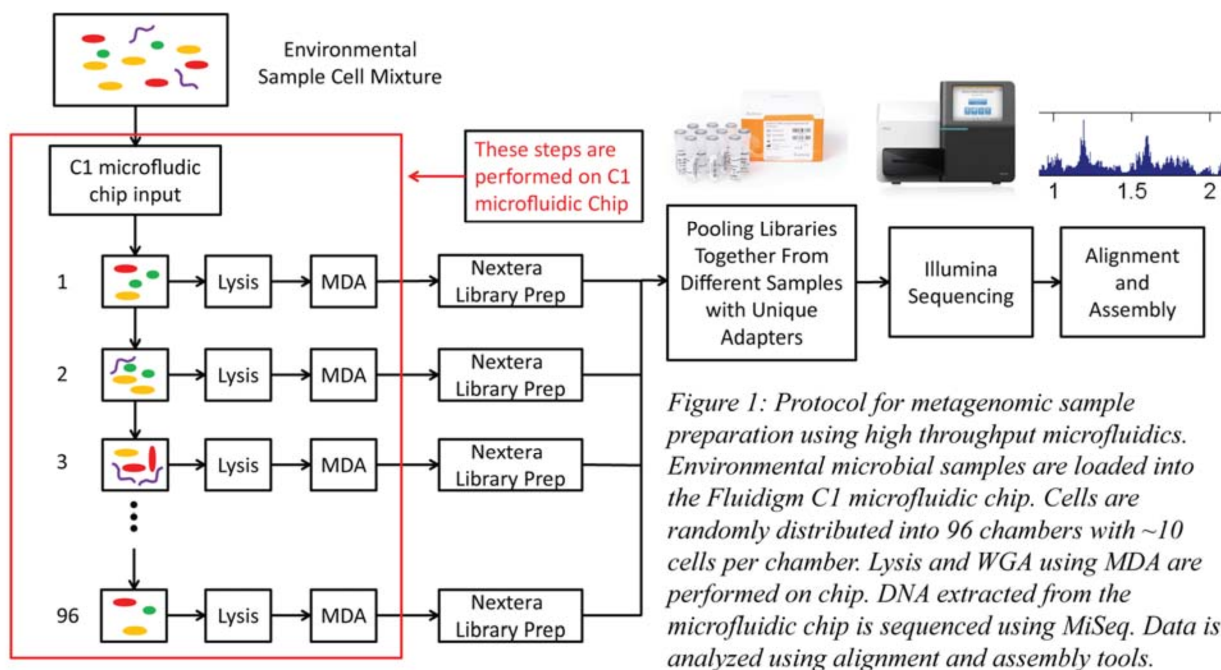

Figure 1: Protocol for metagenomic sample preparation using high throughput microfluidics. Environmental microbial samples are loaded into the Fluidigm C1 microfluidic chip. Cells are randomly distributed into 96 chambers with ~10 cells per chamber. Lysis and WGA using MDA are performed on chip. DNA extracted from the microfluidic chip is sequenced using MiSeq. Data is analyzed using alignment and assembly tools.

Investigating complex microbial communities is challenging. Most bacterial species from these communities are unculturable in laboratory conditions. The large number of different species and their highly variable abundance adds to the difficulty of investigating the genetic makeup of environmental samples. Traditional shotgun metagenomic approach sequences DNA extracted from large sample volumes containing hundreds of cells. Even though such an experimental process is simple, determining which sequencing reads originated from which cell is a complex if not impossible task [4]. Furthermore, most sequencing reads are derived from abundant species, making detection of rare species difficult. Another approach isolates single cells in plates or microfluidic chambers using FACS [5] or an optical

tweezers [6]. These single cell methods guarantee that all sequencing reads originate from the same cell but suffer from low throughput, which also result in difficulty to detect rare species.

METHODS

We present a sample preparation method using a microfluidic platform that performs high throughput metagenomic analysis while significantly reducing the complexity required by bioinformatics analysis. The microfluidic platform described here takes one mixed microbial population, generates 96 independent samples on-chip with a specified concentration of cells dictated by the initial dilution. In this case, we use ~10 cells per sample. Lysis and WGA (Whole Genome Amplification) are performed on-chip. WGA is performed using MDA (Multiple Displacement Amplification). Finally, amplified DNA product is harvested from the chip for downstream sequencing and bioinformatics analysis (Fig. 1).

More specifically, we modify a Fluidigm© C1 microfluidic platform. The C1 microfluidic chip contains 96 replicate reaction sequences in parallel. Each reaction sequence contains 5 chambers where reagents can be added sequentially. For our application, we use the first reaction chamber for lysis, the second and third reaction chambers for neutralization, and the final two reaction chambers for MDA. Since each replicate contains on the order of 10 cells, multiplexing in 96 independent reactions ensures that ~1000 cells are analyzed, increasing the discovery probability of rare microbial species.

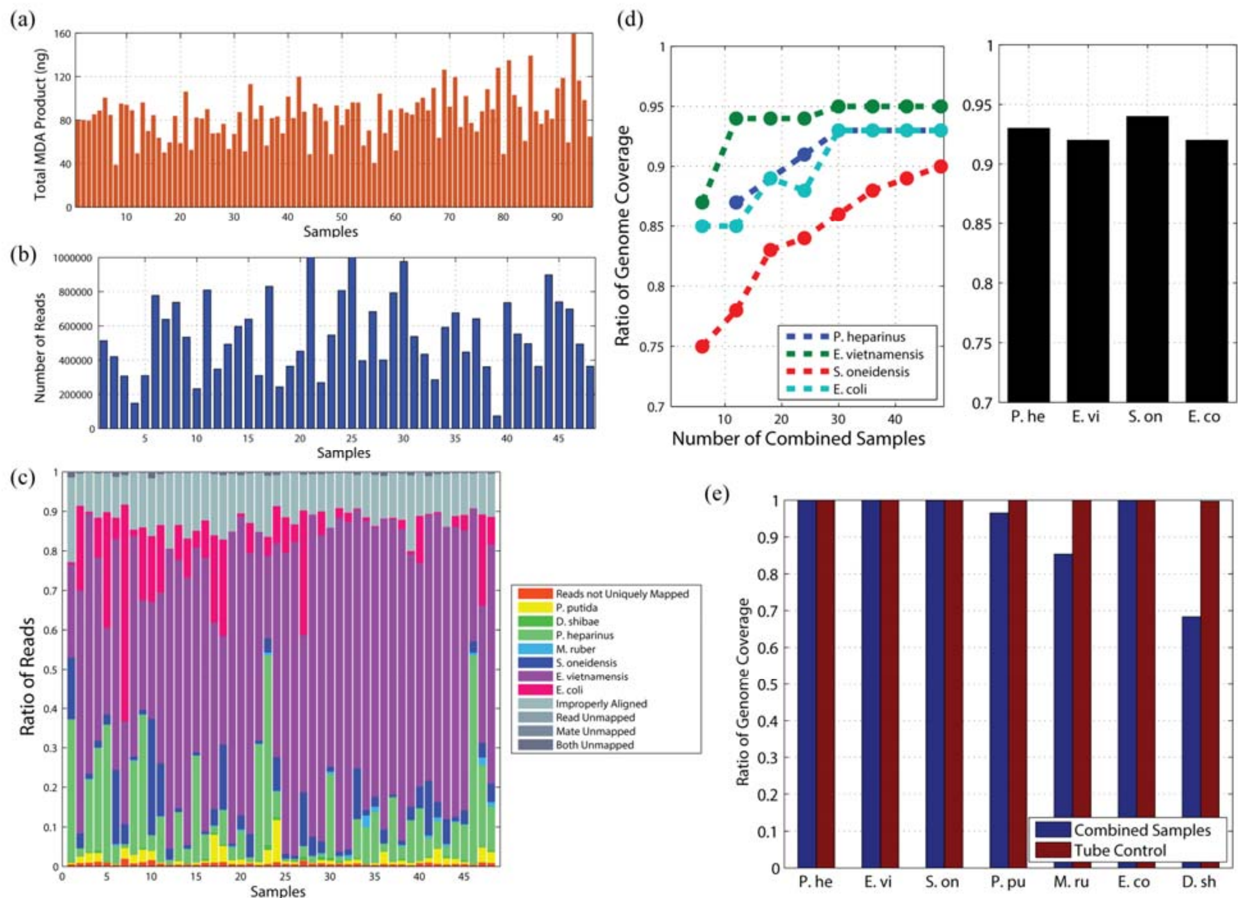

Figure 2: (a) Total MDA product from a C1 microfluidic chip. (b) Sequencing reads obtain for the 48 sample libraries. (c) Most of the reads map uniquely to one of the seven species, with low contaminants and few chimeras. (d) MDA amplification bias can be reduced by pooling reads from multiple samples, yielding a more uniform coverage. Left panel represents pooled reads down-sampled to 5X depth. As more samples are pooled, coverage at 5X depth increases and approaches tube controls coverage at 5X depth (shown in the right panel). (e) Reads pooled from all samples cover the entire genome. This demonstrates that, provided enough sequencing depth for each species, MDA amplified DNA contain information of the entire genome.

RESULTS AND DISCUSSION

When analyzing small number of bacterial cells, DNA quantity is the limiting factor. MDA is the most common method to amplify unknown DNA materials using random primers and *phi29* polymerase. However, MDA creates spatially biased amplification profiles whose patterns appear to be random. Research have demonstrated that MDA bias may be mitigated by reducing the total amount of amplification through limiting reaction volume [7]. Nanoliter volumes of a microfluidic reaction chamber, therefore, represent the best mechanism. We tested our metagenomic sample preparation method using a mixture of known, gram negative species. From eight hour MDA reactions, we obtained 85 ng of DNA on average from 300 nL reactions (Fig. 2a). Due to considerations for sequencing depth, only 48 samples are sequenced. The number of reads obtained per sample is illustrated in Fig. 2b.

Mapping sequencing reads to known genomes, we observe that most reads uniquely align to one of the seven species (Fig. 2c). Due to the closed environment offered by microfluidic based amplification, there is very little contaminants, resulting in less than 2% unaligned reads and 8% improperly aligned reads. Improperly aligned reads are mostly composed of short libraries and chimeras. All species included in the original mix are represented in at least one sample. Bias observed is most likely due to preferential lysis and MDA amplification. Since MDA bias seems random, we test if pooling reads belonging to the same species from multiple samples will result in more uniform genome coverage. Using four most represented species from the sequencing run, we pool different number of samples together and observed an increase in ratio of genome coverage after randomly down-sampling to 5X sequencing depth (Fig. 2d). We notice further that genome coverage saturate at tube control levels, confirming our earlier hypothesis that, for rare species, pooling reads from different samples generated by the C1 microfluidic chip increase coverage uniformity. Finally, pooling reads from all 48 samples results in the complete coverage of bacterial genomes for the four most represented species (Fig. 2e). This result demonstrate that, with sufficient sequencing, MDA amplified DNA contain information of the entire genome.

The advantages of our microfluidic metagenome approach include automation, throughput, and reduced bioinformatics complexity. Small reaction volumes not only reduce reagent cost but also decrease MDA bias. In addition, large numbers of analyzed cells per run provide higher probability for the detection of rare bacterial species. Taken together, the microfluidic sample preparation method is ideal for metagenomic applications.

REFERENCES

[1] Human Microbiome Project Consortium, "Structure, function and diversity of the healthy human microbiome," *Nature*, 486, 207-214, 2012.

[2] R.C. Everroad, H. Otaki, K. Matsuura, and S. Haruta, "Diversification of bacterial community composition along a termperature gradient at a thermal spring," *Microbes Enviorn.*, 27, 4, 374-381, 2012.

[3] D. Bhaya et al, "Population level functional diversity in a microbial community revealed by compariative genomic and metagenomic analyses," *The ISME Journal*, 1, 703-713, 2007.

[4] V. Iverson et al, "Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota," *Science,* 335, 587-590, 2012.

[5] C. Rinke et al, "Insights into the phylogeny and coding potential of microbial dark matter," *Nature*, 499, 431-437, 2013.

[6] I.P. Marshall et al, "A single-cell genome for Thiovulum sp," *Appl. Environ Microbiol*, 78, 8555-8563, 2012.

[7] J. Gole et al, "Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells," *Nat. Biotech*., 31, 1126-1132, 2013.

CONTACT

* Feiqiao Yu; phone: +1-626-484-2321; brianyu@stanford.edu