prepIT®•MAX

Comparison of extraction methods for whole genome sequencing of *Mycobacterium tuberculosis*

Jacques Niles, Christina Dillane, Mike Tayeb, Pat Curry and Cassandra Kelly-Cirino
DNA Genotek, Ottawa, Ontario, Canada

Introduction

With the global rise of multi-, extensive, and total drug resistance to *Mycobacterium tuberculosis* (MTb), it is crucial to have a technique that can detect resistance markers quickly and accurately[1]. The traditional method of bacterial culture testing is slowly making way for molecular-based tests, such as Cepheid® GeneXpert®, line probe assays, targeted sequencing, and whole genome sequencing, which can identify and characterize all resistance mutations simultaneously[2]. These techniques demand quick and simple methods of DNA extraction that reliably yield large amounts of DNA from samples. The standard methods of DNA extraction from MTb cultures, such as bead beating and sonication, can be time-consuming, provide inconsistent DNA quality and can produce hazardous aerosols. prepIT®•MAX is a fast and simple extraction method that maximizes the MTb DNA yield from cultures without the need for potentially hazardous bead beating or sonication. The purpose of this study was to compare the quality and sequencing performance of DNA extracted using the prepIT•MAX extraction kit versus DNA extracted using InstaGene™Matrix (Bio-Rad) and a bead beating-based method.

Methods

DNA isolation: DNA was obtained from attenuated *Mycobacterium tuberculosis* (aMTb) cultures (10^8 CFU/mL in phosphate-buffered saline [PBS]) using three different extraction methods (n=5 per method). In the first method, DNA was extracted using the prepIT•MAX protocol[3]. In the second method, the samples were centrifuged, the PBS was removed, and the mycobacterial pellet was resuspended in 200 µL of InstaGene Matrix and was prepared according to the manufacturer's bacteria protocol[4]. In the third method, samples were processed using a bead beating extraction protocol, which consisted of incubating cultures at 80°C for 1 hour, mixing with 150 µL of glass beads (105-150 µM) and then placing the mixture in a bead beater for 1 minute followed by 1 minute on ice. The resulting DNA was quantified using the PicoGreen® technique (Thermo Scientific) and then quality-tested by TaqMan® real-time PCR assay and by running the products on an agarose gel.

DNA fragmentation, library preparation and sequencing: DNA was mechanically fragmented using a Covaris S220 focused-ultrasonicator, targeting a peak fragment size of 400 bp (Duty Factor = 10%, Peak Incident Power = 105 W, Cycles per Burst = 200, Time = 55 s). Sequencing libraries for the prepIT•MAX- and bead beating-extracted DNA were prepared using 10 ng of the fragmented DNA. Libraries for the InstaGene-extracted DNA were prepared using the maximum volume (32 µL) of fragmented DNA due to low yields. All libraries were prepared using the NEXTflex™ Rapid-DNA Seq Kit with NEXTflex ChIP-seq Barcodes (BIOO Scientific). Briefly, DNA fragments were end-repaired and adenylated, adapters were added by ligation, and product was size-selected using AMpure XP beads (Beckman Coulter Genomics). The libraries were then amplified via PCR (12 cycles for the prepIT•MAX- and bead beating-extracted DNA; 14 cycles for the InstaGene-extracted DNA due to low yields) and a final bead-cleanup was performed. Libraries were qualified on the Bioanalyzer (Agilent Technologies) and quantified by PicoGreen. All individual libraries were normalized to the same nanomolar amount and the pool was quantified by qPCR using the NEBNext® Library Quant Kit for Illumina® (New England Biolabs) before being sequenced on the Illumina MiSeq (2×250 bp paired-end).

PD-WP-00049 Issue 1/2016-04
© 2016 DNA Genotek Inc., a subsidiary of OraSure Technologies, Inc., all rights reserved.

Bioinformatics: Adapter trimming was performed by the MiSeq control software (v2.5.0.5) on-instrument. Bases with a Q-score less than 30 were trimmed from the reads using Sickle[5] (v1.21). Trimmed reads were aligned to the *Mycobacterium tuberculosis* H37Ra reference genome (Genbank Accession GCA_000016145.1) using BWA[6] (v0.5.9-r16). Sequencing and alignment summary metrics were generated using FastQC[7] (v0.10.1), Picard Tools[8] (v1.134) and Qualimap[9] (v2.1.1).

Results

The prepIT•MAX method generated a significantly higher yield of total DNA than the InstaGene and bead beating methods (both $p<0.05$) (Table 1). As well, the prepIT•MAX-extracted DNA was of higher molecular weight than that extracted with InstaGene or bead beating (Figure 1). In addition, Ct values below 35 cycles from the TaqMan real-time PCR assay demonstrated DNA functionality for each extraction (data not shown).

Method (n=5 for each)	Average total yield (ng)
prepIT•MAX	116.68 + 21.70
InstaGene	27.33 + 14.06
Bead beating	64.94 + 6.74

Table 1*: Average total yields for the three methods.*

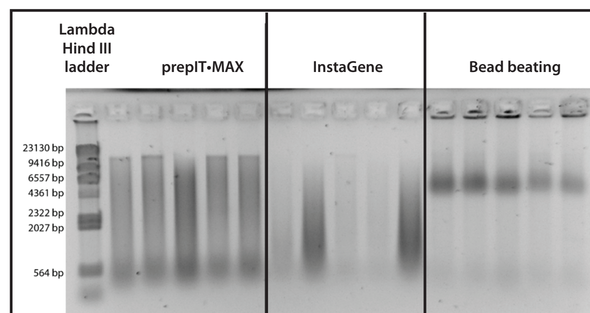

Figure 1*: DNA from the three extraction methods run on a 0.8% agarose gel.*

Bioanalyzer traces of the prepared libraries demonstrated the effectiveness of the Covaris shearing with the prepIT•MAX prepared DNA. The libraries prepared from the InstaGene- and bead beating-extracted DNA showed more off-target-size fragments compared to prepIT•MAX DNA. The InstaGene samples also had smaller average fragment size than was targeted (300 bp), whereas the bead beating samples had larger average fragment size than was targeted (600 bp). The prepIT•MAX samples showed a good distribution of fragment lengths, with average size approximately 450 bp (Figure 2). The Bioanalyzer traces also reveal that the InstaGene libraries had concentrations two- to three-fold lower than both the prepIT•MAX and bead beating libraries despite using an extra two PCR cycles during library preparation.

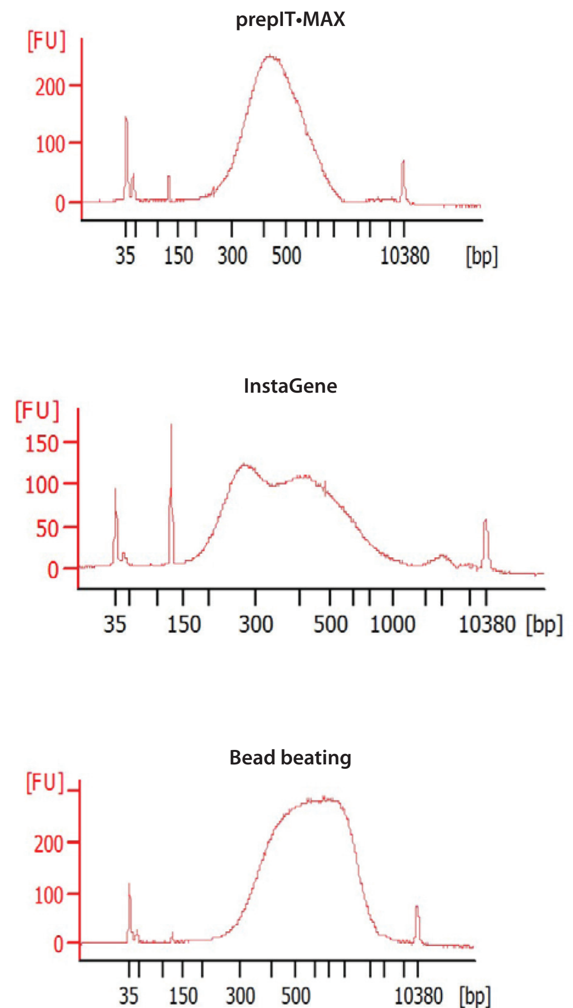

Figure 2*: Representative Bioanalyzer traces for each of the three methods.*

© 2016 DNA Genotek Inc., a subsidiary of OraSure Technologies, Inc., all rights reserved.

Despite the fact that all samples were library-prepped together and equivalent nM amounts of each sample were loaded on the MiSeq flow cell, the prepIT•MAX-extracted samples had, on average, four-fold more reads sequenced than the InstaGene-extracted samples. The number of reads for the bead beating-extracted samples was similar to that for the prepIT•MAX samples (Table 2). With the prepIT•MAX- and bead beating-extracted samples, 99.0% of the reads mapped to the reference genome (aMTb), whereas the average for the InstaGene-extracted samples was 89.7%. Genome coverage was correlated with the number of reads obtained for each sample. Samples extracted using prepIT•MAX and bead beating had greater than 90× coverage, on average, throughout the genome, whereas the corresponding value for the InstaGene was 20× (Table 2). Examination of coverage for three drug resistance genes (*rpoB*, *inhA* and *katG*) revealed similar depths of coverage for the prepIT•MAX- and bead beating-extracted samples (approximately 100×), whereas there was less coverage for the InstaGene-extracted samples (approximately 20×) (Table 3).

Method (n=5 for each)	Number of reads	Mapped reads (%)	Coverage depth (X)
prepIT•MAX	2,084,632 + 89,987	99.0 + 0.1	90.0 + 3.8
InstaGene	524,436 + 320,287	89.7 + 7.0	20.8 + 14.6
Bead beating	2,246,128 + 219,076	99.2 + 0.1	98.5 + 9.7

***Table 2**: Average sequencing statistics for the three extraction methods.*

Method (n=5 for each)	Average coverage depth of resistance genes (X)		
	rpoB	*inhA*	*katG*
prepIT•MAX	108.0 + 10.1	100.8 + 7.6	101.9 + 6.8
InstaGene	20.7 + 14.6	21.8 + 16.6	21.5 + 13.8
Bead beating	110.2 + 12.7	111.5 + 9.2	109.2 + 8.1

***Table 3**: Average coverage of three resistance markers (rpoB, inhA and katG) for each of the three extraction methods.*

Discussion

The data demonstrate that DNA obtained using the prepIT•MAX extraction kit is suitable for whole genome sequencing, and is compatible with a commonly used method of DNA fragmentation (sonic shearing using a Covaris instrument) and a ligation-based library preparation method. Findings for number of reads and percentage mapped to the reference genome were dependent on the DNA extraction method used. Coverage beyond that suggested for drug resistance identification (approximately 20× to 30×) was achieved in this study, indicating that additional samples could be included in the sample pool on the MiSeq[10].

Number of reads, coverage depth of the whole genome, and coverage of the individual drug resistance genes evaluated were greater for samples extracted using prepIT•MAX than for those extracted using the InstaGene method. This could be attributed to the low yield and greater extent of degradation observed in the InstaGene samples. It is likely that the increased DNA degradation contributed to the ineffective library prep, as indicated by the Bioanalyzer results (Figure 2) and library concentrations.

Bead beating and prepIT•MAX extractions yielded similar numbers of sequencing reads and coverage depth. However, the prepIT•MAX extraction method is much safer for laboratory technicians than the bead beating method, as prepIT•MAX extraction does not generate hazardous aerosols. DNA Genotek also obtained higher molecular weight DNA and better yields with the prepIT•MAX kit than with conventional bead beating.

It is also important to note that total DNA yield, total number of reads, and coverage depth are reproducible when samples are extracted with the prepIT•MAX method. In contrast, the DNA yield and total number of reads are inconsistent for the samples extracted with InstaGene and bead beating.

Conclusion

The results indicate that the prepIT•MAX extraction kit is a simple, fast and safe method of consistently extracting large amounts of high quality DNA that performs equal to or better than DNA extracted by other methods commonly used in whole genome sequencing.

© 2016 DNA Genotek Inc., a subsidiary of OraSure Technologies, Inc., all rights reserved.

References

[1] Brown AC, Bryant JM, Einer-Jensen K, et al. Rapid whole-genome sequencing of *Mycobacterium tuberculosis* isolates directly from clinical samples. *J Clin Microbiol 2015*; 53: 2230–37.

[2] Takiff HE, Feo O. Clinical value of whole-genome sequencing of *Mycobacterium tuberculosis*. *Lancet Infect Dis. 2015*;15(9):1077–90.

[3] prepIT®•MAX for TB. Laboratory protocol for manual purification of DNA from sputum. DNA Genotek Inc., PD-PR-00497

[4] InstaGene™Matrix. DNA Preparation From Bacteria. *Bio-Rad*. LIT544 Rev G

[5] Joshi NA, Fass JN. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle.

[6] Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, 25:1754-60. [PMID:19451168]

[7] FastQC: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

[8] Picard: https://github.com/broadinstitute/picard

[9] Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2015). "Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data." *Bioinformatics*, btv566

[10] Cuomo, A.M., Zucker, H.A., Dreslin, S. Validation of Next Generation Sequencing (NGS)- Based Methods for Identification and/or Characterization of Infectious Agents (isolates only). New York State Department of Health.

Some DNA Genotek products may not be available in all geographic regions.
®OMNIgene and prepIT are registered trademarks of DNA Genotek Inc. All other brands and names contained herein are the property of their respective owners.
All DNA Genotek protocols, white papers and application notes, are available in the support section of our website at www.dnagenotek.com.

© 2016 DNA Genotek Inc., a subsidiary of OraSure Technologies, Inc., all rights reserved.