Strain-Typing and Antibiotic Resistance Profiling From Research Samples Using Highly Multiplexed Targeted Library Construction with High Throughput Semi-Conductor Based Sequencing.

John Bishop, Loni Pickle, Miro Dudas, Guoying Liu, Melanie Baumann, and Astrid Ferlinz

Thermo Fisher Scientific, Carlsbad, CA, USA

ion torrent by Thermo Fisher Scientific

Overview

We describe here a fast, sensitive, highly multiplexed targeted amplification method that can genotype up to thousands of specific loci in microbial genomes from heterogeneous samples. We used this method with semi-conductor based sequence to identify *Mycobacterium* from cultures and sputum samples, and further to genotype specific loci associated with antibiotic resistance in *M. tuberculosis*.

Introduction

Recent advances in high throughput sequencing have enabled sequencing of microbial genomes. However, pure microbial isolates are often not available for sequencing, and furthermore, full genome sequencing often is not necessary, such as when only specific regions of the genome are of interest. For example, identifying strains of *Mycobacterium tuberculosis* with specific antibiotic resistance genotypes has become increasingly important in combating multi- and extensively- drug resistant (MDR and XDR) strains. There are many known and suspected antibiotic resistance loci in *M. tuberculosis* , but *M. tuberculosis* grows extremely slowly in culture, taking weeks to provide complete antibiotic-sensitivity information based on phenotypic assays or on whole genome sequencing. We therefore developed a highly multiplexed amplification-based method to interrogate the major antibiotic resistance loci in *M. tuberculosis* starting with either cultures or sputum samples. This method is scalable to thousands of individual loci and can be generalized to any organism.

Methods

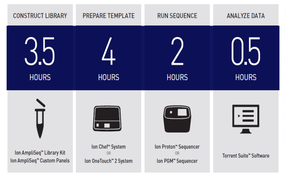


Figure 1. Workflow for sequencing of multiplexed amplification libraries:

Using AmpliSeq.com we designed a multiplex panel tiling across the full coding region of eight known antibiotic-resistance loci in *M. tuberculosis* (see table below). Libraries were created from cultures and sputum samples using this panel with the Ion AmpliSeq™ Library reagents. Libraries were sequenced on the Ion Torrent PGM Sequencer, and analyzed using the Torrent Server software and custom scripts

Table 1. Genes covered in the Ion AmpliSeq™ panel. The full protein coding regions and some upstream and downstream bases are covered for each gene to maximize discovery of known and novel variants. The total number of targeted bases is scalable to 5 Mb.

Gene Target	Function	Major Drug Resistance	Size of Full Genes Covered (bp)	Number of Amplicons
embB	Membrane indolylacetylinositol arabinosyltransferase B	Ethambutol	3321	22
eis	Aminoglycoside acetyltransferase	Kanamycin	1335	10
gyrA	DNA gyrase subunit A	Fluoroquinolones	2552	17
inhA	NADH-dependent enoyl-[acyl-carrier-protein] reductase	Isoniazid	1010	7
katG	Catalase-peroxidase-peroxynitritase T	Isoniazid	2413	20
pncA	Pyrazinamidase/nicotinamidase	Pyrazinamide	708	5
rpoB	DNA-directed RNA polymerase β-chain	Rifampicin	3552	24
rpsL	30S ribosomal protein S12	Streptomycin	549	4
		Totals	15 440	109

Results

Figure 2. Limit of detection for *M. tuberculosis*.

To ascertain a limit of detection for *M. tuberculosis* (Mtb) we made libraries from H37rv gDNA titrated from 0 to 100 pg into a background of 10 ng DNA from sputum. High quality libraries were made from at least 0.1 to 1 pg of Mtb DNA (~50 to 500 genomic equivalents).

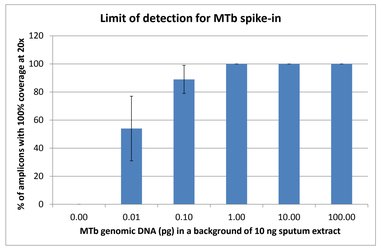


Figure 3. Uniformity of amplicon coverage

All 109 amplicons in the panel were produce at highly even levels from both cultures and sputum samples.

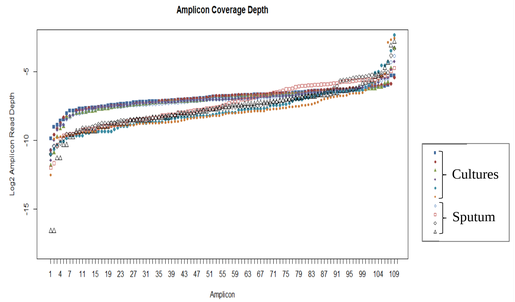
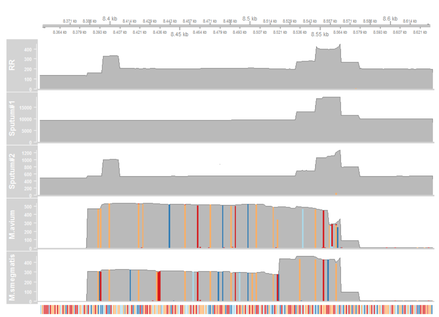


FIGURE 4. *M. avium* and *M. smegmatis* differentiation at sequence level: Non-tuberculosis *Mycobacterium* may be identified by the abundant variants present in the minority of targeted amplicons that amplify from these genomes. Sequence traces below are shown for (top to bottow) a rifampicin resistant Mtb strain, Mtb-containing sputum samples #1 and #2, *M. avium*, and *M. smegmatis*. One amplicon from gyrA is shown on the left, and one amplicon from inhA on the right.



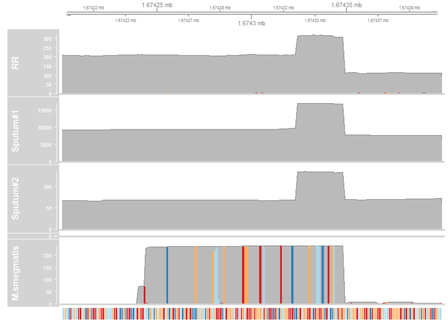
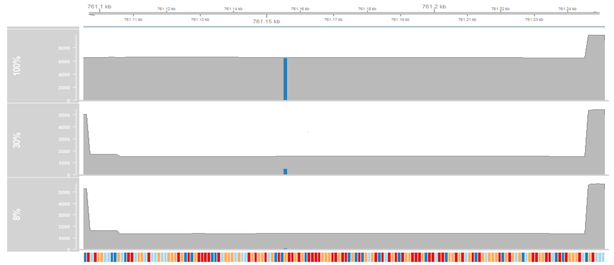


FIGURE 5. Minor variants can be called in mixed populations: We made libraries from 10 ng sputum mixed with 1 pg gDNA from a rifampicin-resistant strain containing a C→T variant in the rpoB gene at 100% (top trace), 30% (middle trace), or 8% (bottom trace) mixture with H37rv strain. The single variant is evident (blue line) and called at both 30% and 8% minor frequency.



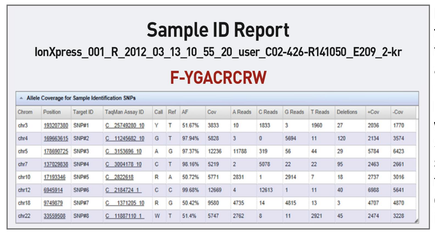


FIGURE 6. Multiple organisms can be targeted simultaneously. Panels can target up to ~5 Mb spread across thousands of individual amplicons, which can be targeted to loci in multiple different organisms in the same sample. For example, combining the SampleID panel with the Mtb panel ascertains sex and unique SNP profile for human DNA present in the sputum sample. Similarly, genotypic markers for host-microbe interaction, pharmaco-genomics, or microbial co-infection could be included in a single panel.

Conclusion

We've described a fast and sensitive highly multiplexed amplification method for producing sequencing libraries specific to targeted loci in heterogeneous microbial samples. We demonstrated this process with a panel specific to full gene coverage of eight known antibiotic resistance loci in *M. tuberculosis*, however this process is scalable and extensible to simultaneously assess thousands of loci across multiple organisms per sample.

For Research Use Only. Not intended for any animal or human therapeutic or diagnostic use