

(51) International Patent Classification:
C12Q 1/68 (2006.01)(21) International Application Number:
PCT/US2013/040302(22) International Filing Date:
9 May 2013 (09.05.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/644,876 9 May 2012 (09.05.2012) US
61/646,060 11 May 2012 (11.05.2012) US
61/695,960 31 August 2012 (31.08.2012) US
61/737,250 14 December 2012 (14.12.2012) US(71) Applicant: **LONGHORN VACCINES AND DIAGNOSTICS, LLC** [US/US]; 2 Bethesda Metro Center, Suite 910, Bethesda, MD 20814 (US).(72) Inventors: **DAUM, Luke, T.**; 318 Larkwood Drive, San Antonio, TX 78209 (US). **FISCHER, Gerald, W.**; 6417 Lybrook Drive, Bethesda, MD 20817 (US).(74) Agents: **SMITH, Matthew** et al.; Remenick Pllc, 1025 Thomas Jefferson St., Nw, Suite 175, Washington, DC 20007 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

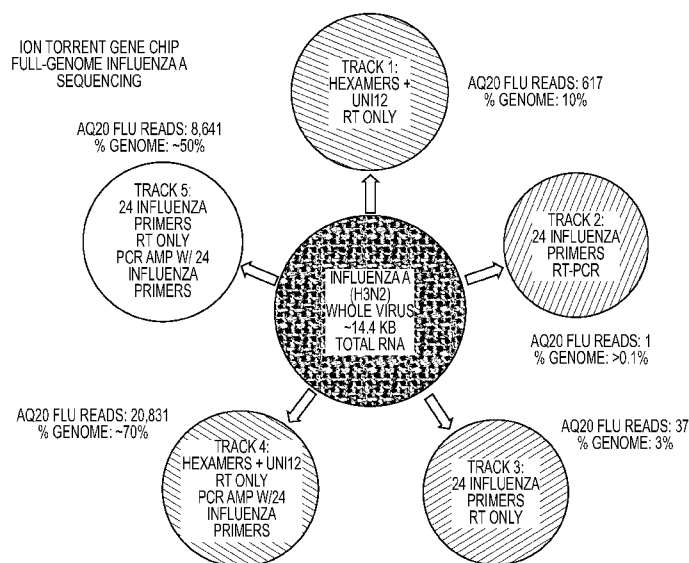
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

[Continued on next page]

(54) Title: ION TORRENT GENOMIC SEQUENCING

**FIG. 9**

(57) Abstract: Disclosed is an enhanced method for rapid and cost-effective analysis of sequences of a microorganism by semiconductor sequencing, preferably ion-torrent sequencing. This method provides for full length analysis and of multiple areas (e.g. genes) of multiple genomes. These methods identify genetic mutations of a particular gene that are responsible for conferring resistance or sensitivity to an antibiotic or other chemical compound. Multiple different species, strains and/or serotypes of a particular organism are rapidly and efficiently screened and mutations identified along with the complete genome of an organism. By selecting primers pairs of similar size and GC content that produce amplicons with sequences spanning the entire genome, a single PCR reaction analyzed by ion torrent methodology can determine the sequence of a complete genome. Methods are useful to sequences the genomes of viral agents, such as influenza virus, and bacterial agents, such as tuberculosis bacteria.



-
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))* — *with sequence listing part of description (Rule 5.2(a))*

ION TORRENT GENOMIC SEQUENCING

Reference to Related Applications

This application claims priority to U.S. Provisional Application No. 61/737,250 entitled "*Ion Torrent Genomic Sequencing*" filed December 14, 2012, U.S. Provisional Application No. 61/695,960 entitled "*Ion Torrent Genomic Sequencing*" filed August 31, 2012, U.S. Provisional Application No. 61/646,060 entitled "*Drug Susceptibility Determination by Ion Torrent Sequencing*" filed May 11, 2012, and U.S. Provisional Application No. 61/644,876 "*Drug Susceptibility Determination by Ion Torrent Sequencing*" filed May 9, 2012, and the entirety of each of which is specifically incorporated by reference.

Sequence Listing

The instant application contains a Sequence Listing which has been submitted in ASCII format via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on May 7, 2013, is named 3022.018.PCT_SL.txt and is 37,927 bytes in size.

Background

1. Field of the Invention

This invention is directed to tools, compositions and methods for identifying genetic mutation and mega-bases of nucleic acid information by semi-conductor sequencing and, in particular, to analyzing genes and complete genomes by ion torrent sequencing.

2. Description of the Background

Mycobacterium tuberculosis (MTB), the causative agent for tuberculosis, is a highly transmissible bacterial pathogen with significant morbidity and mortality, particularly in HIV infected patients. Since 1997 tuberculosis has remained the leading cause of death in South Africa, a statistic linked to this country's growing HIV epidemic. Moreover, effective treatment measures in patients with active MTB have been exacerbated by increasing cases of multidrug resistance (MDR) and extensively drug-resistant (XDR) clinical isolates.

MDR tuberculosis strains are resistant to the first line antibiotics rifampin (RIF) and isoniazid (INH), while XDR MTB strains are resistant to both RIF and INH as well

as any fluoroquinolone and second-line injectable antibiotic drugs (e.g., amikacin, kanamycin or capreomycin). About 6% of all TB cases are MDR strains and South Africa continues to report higher percentages of XDR cases each year. While 7% of patients infected with standard MTB strains succumb to infection, the death rate rises to almost 50% with MDR tuberculosis. The emergence of antibiotic resistant MTB strains underscores an immediate need for rapid and highly accurate diagnosis, particularly in the developing countries of Africa. In addition migratory populations make geographical surveillance and tracking of drug resistance strains more urgent.

Culture-based drug susceptibility testing (DST) for MDR strains is considered the gold-standard, but is time consuming (weeks to months), technically challenging and cost prohibitive, especially in resource limited countries. For example, the BACTEC MGIT 960 (Becton Dickinson Microbiology System, Silver Sparks NV, USA), is an automated continuously culture-based monitoring system that measures bacterial oxygen consumption and can perform DST using prepared kits which are available for susceptibility of strains to a number of antibiotics. DST results obtained with the BACTEC MGIT 960 yield reliable and reproducible but require handling of viable and potentially infectious cultures, days to weeks or delay until results are available, specialized laboratory accommodations and high costs associated with the instrument and consumables.

In recent years, several nucleic acid based assays for determining MTB drug resistance have been developed. One of the most popular commercially available diagnostic assays is the GenoType MTBDR_{plus} Line Probe Assay (LPA) by Hain LifeScience. This test employs nucleic acid extraction, PCR amplification, probe hybridization and colorimetric visualization on lateral strips via an alkaline phosphatase reaction. LPA has been shown to be sensitive and specific, but there are several drawbacks. Sensitivity of the LPA for all resistance-associated mutations will most likely never reach 100% since many mutations that confer resistance have yet to be discovered. Another inherent limitation of the LPA is an inability to detect sample populations that contain a mixture of resistant and susceptible strains. Strains that harbor substitution mutations that change an amino acid to a previously uncharacterized or unknown mutation not presented on the LPA are not detected. Furthermore, the

LPA only allows detection of the most frequent mutations that cause resistance. If a strain were to contain mutations outside of the targeted mutations, the wild-type banding pattern will appear leading to a false negative (susceptible) result.

Thus, there is a need for a rapid, standardized, cost-effective protocol for full length gene analysis of critical genes such as, for example, genes associated with first and second line drug resistance.

Summary

The present invention overcomes disadvantages associated with current strategies and designs, and hereby provides tools, compositions, and methods for analyzing sequence information of nucleic acids including full-length genes and complete genomes.

One embodiment of the invention is directed to analyzing drug resistance mutations by semi-conductor sequencing and, preferably, ion torrent sequencing. Nucleic acid segments containing a gene of interest are amplified by PCR and the amplified products are processed and subsequently analyzed by sequencing. Sequencing is preferably by Ion Torrent, or Next-Generation sequencers including the Ion Torrent Personal Genome Machine (PGM). Preferably, the amplification products represent a common full-length, or multiple overlapping pieces of genes of a number of species, strains and/or serotypes of organisms. The amplified products are sequenced and mutations identified and mapped. Mapping identifies both known and previously unknown mutations and is useful to track the progress and movement of drug resistance across a population. Preferably, the invention analyzes nucleic acids of pathogens such as, for example, virus, bacteria or parasites. Preferably the viral pathogens are the causative agents of influenza or HIV and the bacterial pathogens are the causative agents of tuberculosis. Ion torrent sequencing of the nucleic acid segments provides enhanced sequencing for rapid, efficient, cost-effective protocol for full length gene analysis. Drug resistance and other mutations are immediately determined.

Another embodiment of the invention is directed to tools, compositions and methods for performing semi-conductor sequencing, preferably ion torrent sequencing, on complete genomes. The invention comprises obtaining a DNA sequence of an organism of interest and performing polymerase chain reaction analysis using multiple

pairs of nucleic acid primers. Each pair of primers is designed to simultaneously amplify overlapping segments of the genome under similar PCR conditions. Preferred primers possess similar GC content and overall size. A single PCR amplification of the genome produces hundreds of amplification products whose sequences include the full-length gene, large gene and noncoding segments or the entire genome of the organism. These products are analyzed, preferably by ion torrent sequencing, and the sequences matched to create a sequence map of the entire genome.

Another embodiment of the invention is directed to methods of identifying a sequence motif in the genome of a microorganism that confers resistance to an antimicrobial compound, comprising: providing multiple nucleic acid samples obtained from multiple different strains or serotypes of the microorganism; amplifying the sequences of the multiple nucleic acid samples by a polymerase chain reaction; obtaining sequence information of the amplified sequences by ion torrent sequencing; identifying a polymorphism in the genome of at least one microorganism strain or serotype from the sequence information obtained; and correlating the polymorphism identified with a phenotype or genome location of the at least one microorganism strain or serotype to identify the sequence motif that confers resistance to the antimicrobial compound. Preferably, the microorganism is a virus, a bacterium, a fungus or a parasite, and the virus is influenza virus and the bacterium is *Mycobacterium tuberculosis*. Preferably, the nucleic acid samples are provided in an aqueous molecular transport medium that contains a chaotrope, a detergent, a reducing agent, a chelator, a buffer, and an alcohol, together present in an amount sufficient to lyse cells, denature proteins, inactivate nucleases, kill pathogens, and not degrade nucleic acid. Preferably, amplifying is performed in a one step polymerase chain reaction utilizing a primer pair that amplifies a gene or nucleic acid segment associated with resistance to an antimicrobial compound, and the polymerase chain reaction is carried out in an aqueous mix comprising: a heat-stable polymerase; a mix of deoxynucleotide triphosphates comprising about equivalent amounts of dATP, dCTP, dGTP and dTTP, a chelating agent, an osmolarity agent, an albumin, a magnesium salt; and a buffer. Preferably the antimicrobial compound is an antibiotic.

Another embodiment of the invention is directed to methods of treating a disease or disorder caused by the at least one microorganism strain or serotype with the antimicrobial compound identified by the methods of the invention.

Another embodiment of the invention is directed to methods for determining a complete sequence of a genome of an microorganism comprising: producing a series of amplicons by performing a single polymerase chain reaction (PCR) of the genome in an aqueous mixture containing a heat-stable polymerase; a mix of deoxynucleotide tri phosphates comprising about equivalent amounts of dATP, dCTP, dGTP and dTTP; a chelating agent; a salt; a buffer; a stabilizing agent; and a plurality of primer pairs wherein each primer of the plurality of primer pairs has a similar annealing temperatures; sequencing each of the series of amplicons produced by semi-conductor sequencing, and correlating the sequences of the amplicons and constructing the complete sequence of the genome. Preferably, each of the primers of the multiple primer pairs comprise primers that are from 15 to 25 nucleic acids in length and each has a GC content of about 25-50%. Preferably, each primer pair is designed to PCR amplify an amplicon, and the collection of amplicons that are PCR amplified encompass overlapping segment of the complete genome sequence. Preferably, the plurality of primer pairs hybridize to the genome and are spaced along the genome at about every 500 to 2,000 nucleotides. Preferably, the microorganism is a virus, a bacterium, a fungus, a parasite or a cell, and the virus is influenza virus and the bacterium is *Mycobacterium tuberculosis*.

Another embodiment of the invention is directed to methods for determining the sequence of a nucleic acid segment in one step comprising: performing a polymerase chain reaction on the nucleic acid segment to produce a series of amplicons, wherein the PCR comprises: a heat-stable polymerase; a mix of deoxynucleotide tri phosphates comprising about equivalent amounts of dATP, dCTP, dGTP and dTTP; a chelating agent; a salt; a buffer; a stabilizing agent; and a plurality of primer pairs wherein each primer of the plurality of primer pairs has an annealing temperature within 5°C; sequencing each of the series of amplicons produced by semi-conductor sequencing, and correlating the sequences of the amplicons and constructing the sequence of the nucleic acid segment. Preferably the nucleic acid segment is 1 Mb or greater in length,

more preferably greater 2 or more Mb in length, more preferably 5 or more Mb in length and more preferably 10 or more Mb in length. Preferably, each of the primers of the multiple primer pairs is of from 16 to 24 nucleotides in length, has a GC content of about 28-35%, and has an annealing temperature of within 3°C of each other primer.

5 Preferably, each primer pair is designed to PCR amplify an amplicon representing a portion of the sequence of the nucleic acid segment, and the collection of amplicons that are PCR amplified represent overlapping portions of the complete sequence of the segment. Preferably, the plurality of primer pairs hybridizes to the segment at a spacing of about 800 to 1,200 nucleotides in length.

10 Another embodiment of the invention is directed to mixtures comprising multiple pairs of nucleic acid primers wherein, upon subjecting the collection to a polymerase chain reaction in association with a nucleic acid segment, the collection of primer pairs generates a collection of amplicons, wherein each amplicon is about 500 to 2,000 nucleotides in length, such that the entire sequence of the segment is represented
15 in the resulting collection of amplicons. Preferably, each primer of the collection of primer pairs is about 15 to 25 nucleotides in length, has a GC content of about 25-45%, and an annealing temperature within 3°C of each other primer, and each primer of the collection of primer pairs contains a sequence that hybridizes to the genome of the same microorganism. Preferably, the microorganism is a virus, a bacterium, a parasite,
20 or a fungus. Preferably, the mixture contains a heat-stable polymerase; a mix of deoxynucleotide tri phosphates comprising about equivalent amounts of dATP, dCTP, dGTP and dTTP; a chelating agent; a salt; a buffer; a stabilizing agent and nuclease-free water.

Other embodiments and advantages of the invention are set forth in part in the
25 description, which follows, and in part, may be obvious from this description, or may be learned from the practice of the invention.

Description of the Figures

Figure 1 Illustrates the *pncA* gene sequence plus 100 flanking base pairs as well as the reverse compliment sequence, the protein sequence, and the primers sequences.

30 Figure 2 Illustrates the nucleotide sequence of H37RV Gene strain as well as the sequences of the TB 16S ribosomal RNA gene sequencing primers.

Figure 3 Illustrates the *rpoB* gene conferring sensitivities/resistance to Rifampin as well as the forward and reverse primer sequences for *rpoB*.

Figure 4 Illustrates the *Mycobacterium tuberculosis* H37Ra, complete genome (GenBank: CP000611.1) *GyrA* Gene and three sets of forward and reverse primers.

5 Figure 5 *Mycobacterium tuberculosis* H37Ra, complete genome (GenBank: CP000611.1) catalase-peroxidase-peroxynitritase T *katG* and three sets of forward and reverse primers.

Figure 6 Illustrates the cycle threshold of Gyrase A and IS 6110 assays.

Figure 7 Illustrates Gyrase A assay and the IS 6110 assay using sequence isolates
10 by cycle number vs. Ct value.

Figure 8 Illustrates Gyrase A assay and the IS 6110 assay using sequence isolates vs cycle threshold (ct).

Figure 9 Summary of results achieved in sequencing the influenza A genome using various primer pair collections with ion torrent sequencing methodology.

15 Figure 10 Characterization of primer pairs for whole-genome ion torrent sequencing of influenza A (H3N2).

Figure 11 (A) Gene sequence of *pncA* showing coding regions as shaded, and (B) *pncA* forward and reverse primers utilized in PCR tiling and *pncA* regions P1-P4.

Description of the Invention

20 Rapid analysis of genes associated with drug resistant strains is a major challenge for successful treatment of many diseases and disorders. Real-time geographical surveillance of emerging MTB drug resistance would facilitate more appropriate treatment strategies (e.g., drug, antibiotic, chemical). Currently, available molecular methods such as the GenoType® MTBDR_{plus} LPA offer limited detection
25 capabilities, particularly when novel/uncommon amino acid substitutions are within known drug resistance regions or when undiscovered amino acid mutations impact drug resistance. Also, current methodology including Ion Torrent protocol requires multi-steps, ancillary equipment and increased expense, and is labor intensive.

A simplified semiconductor sequencing protocol for rapid characterization of
30 full-length genes and genome has been surprisingly discovered. The invention comprises a standardized protocol utilizing semiconductor sequencing and preferably

Ion Torrent sequencing of full-length genes. The protocol enables sequencing of entire coding regions implemented allowing characterization of known mutations and discovery of new polymorphisms. This protocol also enables the sequencing of megabases of nucleotide information such that complete genomes of cells and organisms can be determined and the genetic polymorphism readily mapped and identified. Preferably the cells or organisms are disease causing prokaryotic or eukaryotic cells, or yeast or fungal cells. Preferred disease causing organisms include strains of bacteria, virus, fungus, and parasites. Exemplary organisms include, but are not limited to Mycobacteria (e.g., *M. tuberculosis*), Influenza virus (e.g., H5N1, H1N1, H7N9), Respiratory Syncytial Virus, HIV, and Plasmodium (e.g., *Plasmodium falciparum*).

This relatively rapid (e.g., 1, 2, or 3 days, or less), standardized, cost-effective protocol allows for full length analysis of genes such as, for example, *rpoB*, *katG*, *gyrA*, *pncA* and *rrs* MTB genes associated with first and second-line MTB drug resistance (see Figures 1-5). The protocol was evaluated using 26 geographically diverse clinical isolates collected in South Africa including MDR and XDR strains with next-generation Ion Torrent Personal Genome Machine (PGM). The sequencing data obtained from this developed methodology were compared to the HAIN LPA and genotyping DST data from culture. This methodology for the first time enables sequencing entire coding regions for genes implemented in resistance allowing characterization of known mutations and discovery of new polymorphisms. Previously uncharacterized substitution mutations were identified on the *rrs*, *rpoB*, *katG*, *pncA* and *gyrA* genes.

The present invention offers significant potential for next-generation instruments to be more utilized in resource deprived environments such as Africa, Asia, and India. Specifically, the current invention improves and streamlines the up-front library preparation process. Methodology of the invention does not require the use of expensive ancillary equipment pieces typically utilized or required by the manufacturer. Specifically, the standardized procedure of the invention does not require an Agilent Bioanalyzer for DNA quantifications; the OneTouch ePCR system for emulsification PCR step, or the PipinPrep for gel excision. Additionally, since the protocol of the invention involves re-sequencing full-coding genes (not necessarily full genomes) the

Bioruptor is not required for shearing DNA into smaller pieces. Additionally, it is not necessary to sequence the entire genome and then identify genes. The method and tools of the invention allow for pre-selection of the genes and/or regions of interest that are to be sequenced. As the Agilent 2100 BioAnalyzer, OneTouch, PipinPrep, and Bioruptor all require additional training for use, consume valuable laboratory bench space, and are extremely expensive, the invention represents a significant advance and improvement over convention methodologies.

The sequencing protocol of the invention is exemplified herein using Ion Torrent sequencing as this sequencing method has been applied to *M. tuberculosis*. As believed to clear to those skilled in the art, the protocol involves semiconductor sequencing, with is exemplified by Ion Torrent sequencing and, as such, involves the sequencing of large numbers of different regions simultaneously. The sequencing and nucleic acid methodologies are applicable to any series of genes, genomes or nucleic acid sequences.

The invention also includes a methodology for selecting primer pairs for sequencing a target of interest. Primer pairs are preferably selected with matched annealing and melting temperature as to the target. Preferably, melting and annealing temperatures are based on sequence characteristics such as the GC content of the sequence, the possibility of self-hybridization of the primer (e.g., forming hairpin loops within the primer), and possible structures near the binding site. Preferably the primers do not self-hybridize under the conditions of sequencing. Preferably the GC content of primers is between about 25% and 50%, more preferably between about 30% and 40%, more preferably between about 25% and 35%, and also more preferably between about 40% and 50%. Thus, primer sequences of the target are selected for hybridization based on sequence characteristics such that all of the primer pairs utilized for the target will have similar melting and/or annealing temperatures to the target. Preferably primer sequences contain no regions of reasonably possible self hybridization of the primer sequence. Preferably primer pairs are matched for annealing and/or disassociation temperatures which may be within 5°C, within 4°C, with 3°C, within 2°C, with 1°C and more preferably the same annealing temperature, the same melting temperature or both. Primer pairs preferably generate amplicons of between about 500

to about 2,000 nucleotides (NT) in length that represent overlapping segment of the target, more preferably between about 600 and 1,500 NT, more preferably between about 700 and 1,300 NT, more preferably between about 800 and 1,200 NT, more preferably between about 900 and 1,100 NT, and more preferably about 1,000 NT.

5 Primers are generally between 12 NT and 45 NT in length, more preferably between 15 and 35 NT, and more preferably between about 18 and 25 NT. Although not a rule, generally longer primers have a lower GC content. Exemplary primers pairs are identified for the *pncA* gene (see Figure 1), the H37RV gene strain (see Figure 2), the *rpoB* gene (see Figure 3), the *GyrA* gene (see Figure 4, and the *katG* gene (see Figure
10 5).

In one embodiment of the invention, a semiconductor sequencing protocol was determined for five genes of *M. tuberculosis* for determining drug resistance in MDR and XDR strains (e.g., cumulatively sequencing 11.4 kb per isolate). The *M. tuberculosis rpoB* gene encodes a 1,178 amino acid beta subunit for an RNA dependent
15 DNA polymerase enzyme. Mutations within an 81-bp “core region” of the *rpoB* gene are responsible for approximately 95% of rifampin resistance in *M. tuberculosis* strains. Three of these mutations at positions 516 (D→V), 526 (H→Y/D), and 531 (S→L) constitute the majority of mutations within this region. Of the 21 rifampin-resistant strains characterized in this study, 11 (52.4%) carried the S531L mutation, 7 (33.3%)
20 contained an amino acid substitution at position 516, and 3 (14.3%) contained a mutation at position 526 of the *rpoB* gene (Table 1). The most prevalent *rpoB* substitution observed at position 516 is a valine (D516V). Ion Torrent sequencing according to the invention revealed that 6 of 7 strains contained a rarer glycine residue (D516G) at this position (Table 1). These 6 strains were shown as absent for both
25 mutant and wild type bands by LPA (Table 1). Similarly an uncommon amino acid substitution was identified at position 526 in the *rpoB* gene. The most prevalent amino acid substitution reported at position 526 in the *rpoB* gene is a histidine to tyrosine or aspartic acid (H526Y/D). Ion Torrent sequencing revealed 1 of 3 isolates contained an uncommon arginine (R) residue (H526R) that by HAIN LPA was shown to be absent
30 for both wild type and mutant bands (Table 1). While absence of wild type and mutant bands in a sample are interpreted as resistant according to LPA testing, there remains

ambiguity since the type of amino acid change is not directly characterized. This underscores the utility of Ion Torrent sequencing for resistance surveillance, and discovery of novel amino acids in circulating MTB strains.

Table 1

Summary of 10 amino acid mutations in the first 900 amino acid residues* of the *rpoB* gene of 26 (14 MDR, 7 XDR and 5 fully susceptible) *M. tuberculosis* isolates from South Africa deduced by Ion Torrent sequencing, Hain LPA genotyping and culture.

Isolate No.	Amino Acid Substitution(s)**		Rifampin Result by		
	of <i>rpoB</i> gene (3619 bps)	Ion Torrent*	HAIN LPA	Bacter MGIT 960	
9	S531L	Resistant	Resistant	Resistant	
1	S531L, V194I	Resistant	Resistant	Resistant	
1	S531L, Y645H	Resistant	Resistant	Resistant	
15	1 H526D	Resistant	Resistant	Resistant	
1	H526Y, S509R	Resistant	Resistant	Resistant	
1	H526R	Resistant	Resistant	Resistant	
5	wild type **	Sensitive	Sensitive	Sensitive	
6	D516G, L533P	Resistant	Resistant	Resistant	
20	1 D516V	Resistant	Resistant	Resistant	

* There were 5 *rpoB* amino acid substitutions (R908C, Q1042H, P1043A, I1187T, and V1249F) noted in at least 1 strain at the 3' end (residues 900-1253).

** Compared to the sequenced H37Rv reference strain.

The *katG* gene encodes catalase peroxidase, an enzyme that converts isoniazid (INH) into the active form. The majority of isoniazid resistance is associated with *katG* codon 315 (S315T), although mutations in the promoter region of *inhA* and *nod* also contribute to resistance. Of 26 strains assessed, 16 (62%) contained the characteristic serine-to-threonine amino acid substitution at position 315 (S315T) conferring isoniazid resistance (Table 2). These sequencing results exhibited 100% concordance with comparisons made using the HAIN LPA and culture DST.

Table 2

Summary of 4 amino acid mutations in the *katG* gene of 26 (14 MDR, 7 XDR and 5 fully susceptible) *M. tuberculosis* isolates from South Africa deduced by Ion Torrent sequencing, Hain LPA genotyping and culture

Isolate No.	Amino Acid Substitution(s)**		Rifampin Result by		
	of <i>katG</i> gene (2447 bps)	Ion Torrent*	HAIN LPA	Bacter MGIT 960	
40	11 S315T	Resistant	Resistant	Resistant	
5	S315T, R463L	Resistant	Resistant	Resistant	
1	W191R, R463L	Sensitive	Sensitive	Sensitive	
7	wild type**	Sensitive	Sensitive	Sensitive	
1	R463L	Sensitive	Sensitive	Sensitive	
45	1 N138H	Sensitive	Sensitive	Sensitive	

* Rifampin resistance is known to occur in *rpoB* at positions 531(S→T), 526 (H→Y/D), and 516 (D→V).

** Compared to the sequenced H37Rv reference strain.

5

Pyrazinamide (PZA) is a synthetic derivative of nicotinamide that has been used as a first-line drug to fight tuberculosis since 1952. Standard DST for PZA is complicated due to an acidic pH requirement *in vitro*, which inhibits *M. tuberculosis* growth and complicates accurate phenotypic assessment. PZA resistance is attributed to mutations in the *pncA* gene which encodes a pyrazinamidase. These resistance conferring mutations are numerous and include amino acid substitutions, frameshifts and stop codon mutations. Seven mutations were characterized from the 26 South African isolates assessed, including one silent mutation, 5 amino acid substitutions, and one chain termination mutation. The Q122 (Stop) termination mutation (Table 3) observed in one isolate is novel, having not been reported elsewhere. The difficulty in PZA phenotypic assessment and the variability of mutations along the *pncA* gene further underscores the added value of Ion Torrent gene sequencing to assess mutations in this hyper variable MTB gene.

Table 3
Summary of 6 amino acid mutations in the *pncA* gene of 26 (14 MDR, 7 XDR and 5 fully susceptible) *M. tuberculosis* from South Africa deduced by Ion Torrent sequencing and culture

Isolate No.	Amino Acid Substitution(s)** in the <i>pncA</i> gene (3619 bps)	Pyrazinamide Result by Ion Torrent*	Bacter MGIT 960
3	C14R	Resistant	Resistant
1	A102V	Resistant	Resistant
1	Q122 (stop)	Resistant	Resistant
16	wild type **	Sensitive	Sensitive
1	V139G	Resistant	Resistant
1	R154G	Resistant	Resistant
2	L172P	Resistant	Resistant
1	Silent (C195T)	Sensitive	Sensitive

* pyrazinamide resistance is known to occur in 25 mutations described by Mphahlele et al (23).

** Compared to the sequenced H37Rv reference strain. One strain contained a silent (synonymous) nucleotide mutation at position 195 (C→T).

The primary target of fluoroquinolones (FQ) in *M. tuberculosis* is DNA gyrase, a type II topoisomerase composed of two A and B subunits encoded by the *gyrA* and *gyrB* genes, respectively. Amino acid substitutions located within a short region of the

gyrA gene known as the quinolone resistance-determining region (QRDR), account for the majority of known FQ resistant tuberculosis strains. Substitution mutations in the QRDR at positions 88, 90, and 94 were observed in 10 of 26 (38.5%) sequences from this study (Table 4). Three of these 10 strains contained substitutions at position 94 in the *gyrA* gene; two were noted as D94G substitutions, and one was a D94Y substitution. Both D94G and D94Y have been characterized as substitutions and both amino acid substitutions at codon 94 give rise to similar levels of FQ antibiotic resistance. Of the strains assessed, the *gyrA* gene was the most variable containing nine amino acid substitutions in the 26 clinical isolates assessed. Furthermore, two of these *gyrA* codons (549 and 613), exhibited heterogeneous residues (Table 4), an advantage of performing Ion Torrent sequencing over HAIN LPA and DST.

Table 4

Summary of 10 amino acid mutations in the *gyrA* gene of 26 (14 MDR, 7 XDR and 5 fully susceptible) *M. tuberculosis* isolates from South Africa deduced by Ion Torrent sequencing and culture

<u>Isolate No.</u>	<u>Amino Acid Substitution(s)** in the <i>gyrA</i> gene (2664 bps)</u>	<u>Rifampin Result by</u>	
		<u>Ion Torrent*</u>	<u>Bacter MGIT 960</u>
3	E21Q, S95T, G2475S, G668D	Sensitive	Sensitive
2	E21Q, D94G, S95T, G668D	Resistant	Resistant
1	E21Q, G88C, S95T, G668D	Resistant	Resistant
10	E21Q, S95T, G668D	Sensitive	Sensitive
1	wild type**	Sensitive	Sensitive
1	E21Q, S95T, G668D, Q613Q/E+	Sensitive	Sensitive
1	E21Q, S95T, G668D, L549S/L+	Sensitive	Sensitive
1	E21Q, D94Y, S95T, G668D	Resistant	Resistant
6	E21Q, A90V, S95T, G247S, G668D	Resistant	Resistant

* Fluroquinolone resistance is known to occur in *gyrA* at position 88(G→C), 90 (A→V), 91 (S91P) and 94 (D→H).

** Compared to the sequenced H37Rv reference strain.

+ There is a heterozygous nucleotide mutation in a proportion of Ion Torrent reads; the mutation confers a mixed amino acid substitution.

Emerging cases of XDR tuberculosis defined as MDR cases having acquired additional resistance to FQ, i.e., ofloxacin, and at least one of the three injectable 'second-line drugs', i.e., amikacin (AMK), kanamycin (KAN), or capreomycin (CAP), have become a public health threat in developing countries worldwide. The majority of resistance to second line drugs is associated with mutations in codons 1401 (A1401G), 1402 (C1402T), and 1484 (G1483T) in the 16 S ribosomal RNA *rrs* gene. Analysis of African MTB strains revealed that 7 of 26 (27%) were defined as XDR as evident by

nucleotide mutation at position 1401 (A1401G) (Table 4). Three additional nucleotide mutations at positions 492, 514, and 878 were also discovered (Table 5) in strains from this analysis. The G878A is a novel nucleotide mutation but was shown to be sensitive to AMK, KAN, and CAP according to DST.

Table 5

Summary of 4 nucleotide mutations in the *rrs* (16s) gene of 26 (14 MDR, 7 XDR and 5 fully susceptible) *M. tuberculosis* isolates from South Africa deduced by Ion Torrent sequencing and culture.

Isolate No.	Amino Acid Substitution(s)** in the <i>rrs</i> (16s) gene (1680 bps)	Kanamycin Result by	
		<u>Ion Torrent*</u>	<u>Bacter MGIT 960</u>
1	G878A	Sensitive	Sensitive
12	wild type**	Sensitive	Sensitive
1	A514C, A1401G	Resistant	Resistant
6	A1401G	Resistant	Resistant
3	A514C	Sensitive	Sensitive
1	C492T	Sensitive	Sensitive
1	C492T, A514C	Sensitive	Sensitive
1	A514C	Sensitive	Sensitive

* Aminoglycoside resistance is known to occur at positions 1401 (A→G), 1402 (C→T), and 1484 (G→T).

** Compared to the sequenced H37Rv reference strain.

Previous studies have shown that mutations in *katG* codon 463, and *gyrA* codon 95 are genetic markers for categorizing strains into epidemiological genetic Groups 1, 2, and 3, and that these codons have no effect on antibiotic resistance. Group 1 strains are genetic ancestors of Group 2 and Group 3 strains that link the predominately non-human mycobacterium genus (*M. microti* and *M. bovis* strains) with human *M. africanum* and *M. tuberculosis* lineages. As evident by substitution mutations in *katG* codon 463 and *gyrA* codon 95, a total of 7 of 26 (27%), 18 of 26 (69%), and 1 of 26 (4%) of the African isolates characterized in this study were members of genetic Group 1, 2, and 3, respectively. Tracking Group 1 organisms is important in terms of MTB detection since several isolates belonging to genetic Group 1 lack Insertion Sequence 1661 (IS-1661), a common genetic target for several PCR-based MTB detection assays.

The Ion Torrent protocol for MTB drug resistance can be easily integrated into low resource settings throughout countries and regions such as Africa, India, and China. The Ion Torrent methodology does not require the use of expensive ancillary equipment such as Agilent 2100 BioAnalyzer, DiaGenode Bioruptor® Sonication

System, Ion OneTouch System™, ultracentrifuges, or the Pippin Prep™ Workstation as current Ion Torrent protocols recommend. This is significant since these instruments and needed accessories and consumables can be expensive, require large laboratory footprints, and often require routine maintenance.

5 In contrast to the GenoType® MTBDR*plus* or MTBDRsl Line Probe Assay (LPA), the Ion Torrent PGM protocol provides full-length characterization of genes, making possible discovery of new amino acid substitutions that could potentially be missed by LPA since LPA is limited to only known mutations. Using the protocol, several uncommon amino acid changes in clinical field isolates have been found.
10 Furthermore, the extensive depth of sequence coverage from the Ion Torrent allows for discovery of heterogeneous or mixed strain genetic populations within an isolate.

The scalability of Ion Torrent sequencing permits expansion to include megabases of additional genes on a single chip. The methodology of the invention is expandable beyond the five full-length MTB genes to include all 16 plus genes that
15 currently constitute MTB drug resistance. Full-length gene analysis using the Ion Torrent PGM will identify novel mutations that, when correlated to phenotypic minimal inhibitory concentration (MIC) testing, identify new tuberculosis resistant residues as well as the cumulative inhibitory effect of multiple mutations.

Another embodiment of the invention is directed to megabase sequence
20 identification utilizing semiconductor sequencing protocols. Megabase sequencing according to the invention involves selection of primer pairs that amplify different sections of the target sequence whereby the collection of sections represent the entirety of the target sequence. Preferably the sections overlap to a degree that permits alignment of the resulting amplicons forming the complete target sequence. Primer
25 pairs are preferably designed to form amplicons with lengths of about 0.5k to about 5k nucleotides, preferably about 0.6k to about 3 k nucleotides, more preferably about 0.7k to about 2k nucleotides, and more preferably about 0.8k to about 1k nucleotides. Primer pairs are preferably of similar GC content such that the annealing or hybridization temperatures are as similar or preferably within about 5°C, more
30 preferably within about 2°C, and more preferably within about 1°C. Also preferred is that the hybridization disassociation temperatures be similar, such that annealing and

disassociation occur at very similar temperature for polymerization and PCR. In annealing and disassociation, the length of the primer influences the temperature profile, thus similar length for the all or at least most of the primers is preferred. Primer lengths are preferably about 15-30 nucleotides, more preferably about 20-28
5 nucleotides, and more preferably about 18 to 25 nucleotides. Although it is preferred that all of the primers have such similar characteristics, megabase sequencing can be performed when greater than about 80% of the primers share one or more characteristics, more preferably 85% or more, more preferably 90% or more, and even more preferably 95% or more. In accordance with the utilization of such similar
10 primers, a PCR reaction is performed with one target nucleic acid to be amplified with a mixture of all primer pairs. Also preferred is performance of duplicate PCR analysis on identical mixtures. The number of cycles can range from 10 to 50 or more and, preferably temperature cycling is performed in accordance with convention PCR temperature and reaction conditions.

15 The following examples illustrate embodiments of the invention, but should not be viewed as limiting the scope of the invention.

Examples

Clinical Isolates. A total of 26 geographically diverse clinical isolates, representing drug-sensitive, MD, and XDR tuberculosis strains were obtained from sample archives
20 of the University of Pretoria, South Africa, and the National Institute for Communicable Diseases (NICD), Sandringham, South Africa. The H37Rv MTB lab strain was included as a sequencing control throughout the protocol. All MTB isolates used were archived strains from pure culture MGIT™ 960 System tubes (Becton Dickinson, Sparks, MD) with species identification and genotypic resistance to
25 rifampin and isoniazid determined using the Genotype® MTBD*plus* assay (HAIN LifeSciences, Germany) according to manufacturer's instructions. Phenotypic resistance for first and second line drugs was performed using the MGIT™ 960 System as previously described. Critical concentrations for ofloxacin and kanamycin (second line drugs) were 2.0 µg/mL and 5.0 µg/mL, respectively. Resistance to first and second
30 line drugs was determined using standard diagnostics algorithms.

DNA preparation. MTB isolates were handled in blinded fashion throughout. MTB samples (0.5 mL) were pipetted into cryovial tubes containing 1.5 mL PrimeStore Molecular Transport Medium® (a molecular transport medium) (Longhorn Vaccines & Diagnostics, San Antonio, TX). Inactivated samples were transported from South Africa to San Antonio, Texas, USA at ambient temperature (3-4 days) and stored at 5°C until used. Total DNA (50 µl) was purified from a 200 µl aliquot of PrimeStore MTM® containing inactivated culture using a Qiagen® EZ1® Advanced Robot and EZ1 DNA Tissue Kit (Cat No. 953034) according to manufacturer's recommendations (Qiagen Inc., Germantown, MD).

Primer Design. Novel PCR primers were designed for amplification of full-length coding regions for each MTB gene of interest (Table 6).

Table 6
PCR amplification primers used for full length analysis of MTB Genes.

Amplification Target	Forward	Reverse	Amplicon (bp)
<i>rpoB</i>	5' - TCCTCTAAGGGCTCTCGTT -3' 19nt (SEQ ID NO 1)	5' - GTCAGGTACACGATCTCGT -3' 19nt (SEQ ID NO 2)	1625
<i>rpoB</i> ll (2 half)	5' - ATCGAAACGCCGTACCGCAA -3' 20nt (SEQ ID NO 3)	5' - TGACGTCGAGCACGTAACCTCCCT -3' 23nt (SEQ ID NO 4)	2056
<i>katG</i>	5' - ACACCAACTCCTGGGAAGGAAT - 3' 22nt (SEQ ID NO 5)	5' - TGATCGCACATCCAGCACATTT -3' 22nt (SEQ ID NO 6)	2447
<i>pncA</i>	5' - GACGGATTTGTCGCTCACTAC -3' 21nt (SEQ ID NO 7)	5' - GCCGGAGACGATATCCAGAT -3' 20nt (SEQ ID NO 8)	960
<i>gyrA</i>	5' - AAGGATGTTTCGGTTCCTGGAT -3' 21nt (SEQ ID NO 9)	5' - TAACACTCGTACCCGGCT -3' 18nt (SEQ ID NO 10)	2664
<i>rrs</i> (16s)	5' -TTCTAAATACCTTTGGCTCCCT -3' 22nt (SEQ ID NO 11)	5' -TGGCCAACTTTGTTGTCATGCA -3' 22nt (SEQ ID NO 12)	1680
(5 genes) 11,432 BP total			

Primer pairs for *rpoB* (2 sets of primers), *katG*, *pncA*, *gyrA*, and *rrs* (16s) gene amplification were designed using the genome sequence of *M. tuberculosis* H37Rv strain as reference (GenBank accession no. NC_000962). Primer secondary structure, melting temperature, and potential primer-dimer formation were determined using LaserGene 9.1 (DNASar, Madison, WI) and PrimerExpress 3.0 (Life Technologies, Foster City, CA). All oligonucleotides were synthesized using standard, de-salted primers (Integrated DNA Technologies (IDT), San Diego, CA).

PCR Amplification. Amplification reactions for all MTB gene targets were designed and optimized to be used under one standardized set of thermocycling parameters. All

PCR ‘mastermixes’ were prepared using Platinum Taq DNA Polymerase, 10X Buffer, and 50 mM MgCl₂ (P/N 10966-034; Life Technologies, Foster City, CA). Amplification was carried out in a 50 µl final volume reaction mixture containing 24.1 µl Ambion Nuclease-Free Water (Cat No. AM 9932; Life Technologies, Foster City, CA), 5 µl 10X PCR Buffer, 2 µl 50 mM MgCl₂ (2 mM final), 0.4 µl PCR Nucleotide Mix Ultrapure dNTPS (200 µM final for each dNTP; P/N 77119; USB, Santa Clara, CA), 0.5 µl Platinum Taq DNA Polymerase (2.5 Units final), and 2 µl primer blend (*rpoB*, *katG*, *pncA*, *gyrA*, or *rrs* genes; 0.4 µM final for each primer). To each 34 µl ‘mastermix’ reaction mixture, 16 µl extracted DNA was added to bring the total volume to 50 µl. Reactions were carried out in MicroAmp Optical 96-Well Reaction Plates (P/N N801-0560, Life Technologies, Foster City, CA) and capped using MicroAmp 8-Cap Strips (P/N 4323032, Life Technologies, Foster City, CA). Amplification was performed using an ABI 2720 thermocycler (Life Technologies, Foster City, CA). Thermocycling parameters were 95°C for 2 minutes, followed by 40 cycles at 95°C for 30 seconds, 55°C for 15 seconds, and 72°C for 2 minutes with final extension at 72°C for 5 minutes. Resulting amplicons were confirmed by addition of 5 µl PCR product with 1 µl GelPilot Loading Dye 5X (P/N 1037649; Qiagen, Germantown, MD) on 1% (wt/vol) Molecular Biology Grade Agarose (Cat No. BP1356; Fischer Scientific, Pittsburg, PA) with ethidium bromide (0.1 µg/mL final; Cat No 161-0433; Bio-Rad, Hercules, CA). Electrophoretic separation of products was carried out for 60 minutes at 0.4 mV cm² in 1X Tris Borate-EDTA (TBE) Buffer (Cat No. 1B70153; IBI Scientific, Peosta, IA). Amplicons were visualized under UV transillumination, and size estimation made using a TrackIt 1 kb Plus DNA Ladder (P/N 10488-085; Life Technologies, Foster City, CA). After visualization, the remaining PCR reaction for each clinical isolate gene amplification (~45µL) corresponding to *rpoB*, *katG*, *pncA*, *gyrA*, and *rrs* (16s) targets were transferred to a single microcentrifuge tube. Pooled genes corresponding to each clinical isolate were subjected to PCR purification and eluted in 50 µl Low Tris-EDTA (TE) (Cat No. 602-1155-010; Life Technologies, Foster City, CA) using the MinElute Reaction Cleanup Kit (Cat No. 28204; Qiagen, Germantown, MD) according to manufacturer’s instructions. The concentration and purity of DNA was determined

spectrophotometrically using a NanoDrop ND 1000 (Thermo Fischer Scientific, Wilmington, DE).

Ion Torrent Library Preparation. Barcoded libraries were generated using the Ion Xpress Plus Fragment Library Kit (Cat No. 4471269, Life Technologies, Foster City, CA) and the Ion Xpress DNA Barcoding 1-16 Kit (Cat No. 4468654, Life Technologies, Foster City, CA) according to a modified version of the protocol outlined in the Ion Xpress Plus gDNA and Amplicon Library Preparation.

Amplicon Shearing. Chemical shearing was performed using 1-3 µg DNA containing an approximate equimolar pool of *rpoB*, *katG*, *pncA*, *gyrA*, and *rrs* (16s) gene amplicons. DNA shearing was performed in a 50 µl total reaction volume by combining 5 µl Ion Shear Plus 10X Reaction Buffer, 10 µl enzyme, and 35 µl pooled DNA template (Ion Xpress Plus Fragment Library Kit, Cat No. 4471269, Life Technologies, Foster City, CA). The reaction mixture was incubated at 37°C for 45 minutes, terminated using 5 µl Ion Shear Stop Buffer, and stored on ice until purification. Sheared DNA was purified using Agencourt Ampure XP-PCR Purification beads (P/N A63880; Beckman Coulter, Brea, CA) with Dynal magnetic bead stand (Cat No. 123-21D; Life Technologies, Foster City, CA) according to manufacturer's recommendations. Briefly, 99 µl Agencourt beads was mixed with 50 µl ion shear reaction, incubated for 5 minutes at room temperature, placed on a magnetic stand, washed twice with 70% (v/v) ethanol, and eluted using 12 µl Low TE Buffer (Cat No. 602-1155-010; Life Technologies Inc., Foster City, CA).

Adaptor Ligation. Adaptor ligation was performed in a 0.2 mL low bind PCR tube (P/N PCR-02-L-C; Axygen Inc., Union City, CA) by combining 12 µl sheared amplicon with 1.25 µl Ligase Buffer, 1.25 µl P1-IA Adaptor Mix (Ion DNA Barcoding 1-16 Kit, Cat No. 4468654 Life Technologies, Foster City, CA) and 0.2 µl DNA Ligase (Ion Xpress Plus Fragment Library Kit, Cat No. 4471269, Life Technologies, Foster City, CA). The mixture was pipetted up and down 5 times and incubated at room temperature (22-25°C) for 30 minutes. Adaptor ligation reactions were purified and eluted in 10 µl Low TE Buffer using the Agencourt Ampure XP-PCR Purification beads (P/N A63880; Beckman Coulter, Brea, CA) with the Dynal magnetic bead stand

(Cat No. 123-21D; Life Technologies, Foster City, CA) according to manufacturer's recommendations..

Nick Translation and Barcode Amplification. Amplicon pools from each patient sample were barcoded using the Ion DNA Barcoding 1–16 Kit and Ion Xpress
5 Fragment Library Kit (Part Nos. 4468654 and 4471269, respectively; Life Technologies, Foster City, CA). To maximize yields reactions were scaled 2x by combining 40 µl Platinum PCR SuperMix High Fidelity, 4.4 µl of Ion Primer Mix (BC X where X= barcode 1-16) and 10 µl of ligated DNA. Amplification was performed using an ABI 2720 thermocycler (Life Technologies, Foster City, CA). Thermocycling
10 parameters comprised 72°C for 20 minutes, 95°C for 5 minutes, followed by 10 cycles of 95°C for 15 seconds, 58°C for 15 seconds and 68°C for 1 minute. Following amplification, bar-coded samples were purified and eluted in 50 µl of Low TE (Cat No. 602-1155-010; Life Technologies, Foster City, CA) using the MinElute Reaction Cleanup Kit (Cat No. 28204; Qiagen, Germantown, MD) according to manufacturer's
15 instructions. DNA concentration and purity was determined by spectrophotometric analysis using a NanoDrop ND 1000 (Thermo Fischer Scientific, Wilmington, DE). Ranges for purified bar-coded samples are typically 150-300 ng/µl with A260/280 purity of 1.7-1.9. Equimolar concentrations (~2-3 µg of each bar-coded sample) were combined into a single 1.5 mL nuclease-free microcentrifuge tube and used for size
20 selection.

Size Selection. The appropriate volume of GelPilot 5X Loading Dye (P/N 1037649; Qiagen, Germantown, MD) was added to the pooled bar-coded MTB library tube and loaded onto a 1 % (w/v) agarose gel (Cat No. BP1356; Fischer Scientific, Pittsburg, PA) containing ethidium bromide (0.1 µg/mL final; Cat No 161-0433; Bio-Rad,
25 Hercules, CA). The bar-coded library was electrophoresed for 60 minutes at 0.4 mV cm² in 1X TBE Buffer (Cat No. 1B70153; IBI Scientific, Peosta, IA) and visualized under UV transillumination. Size estimations were determined using a TrackIt 1 kb Plus DNA Ladder (P/N 10488-085; Life Technologies, Foster City, CA). Gel excision was performed under UV transillumination using a sterile scalpel blade excising out a
30 target region between 75–200 bp. Excised agarose gel slices were placed into sterile 1.5 mL microcentrifuge tubes and subjected to DNA purification using the PureLink

Quick Gel Extraction Kit (Cat No. K210012; Life Technologies, Foster City, CA) according to manufacturer's instructions. Concentration and purity values for the barcoded DNA library were determined spectrophotometrically using a NanoDrop ND 1000 (Thermo Fischer Scientific, Wilmington, DE). The recommended library input for emulsion PCR is $\sim 140\text{-}560 \times 10^6$ molecules per 18 μL . This range was achieved by a 1:1000 dilution using library stock and nuclease-free water.

Emulsion Polymerase Chain Reaction (emPCR). Emulsion Polymerase chain reaction was performed in a 1 mL reaction volume using the Ion Template Preparation Kit (Cat No. 4469000; Life Technologies, Foster City, CA) by adding 582 μL nuclease-free water, 200 μL 5X PCR Reagent Mix, 100 μL 10X PCR Enzyme Mix, 100 μL Ion Sphere Particles, and 18 μL diluted library template. The preparation was mixed thoroughly followed by brief centrifugation in a microcentrifuge. Emulsion was achieved using the Ultra-Turrax Tube Drive (Life Technologies, Foster City, CA). A total of 9 mL chilled Emulsion Oil (Ion Torrent Preparation Kit; Cat No. 4469000, Life Technologies, Foster City, CA) was added to an Ion Template Preparation Tube (Cat No. 4467226, Life Technologies, Foster City, CA). The emulsion tube was placed and locked onto the IKA Ultra-Turrax Tube Drive and initiated. While the tube was in motion, the entire 1 mL PCR master mix solution was dispensed into the cap port and mixed for 5 minutes. The mixed emulsion was transferred to a 96-well PCR plate and amplified using an ABI 2720 thermocycler (Life Technologies, Foster City, CA) using the following thermocycling parameters: 94°C for 6 minutes, followed by 40 cycles at 94°C for 30 seconds, 58°C for 30 seconds, and 72°C for 90 seconds; followed by 5 cycles at 94°C for 30 seconds, and 68°C for 6 minutes.

Ion Sphere Particle (ISP) Recovery and Qubit Measurement. Ion Sphere Particles were recovered using reagents supplied in the Ion Xpress Template Kit (Cat No. 4469001, Life Technologies, Foster City, CA) according to manufacturer's protocol (Ion Xpress Template Kit User Guide v2.0, pages 18-19). Quantification of recovered particles was performed using a Qubit 2.0 Fluorometer (Life Technologies, Foster City, CA) and an Ion Sphere Quality Control Kit (Cat No. 4468656, Life Technologies, Foster City, CA) according to manufacturer's recommendations (Ion Xpress Template Kit User Guide, page 25-26). The optimal amount of template-positive ion sphere

particles (ISPs) is between 4-50%. Relative fluorescent unit (RFU) values obtained outside of this range were not pursued into subsequent ISP enrichment.

ISP Enrichment. ISPs were enriched using reagents supplied in the Ion Xpress Template Kit, Ion Sequencing Kit, and DynaBeads MyOne Streptavidin C1 beads (Cat Nos. 4469001, 4468997 and 650.01 respectively; Life Technologies, Foster City, CA) according to the manufacturer's protocols (Ion Xpress Template Kit User Guide v2.0, pages 15-17).

Ion Torrent 314 Chip Preparation and PGM Sequencing. Ion Torrent 314 Chips (Cat No. 4462923; Life Technologies, Foster City, CA) were prepared and loaded according to manufacturer's recommendation (Ion Sequencing Kit User Guide v 2.0). The Ion Torrent PGM was run according to Ion Torrent 314 Chip specifications including a 65-cycle sequencing protocol, use of 18 megaOhm purified water, and standard compressed argon gas to drive fluidics through the PGM system. All *rpoB*, *katG*, *pncA*, *gyrA* and *rrs* genes and corresponding proteins were deposited into GenBank (accession numbers JX303203-JX303332).

Gyrase PCR for the Detection of TB vs. 6110 PCR assay. The gyrase target for OCR and whole Gyrase gene sequencing on the Ion Torrent PGM can also be used to identify TB mutations that lead to resistance. This second PCR target allows for the accurate analysis of TB strains that may not include the entire IS6110 insertion element. While the IS6110 assay has multiple gene copies in most strains, some have only one. As shown in Figures 6, 7 and 8, this Gyrase assay has a generally higher cycle threshold in comparison to the IS6110 assay due to multiple IS6110 gene copies in those isolates and thus more sensitivity. Thus any possible TB mutation can be followed- even away from the detection site by this method of full gene sequencing.

Phenotypic and genotypic results. Amino Acid characterization of 26 *M. tuberculosis* isolates by Ion Torrent sequencing of *rpoB*, *katG*, *pncA*, *gyrA*, and *rrs* (16s) genes are summarized in Tables 1-5, respectively, and compared to BACTEC™ MGIT™ 960 (phenotypic), and/or HAIN GenoType® MTBDRplus (genotypic) LPA. Of the 26 MTB clinical isolates, 14 (54%) were MDR, 7 (27%) were XDR, and 5 (19%) were sensitive to drugs by BACTEC™ MGIT™ 960 phenotypic analysis. The Ion Torrent PGM sequencing method showed 100% (26/26) concordance to both

phenotypic resistance obtained by MGIT™ 960 culture (Tables 1-5) and genotypic *rpoB* and *katG* data obtained by Hain LPA (Table 1, 2).

***rpoB* gene mutations.** A total of 10 *rpoB* amino acid substitutions were identified in the 26 clinical isolates compared to the H37Rv wild type strain. The common S531L mutation was the most prevalent, but mutations in codons 516 and 526, also known to confer resistance to rifampin were observed (Table 1). Additionally, mutations were observed within the *rpoB* open reading frame but outside of the 81-basepair rifampin resistance-determining region (RRDR; Table 1). The V194I mutation observed outside of the RRDR in one strain is a unique substitution that is likely not associated with rifampin resistance. Five amino acid substitutions were noted in at least one strain beyond residue position 900 of the *rpoB* protein. There were seven strains with an *rpoB* mutation (6 at position 516 and 1 at position 526) where a wild type band was absent without a corresponding mutation band according to LPA. In six of these seven isolates, Ion Torrent sequencing revealed an uncommon amino acid substitution (i.e., glycine) within a known mutation site at position 516 where a valine (V) substitution (D516V) is typically known to occur (Table 2). Similarly, in one isolate Ion Torrent sequencing revealed an arginine (R) within a known mutation site at position 526 where tyrosine (Y) or aspartic acid (D) substitutions (H526Y/D) typically occur.

***katG* gene mutations.** Four amino acid substitutions were observed in the *katG* gene with S315T which is known to confer isoniazid resistance present in all resistant strains (Table 2). Clinical strains harboring R463L, W191R, and N138H mutations were detected by DST (Table 2) and have been previously characterized. A substitution at position 463 (R463L) in *katG* has been previously shown to have no effect on antibiotic resistance and can be used to categorize *M. tuberculosis* isolates into genetic Groups 1 (Arg463) or 2/3 (Leu463). Of 26 clinical isolates assessed, 7 (27%) were members of genetic Group 1 as evident by this R463L substitution.

***pncA* gene mutations.** Seven nucleotide mutations were noted in at least one strain among 561 bps comprising the full-length coding region for the *pncA* gene (Table 3). Nine of 26 strains (34.6%) contained an amino acid mutation conferring pyrazinamide resistance (Table 3). In one strain, a silent (synonymous) nucleotide mutation was characterized at nucleotide position 195 (C195T). Five strains contained previously

characterized amino acid substitutions (C14R, A102V, V139G, R154G, and L172P) known to confer resistance to pyrazinamide. A novel mutation, not previously reported elsewhere, encoding a termination *stop* codon was found in one isolate at residue 122 (Q122Stop) in the *pncA* protein (Table 3).

5 ***gyrA* gene mutations.** Nine unique mutations were observed in the 2,517 bp full-length *gyrA* gene encoding subunit A of the DNA gyrase enzyme. Resistance to fluoroquinolones (FQ) was only noted in strains harboring mutations in the quinolone resistance determining region (QRDR) defined by substitutions in *gyrA* at codons 88, 90, and 94. A number of additional mutations were also observed in regions outside of
10 the QRDR including two ‘mixed strain’ mutations at position 549 and 613 in the *gyrA* protein (Table 4). Mutation at position 95 (S95T) is known to have no effect on FQ resistance but can be used to categorize strains in genetic Groups 2 or 3. Of the 19 total clinical isolates belonging to genetic Groups 2/3, 18 (96%) were Group 2, and 1 (4%) was Group 3 according to assessment of *gyrA* position 95 (T = genetic Group 2, and S
15 = genetic Group 3).

***rrs* (16s) gene mutations.** Four nucleotide mutations were noted among the 1,540 bps comprising the full length 16s *rrs* gene. Seven of 26 (27%) clinical isolates were shown to be resistant to aminoglycosides by DST, and all strains harbored an A1401G mutation known to confer resistance (Table 5). Two other amino acid mutations
20 (C492T and A514C) were observed, but have been previously shown to not inhibit aminoglycoside efficacy. A previously uncharacterized G878A nucleotide mutation was observed, but the isolate was shown to be sensitive according to DST (Table 5).

Megabase Sequencing. Ion torrent gene chip sequencing was performed on the complete genome of Influenza virus A under five distinct conditions, identified in
25 Figure 9 as Tracks. Whole viral nucleic acid of Influenza A, strain H3N2 (about 14.4 kb total RNA) was prepared as discussed above and either reverse transcribed only, or reverse transcribed and PCR amplified as indicated in Figure 9. Influenza virus genome was mass amplified by reverse transcription (RT) and certain amplified cDNA populations subjected to PCR. Each result was then analyzed using the Ion Torrent
30 sequencing protocol. RT and/or RT-PCR analysis was performed with uniform hexamers, Uni 12, and/or 24 different influenza-specific primers (different in both

length and sequence). Uniform hexamers comprise a collection of primers, each six nucleotides in length whereby the collections contain all of sequence iterations of the six nucleotides. Uni 12 is primer that contains a sequence that is complimentary to 12 nucleotides at the 3' terminus of each of the segments of the influenza H3N2 viral genome (5'-ACGCGTGATCAGCAAAAGCAGG; SEQ ID NO 13). As shown in Figure 9, Track 4 amplification and sequencing with hexamer primers and Uni 12 followed by PCR amplification with the 24 influenza-specific primers and Ion Torrent protocol sequencing identified about 70% of the influenza genome.

Additional experiments were performed to achieve one-step sequencing of the complete Influenza genome. A series of influenza-specific primers were developed that would allow for uniform conditions to be performed for a PCR reaction. The primers that were developed are listed in Figure 10. These primers are all specific for the influenza virus genome with primer pairs spaced along the genome about every 800 to 1,000 bases in length (see Figure 10, amplicon length and start and stop positions for primer placement and sequence). All primers were of similar length, about 18-23 nucleotides and contain a similar GC content, about 22.7% to 38.9%, with nearly about 33% $\pm$ 6% and most about 33% $\pm$ 3%. PCR analysis was performed using different collections of these primers and the amplified products identified using the Ion Torrent sequencing protocol.

Sequencing of pncA gene. The gene sequence of pncA was determined using a series of primers spaced or "tiled" along the pncA gene in accordance with the invention. The coding sequence of the pncA gene is depicted in Figure 11A and the primers utilized are depicted in Figure 11B in bold and underlined. Using these primers in conjunction with Ion Torrent methodology, the entire coding regions of pncA was determined (see P1-P4 of Figure 11B). Expanding the primers utilized to all genes or of specific regions allows for one-step sequencing of the entire genome.

Other embodiments and uses of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein. All references cited herein, including all publications, U.S. and foreign patents and patent applications, are specifically and entirely incorporated by reference. The term comprising, where ever used, is intended to include the terms consisting and

consisting essentially of. Furthermore, the terms comprising, including, and containing are not intended to be limiting. It is intended that the specification and examples be considered exemplary only with the true scope and spirit of the invention indicated by the following claims.

Claims

1. A method of identifying a nucleic acid sequence motif in a microorganism comprising:

providing multiple nucleic acid samples each sample obtained from a different
5 strain or serotype of the microorganism;

amplifying the sequences of the multiple nucleic acid samples by a polymerase chain reaction;

obtaining sequence information of the amplified sequences by ion torrent sequencing;

10 determining a polymorphism in the genome of at least one microorganism strain or serotype from the sequence information obtained; and

correlating the polymorphism identified with a phenotype or genome location of the at least one microorganism strain or serotype to identify the nucleic acid sequence motif.

15 2. The method of claim 1, wherein the microorganism is a virus, a bacterium, a fungus or a parasite.

3. The method of claim 1, wherein the virus is an influenza virus.

4. The method of claim 1, wherein the bacterium is a *Mycobacterium tuberculosis*.

5. The method of claim 1, wherein the nucleic acid samples are provided in an
20 aqueous molecular transport medium.

6. The method of claim 5, where the aqueous molecular transport medium contains a chaotrope, a detergent, a reducing agent, a chelator, a buffer, and an alcohol, together present in an amount sufficient to lyse cells, denature proteins, inactivate nucleases, kill pathogens, and not degrade nucleic acid.

25 7. The method of claim 1, wherein the amplifying is performed in a one-step polymerase chain reaction.

8. The method of claim 1, wherein the nucleic acid motif correlates with resistance of the microorganism to an antimicrobial compound.

9. The method of claim 8, wherein the antimicrobial compound is an antibiotic.

30 10. The method of claim 1, wherein the polymerase chain reaction is carried out in an aqueous mix comprising: a heat-stable polymerase; a mix of deoxynucleotide tri

phosphates comprising about equivalent amounts of dATP, dCTP, dGTP and dTTP, a chelating agent, an osmolarity agent, an albumin, a magnesium salt; and a buffer.

11. The method of treating a disease or disorder caused by the at least one microorganism strain or serotype with the antimicrobial compound identified by the method of claim 1.

12. A method for determining a complete sequence of a target gene or genome of a microorganism comprising:

producing a series of amplicons by performing a single polymerase chain reaction (PCR) of the target in an aqueous mixture containing a heat-stable polymerase; a mix of deoxynucleotide tri phosphates comprising about equivalent amounts of dATP, dCTP, dGTP and dTTP; a chelating agent; a salt; a buffer; a stabilizing agent; and a plurality of primer pairs wherein each primer of the plurality of primer pairs has a similar annealing temperature;

sequencing each of the series of amplicons produced by semi-conductor sequencing, and

correlating the sequences of the amplicons and constructing the complete sequence of the target.

13. The method of claim 12, wherein each of the primers of the multiple primer pairs comprise primers that are from 15 to 25 nucleic acids in length and has a GC content of about 25%-50%.

14. The method of claim 12, wherein each primer pair is designed to PCR amplify an amplicon, and the collection of amplicons that are PCR amplified encompass overlapping segment of the complete sequence of the target.

15. The method of claim 12, wherein the plurality of primer pairs hybridize to the target and are spaced along the target at about every 500 to 2,000 nucleotides.

16. The method of claim 12, wherein the microorganism is a virus, a bacterium, a fungus, a parasite or a cell.

17. The method of claim 12, wherein the virus is an influenza virus.

18. The method of claim 12, wherein the bacterium is *Mycobacterium tuberculosis*.

19. A method for determining the sequence of a nucleic acid target in one step comprising:

performing a polymerase chain reaction on the nucleic acid target to produce a series of amplicons, wherein the reaction comprises: a heat-stable polymerase; a mix of deoxynucleotide tri phosphates comprising about equivalent amounts of dATP, dCTP, dGTP and dTTP; a chelating agent; a salt; a buffer; a stabilizing agent; and a plurality of primer pairs wherein each primer of the plurality of primer pairs has an annealing temperature within 5°C;

sequencing each of the series of amplicons produced by semi-conductor sequencing, and

correlating the sequences of the amplicons and constructing the sequence of the nucleic acid target.

20. The method of claim 19, wherein the nucleic acid target is greater than 1 Mb in length.

21. The method of claim 19, wherein each of the primers of the multiple primer pairs is from 16 to 24 nucleotides in length, has a GC content of about 28-35%, and an annealing temperature of within 3°C of each other primer of the multiple primer pairs.

22. The method of claim 19, wherein each primer pair is designed to PCR amplify an amplicon representing a portion of the sequence of the nucleic acid target, and the collection of amplicons that are PCR amplified represent overlapping portions of the complete sequence of the target.

23. The method of claim 19, wherein the plurality of primer pairs hybridizes to the target at a spacing of about 800 to 1,500 nucleotides in length.

24. A mixture comprising multiple pairs of nucleic acid primers wherein, upon subjecting the mixture to a polymerase chain reaction in association with a nucleic acid target, the nucleic acid primers hybridize to the nucleic acid target at about the same temperature and generates a collection of amplicons, wherein each amplicon of the collection is about 500 to 2,000 nucleotides in length, such that the entire sequence of the target is represented in the resulting collection of amplicons.

25. The mixture of claim 24, wherein each primer of the collection of primer pairs is about 15 to 25 nucleotides in length, has a GC content of about 25%-45%, and an annealing temperature to the target within 3°C of each other primer.

26. The mixture of claim 24, wherein the target is a gene or genome of a microorganism.

27. The mixture of claim 26, wherein the gene is a gene that confers antibiotic resistance of the microorganism.

5 28. The mixture of claim 26, wherein the microorganism is a virus, a bacterium, a parasite, or a fungus.

29. The mixture of claim 24, which contains a heat-stable polymerase; a mix of deoxynucleotide tri phosphates comprising about equivalent amounts of dATP, dCTP, dGTP and dTTP; a chelating agent; a salt; a buffer; a stabilizing agent and nuclease-
10 free water.

30. The mixture of claim 24, wherein the nucleic acid primers do not self-hybridize.

pncA gene + 100 flanking base pairs (1161 nt)

AGGCGGGAATGAACACCGTCAAGCCGAGTCCATCGGGAACCTGAGTTCCAGATCGCGCAGCAACACCGTGCAGAGACGATATCCAGATCGCGATGGAAGTGAATATCCCGGGCCCGGATGAAGGTGTCGTAGAAGCGG
CCGATGGCTCATGCCACCTGGGCTGGGAACCAACCGGCTTCACCGCGGCTACCGGTTGAACACCGACCGGCGGCTGCTGGCGGGCGCGCTTCGGCGAGCGGTCCACCGCGCCACAGTTC
ATCCGGTTCGGGCTGCCATCAGGAGCTGCAACCAACTCGACCTGGGCTGGCATCTCTCCAGCGCGGACCGTGTATCGCGGACACACCGCTGTCAAGTCCACAGACCTTGTGGCCAGCCATTGG
TACCGCTCTCGGCGCTCTGGGCAACAATGATCGGTGGCATACCGACCATATCGACCTCATCGACCGCGCTTCGGCGAGCAATTCAGCAGTGGGTGCGCTTCGTGCACTCCTTCGAAGCGCTGTACGCTC
CGGTGAGCAACCTTGTAGAACACCGCTCGATTGCGCAGCTGTCAGACTGGGATGGGAAGTCCGCGCGCGGAGTACCGTACGCAATTCGGGTGGCCACGACGAGGAATAGTCCGGTGTCCCGGAGAACTGGTCAACCC
GGTCAATGTGAAGTCTTGGTTGCCACGAGCTGATGGTAGTTCGCGCGCTTCGCCAGGTATGTCGCTATGGCGCGGCGCACCGGTTACCGCAGCGAGCCACCTTCGAGAACTGCTTCGACCGTC
GAAGATGATCAACCGCGCATAGTCCACATAGCTTCGGGCACTGCCCGGCACTGCCCGGCACTTCGCTACCGACGCGGACGACAGATAGGTTCCATAGCGCGGACGATCGCGAACAATGACAGCTAGCGGGCGGCCAAC
CAACCGGGGATAGATACCGCGGTGATGTAGTAGCGGACAAATCCGTCCGGTACAGAGGTGTCATCGCGGCTTGGTGGAGCCACGGCTCCACCGGTCAAGCGGCACTCGACCGGCTTAGCGGCGATGCCGAT
CCCCATATACCGCGGGAACATGCAGCCACATCGTCCGTC (SEQ ID NO: 14)

REVERSE COMPLEMENT

GACGACAGTGGGTGATGTTCCGGCGGTGATATGGGGATCGGCATCGCCGTAAGCGGTGCACTGCCGCTGACCTGGGTGGAGCGGTGGGCTCGAACAGGCGCGGATGACACCTCTGTCAACCGTACGGATTTT
GTCTCTACTATACCAACCGGCTGATCTATCCCGCGTTCGGGTGGGATCGCTGCTGCTCATGTTCCGATCTGTCGGCTCATGTCGACCTATATCTGCGCTTCCGCTGCGGTGCGCATACCGTACCGGAGTTCG
CCCCAGCTATGTTGACGTTATCGGGCTTATCTCTGAGCTGCAGAACGACTTCTGGAGGTTGGTCTGCTGGCGGTAAACGGTGGCGCGGCTGCGCGGCTAGGCAAACTGCCCGGCGAGTGC
TACCATAGTCTGGGCAACCAAGGACTTCCACATCGACCGGCTGACCACTTCTCGGCAACACGGACTATCTCTGCTGTGGCAACCGCATTCGCTACCGGTACTCCCGGCGGCACTTCCATCCAGTCTGGACAC
GTGCGCAATCAGCGGCTGTTCAAGGTGCCCTACACCGGCTGACAGCGGCTTCGAAGGAGTCCAGCAACCGGACCGCACTGCTGAATTCGCTGCGGCAACCGGCTGCGATGAGGTTCGATGTTGTCGGTATTG
CCACCGATCATGTTGTCGCGCAGACCGCGGAGGACCGGTAACGAAATGGCTTGGCACACGGGTCTGTCGCGCGATACCAACCGTCCCGGTGGAGGAGATCGGCAACCGGACCGCGGCTGCGGACCGCGGCTGCGGATCG
GAGTTGTTGCACTCTGATGGCACCGCGAAACGGGATGCACTGTTGCGGCTGAGGCTCGCGCAAGCGGCTCGCGCAAGCGGCTGGGTGGGTTGTTACCGGTGACCGCGGCTCGAAGAC
CCGGTGGTTCGACGCGAGTGGGCTAGGCGCATCGGCGCTTCTACGACACCTTCATCGGCGCGGGGATATCAAGTTCCATCGGCACTCTCGGCTACGGTGGTGGCGGCACTCGGACTCGA
GGTCCGATGACTCTGGCTGACGGTGTTCATTCGCGCT (SEQ ID NO: 15)

PROTEIN SEQUENCE

MRALIIVDVFCEGSLAVTGGAALARAISDYLAFAADYHHWATKDFHIDPDHFSGTPDYSSWPPHCVSGTPCADFHPSLDTSALAEAVFYKGAVTGAYSGFEGVDENGTPILNWLRLQRGVDEVVDVGTATDHCVR
QTAEDAVRNLATRIVLDTAGVSADTIIVAALHEMRASVELVCSS (SEQ ID NO: 16)

Primer Length:

Forward 1 (20 nt) : TCATCGACCCATATCTCTG (SEQ ID NO: 17)
Reverse 1 (20 nt) : ATGAACTGTGGCGCGCGTG (SEQ ID NO: 18)
Amplicon Length: 675 nt

Forward 2 (20 nt) : ACGATTTGTCGCTCATTAC (SEQ ID NO: 19)
Reverse 2 (20 nt) : ATCTGGATATGCTCTCCGC (SEQ ID NO: 20)
Amplicon Length: 959 nt

FIG. 1

From H37RV Gene strain

GGCCATGCTCTGATGACCCCGTTGTCGGGGGGCTGGCCCGTTCTTTTGTCTCAGGATATTTCTATAAATACCTTTGGCTCCCTTTCCAAAGGACGTGTTGGCTTTTGTGAGAGTTTGATCCTGGCTC
AGCAGAACGCTGGCGCGTCTTTAATACATGCAAGTCGACGGAAGGTCTCTTCGAGATACTCGAGTCGCAACGGGTGAGTAACACGTTGGTGTATCTCCCTGCACCTTCGGGATAAGCCTGGGA
AACTGGGTCTAATACCGGATAGGACACGGGATGCAATGCTTGTGGTGGAAAGCGCTTAGCGGTGTGGGATGAGCCCGCGGCTATCAGCTTGTGGTGGGTGACGGCTACCAAGCCGACGACGG
GTAGCCGGCTCAGAGGTGTCCGGCCACACTGGGACTGAGATACGGCCAGACTCCTACGGAGGCGAGTGGGGAATATTGCACAATGGCGCAAGCCTGATGCAGCGACGCCGCTGGGGATG
ACGGCTTCGGGTGTAAACCTCTTACCATTCGACGAAGTCCGGTTCTCTCGGATTGACGGTAGGTGGAGAAAGCAACCGCCCAACTACGTCCACGAGCCGCGTAAATACGTAGGTGCGAGC
GTTGTCGGGAATTACTGGCGTAAGAGCTCGTAGGTGTTTGTGCGGTTGTTCTGCAAACTCTACGGCTTAACCTGTGAGCGTCGGGCGATACGGGCAGACTAGAGTACTGCAGGGAGACTGGAAT
TCCTGGTGTACCGGTGGAATCGGCAGATATCAGGAGGAACACCGTGGCGAAGCGGTCTCTGGCAGTAACCTGACGCTGAGGAGCAAGCGTGGGAGCAACAGGATTAGATACCTTGGTAGTC
CAGCCGTAACCGTGGGTACTAGGTGGGTTTCTTCTTGGGATCCGTGCGGTAGCTAACGCAATTAAGTACCCCGCTGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGCC
CGCACAAGCGGGGAGCATGTGGATTAAATTCGATGCAACGGGAAGAACCTTACCTGGGTTTGACATGCCACAGGACGCGTCTAGAGATAGGCGTTCCCTTGTGGCCTGTGTGCAGGTGGTGCATGGCTG
TCGTCAAGTCTGTCTGTGAGATGTTGGTTAAGTCCCGCAACGAGCGCAACCTTGTCTCATGTTGCCAGCAGCTAAATGGTGGGACTCGTGAGAGACTGCCGGGTCAACTCGGAGGAAGGTGGGA
TGACGTCAAGTCATCATGCCCTTATGTCAGGGCTTCACATGCTACAATGGCGGTACAAAGGCTGGCATGCCGAGGTTAAGCGAATCCTTAAAGCCGCTCTCAGTTCGGATCGGGTCTG
CAACTCGACCCGTGAAGTCGGAGTCGCTAGTAAATCGCAGATCAGCAACGCTGGGTGAATACGTTCCCGGGCTTGACACACCCCGCTCAGTCAATGAAGTCGGTACACCCGAGCCAGTGGC
CTAACCTTCGGCAGGAGCTGTGCAAGCTGGGATCGGCGATTGGCACGAAGTCGTAAACAGGTAGCCGTACCGGAAGGTGCGGCTGATCACCCTCCTTTCTAAGGAGCAACGACGAAACGCCCCAACT
GGTGGGGCGTAGCCCGTGAGGGGTTCTTCTGTAATGGGGGAGAGCGGGGTGCATGACAAACAAGTTGGCCA (SEQ ID NO: 21)

TB 16S ribosomal RNA gene sequencing primers

PAIR 1:

For 16s Ver1: 5'-TGGCCGTTGTTTGTCTCAGGAT-3' (SEQ ID NO: 22)

Rev16s Ver1: 5'-TACAGACAAGACCCCTCACGG-3' (SEQ ID NO: 23)

Expected amplicon size: 1699 bp

PAIR 2:

For 16s Ver2: 5'-TTCTAATATACCTTTGGCTCCCT-3' (SEQ ID NO: 11)

Rev16s Ver2: 5'-TGGCCAACTTTGTTGTCAATGCA-3' (SEQ ID NO: 12)

Expected amplicon size: 1680 bp

FIG. 2

(SEQ ID NO: 27) TCGCCGCGATCAAGGAGT = Reverse Primer from Article EID (Emerging Infectious Diseases) - 2001

(SEQ ID NO: 28) **TCGAGGTCCCGCAGCTCCA** = Forward Primer from Article EID - 2001

TTCG = codon 531 Serine to Leucine mutation

CAC = codon 526 Histidine to Asparagine mutation

TCG = codon 522 Serine to Leucine mutation

GAC = codon 516 Aspartic acid to Valine

CAA = codon 513 Glutamine to Lysine

CTG = codon 533 Leucine to Proline

ATC = codon 572 Isoleucine to Phenylalanine

New Forward Primer

TB.rpoB. Forward Primer3: ~~AAATCTGCTCCGCTTCCACCA~~ (SEQ ID NO: 29)

With TB.rpoB.Reverse Primer is 620 bps

FIG. 3
Continue

GenBank: CP000611.1

GyrA Gene

[illegible]

FIG. 4

Primer Pair #1:

TB.GyrA. Forward1: ~~CTAAGCAACCTGCGTCAAT~~ (22 bases) (SEQ ID NO: 31)
TB.GyrA. Reverse1: ~~ATTCCCTCAAGATCGTAGC~~ (21 bases) (SEQ ID NO: 32)

Amplicon: 2605 bp

Primer Pair #2:

TB.GyrA. Forward2: ~~AAGGATGTTGGTTCCCTGAT~~ (21 bases) (SEQ ID NO: 9)
TB.GyrA. Reverse2: ~~TAACTCGATACCCGGT~~ (18 bases) (SEQ ID NO: 10)

Amplicon: 2703 bp

FQ resistance is at codons 94 (Green) and perhaps 90.

TB.GyrA. Reverse3: ~~CGTCGTAGTTAGCGATGAATC~~ (SEQ ID NO: 33)

FIG. 4
Continue

catalase-peroxidase-peroxynitritase T katG

TB.KatG-Reversel: ATGCACGGCGGGTCTGACAAAT (SEQ ID NO: 36)

Amplicon Size: 2538 bp

5' ~~TG~~ ~~KatG-Forward:~~ ~~ACACCAAC~~ ~~TCCTGGGAGGAAT~~ (SEQ ID NO: 37) or ~~ACACCAACTCCTGGGAGGAAT~~ (SEQ ID NO: 5)
3' ~~TG~~ ~~KatG-Reverse:~~ ~~GCATGATGACGACATTT~~ (SEQ ID NO: 6)

Amplicon Size: 2447 bp

The mutation is at Codon 315.

IN Wildtype it is Serine (S)

IN Beijing it is Gycine (G)

In others it is a Threonine (T)

~~New primers,~~

5' ~~TG~~ ~~KatG-Forward:~~ ~~TTCACGCGGTACGCGGATCT~~ (SEQ ID NO: 38)
3' ~~TG~~ ~~KatG-Reverse:~~ ~~TTCGCGAATCTCTCGGCAAT~~ (SEQ ID NO: 39)
~~Amplicon size: 603 bp~~

FIG. 5
Continue

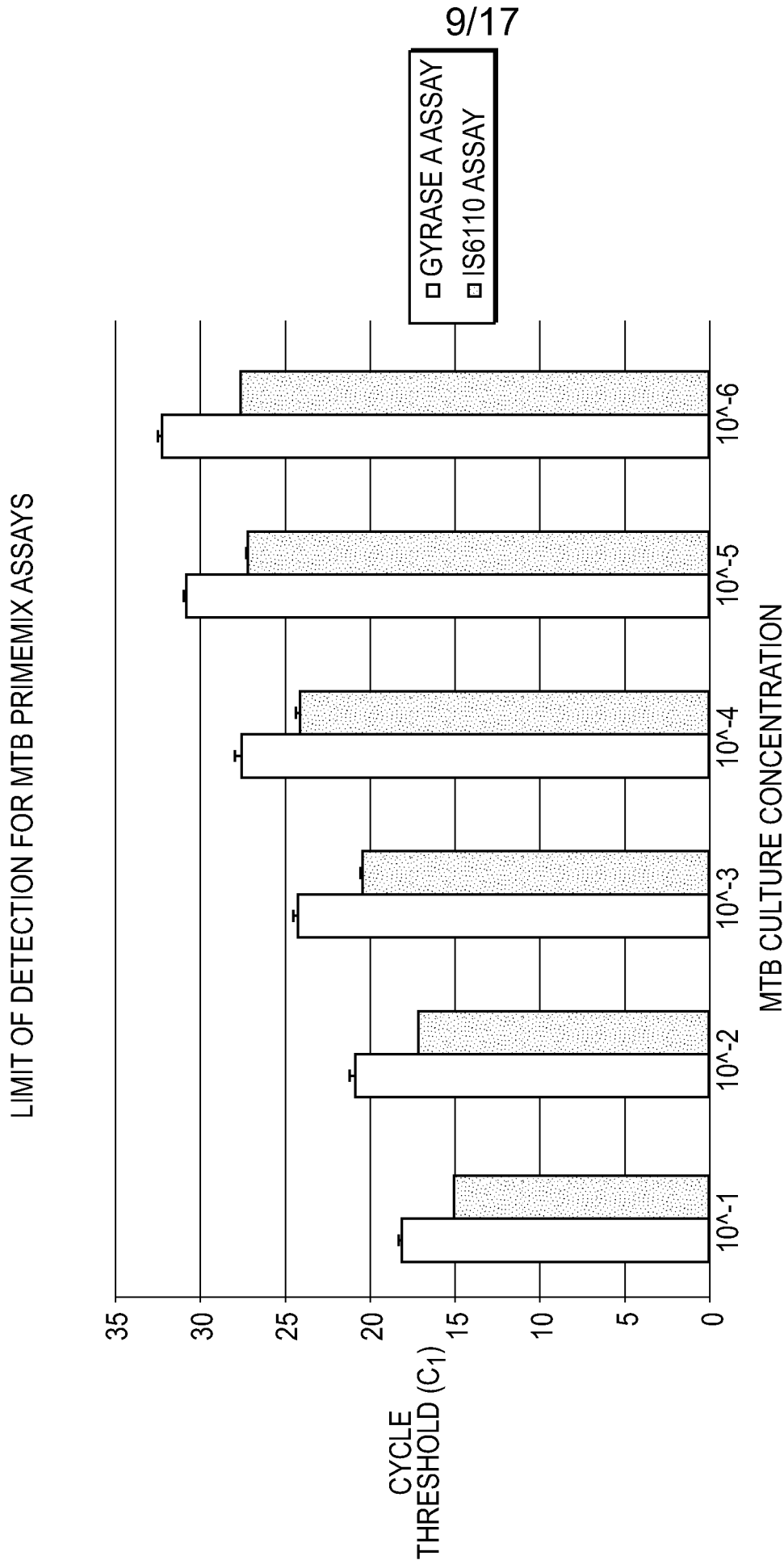


FIG. 6

10/17

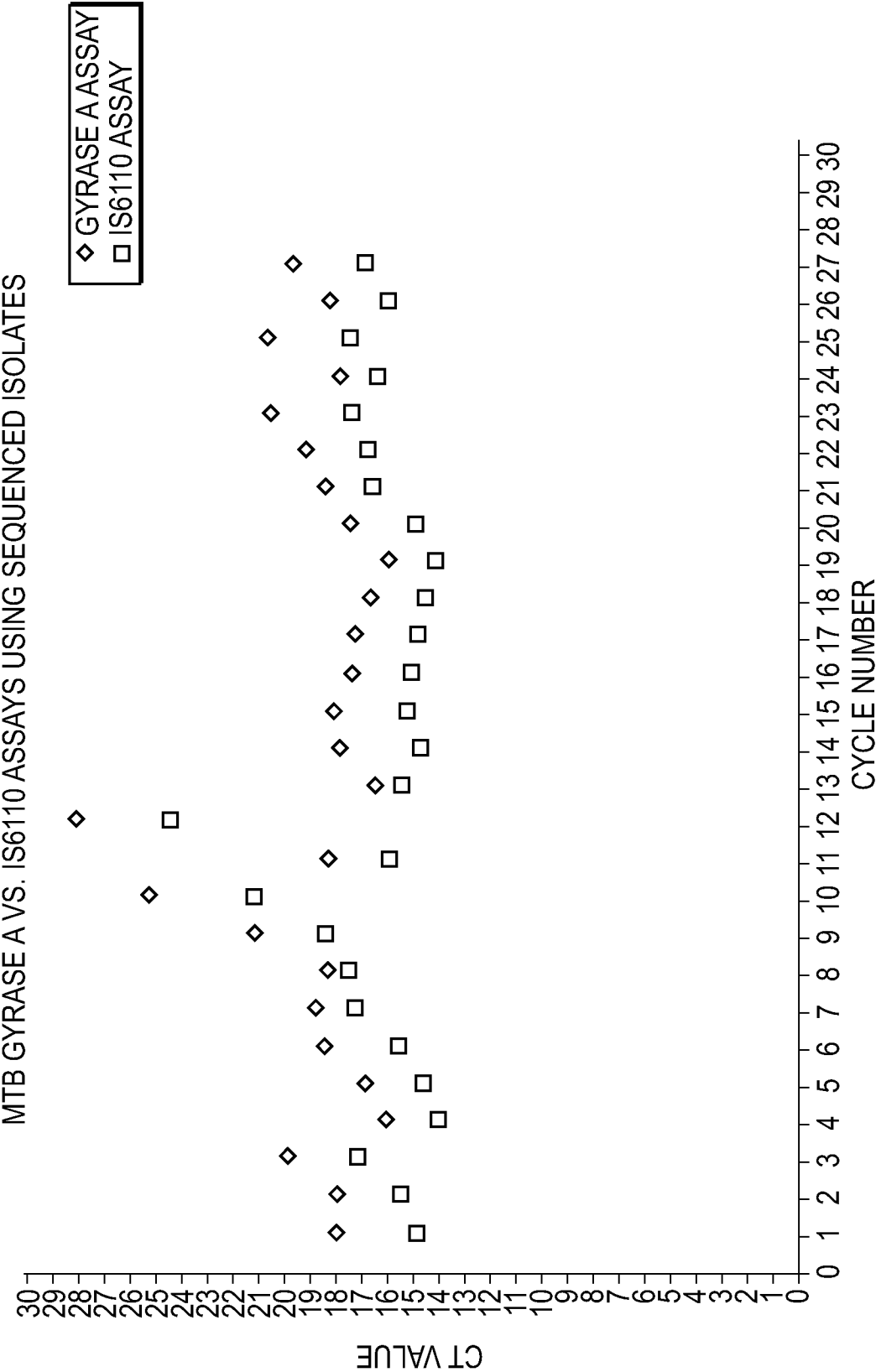


FIG. 7

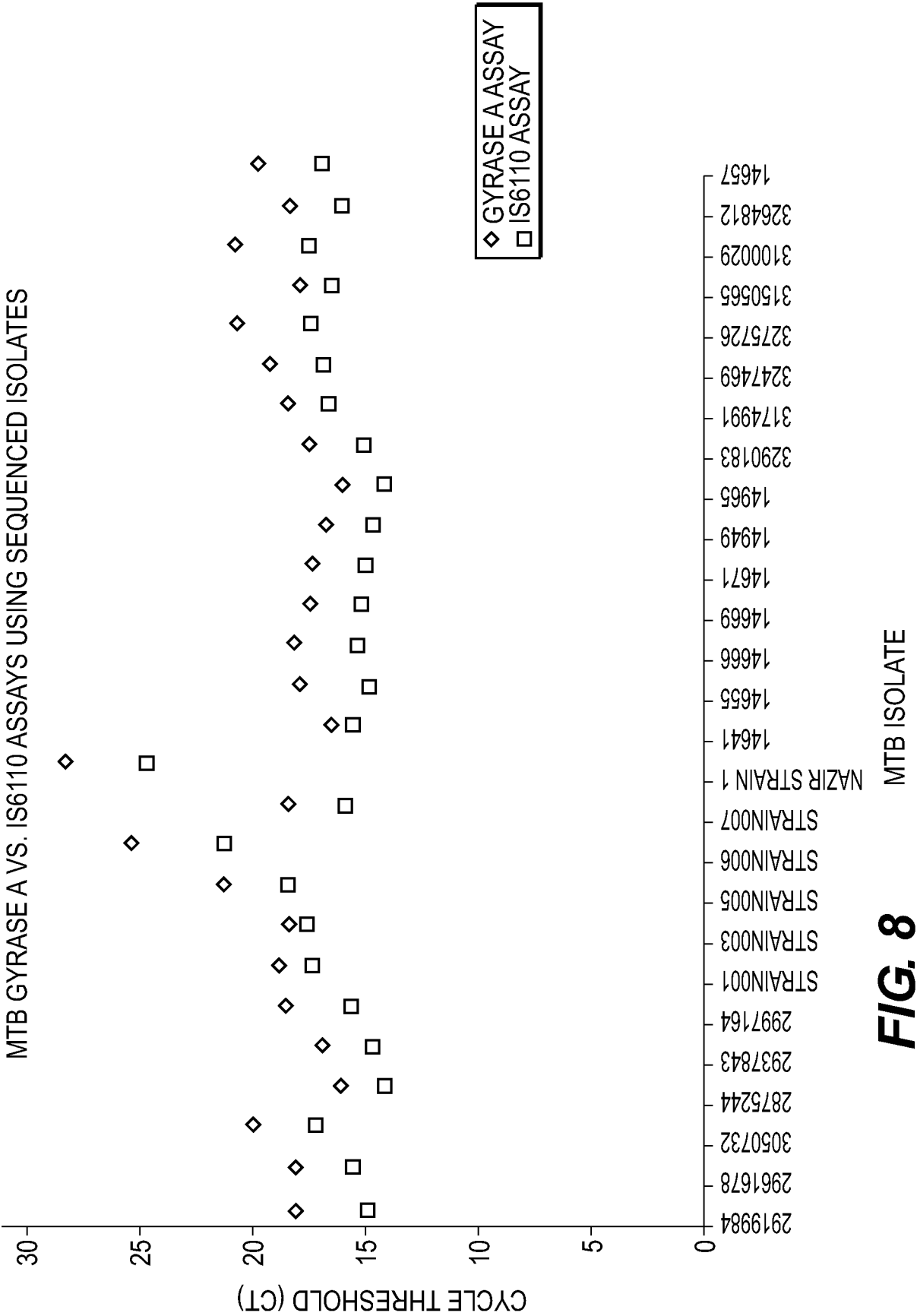


FIG. 8

12/17

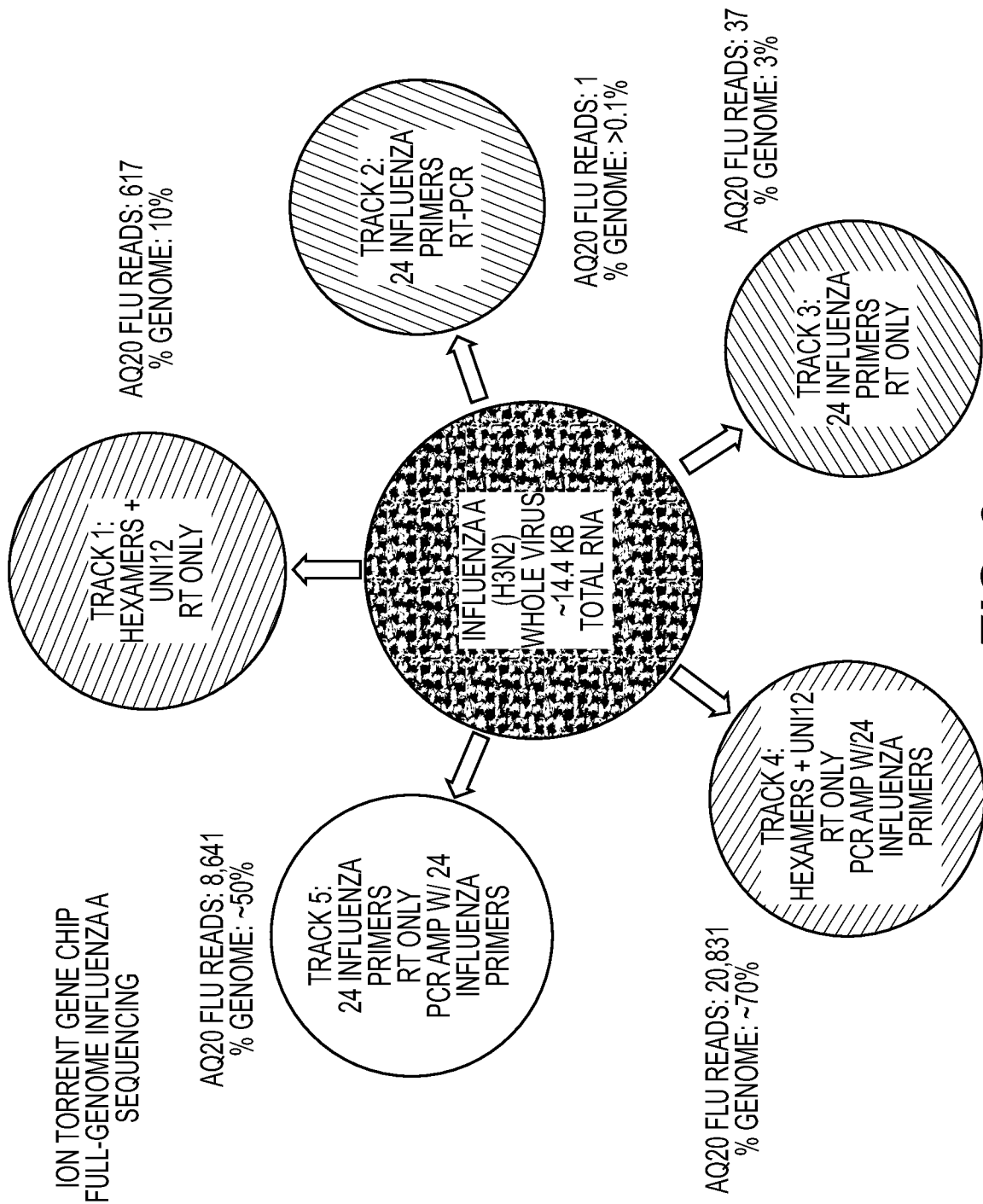


FIG. 9

13/17

INFLUENZAA (H3N2) MASTER PRIMER LIST FOR ION TORRENT

PB2 PRIMER SETS:		SEQ ID NO:	LENGTH (bp):	MELTING TEMP (°C):	G/C %	AMPL LENGTH (bp)	START POSITION:	END POSITION:
FORWARD: ATT ATA TTC AGT AIG GAA AGA A		40	22	44.5	22.7	943	1	
REVERSE: ATA TAT CCA CAG CTT GTT C		41	19	46.6	36.8			943
FORWARD: GAT CCA CTA GCA TCT TTA TT		42	20	46.9	35	997	835	
REVERSE: GTA CGT CTC TCA TTT GTT		43	18	46.5	38.9			1831
FORWARD: AAG GCA TTT TCA GAA AGA T		44	19	47	31.6	998	1336	
REVERSE: GTC GTT TTT AAA CTA TTC AGC		45	21	47.7	33.3			2314
		AVERAGE:		46.5		979		
PB1 PRIMER SETS:								
FORWARD: ACC ATT TGA ATG GAT GTC		46	18	47.2	38.9	996	1	
REVERSE: TTG ATT CTT TGT GAT GTA TGT		47	21	47.5	28.6			996
FORWARD: AAT GAT GAC TAA TTC ACA AG		48	20	45.3	30	997	997	
REVERSE: ATA ATT CTC ATC CAT CAG C		49	19	46.7	36.8			1869
FORWARD: AAA TAC ACC AAG ACA ACA TA		50	20	46.6	30	1009	1316	
REVERSE: CAT GAA GGA CAA GCT AAA T		51	19	47.4	36.8			2311
FORWARD: ATTTTGAATGGATGTCAATC		52	20	46.2	30			
REVERSE: GTGATTGTGAAAGAAAGCT		53	19	48.1	36.8	913	3	916
		AVERAGE:		46.8		1001		
PA PRIMER SETS:								
FORWARD: CTG ATT CGA AAT GGA AGA		54	18	46.6	38.9	989	1	
REVERSE: CGT GTG GTT TGA CTA TAT		55	18	46.3	38.9			989
FORWARD: GAT CAA GTG CAT AAA AAC AT		56	20			1001	931	
REVERSE: ATA GAG TCC TAC AGA CTT T		57	19	46.5	36.8			1931
FORWARD: TGA CGA ACC TGA ATT AAG		58	18	46.5	38.9	998	1212	
REVERSE: GTA CGG ATA ACA AAT AGT AG		59	20	44.9	35			2192
		AVERAGE:		46.2		996		

FIG. 10

14/17

HA PRIMER SETS:									
FORWARD: TAA TTC TAT TAA CCA TGA AGA C	60	22	45.5	27.3	887	1			
REVERSE: GCA TCT GAT CTC ATT ATT G	61	19	45.5	36.8					887
FORWARD: CAT ACT TTT GAT TAA CAG CA	62	20	45.8	30	940	802			
REVERSE: TTAATG CAC TCAAT GCA	63	18	47.1	33.3					1722
AVERAGE:			46.0		914				
NP PRIMER SETS:									
FORWARD: AAT AAT CAC TCA CTG AGT G	64	19	46.5	36.8	803	1			
REVERSE: AAT ATG AGA TCT TCG ATC TC	65	20	46.1	35					803
FORWARD: ACA AGA AGT GCT TAT GAG	66	18	46.5	38.9	859	690			
REVERSE: TTC CTT AAT TGT CGT ACT C	67	19	46.4	36.8					1537
FORWARD: CTG AGT GAC ATCAAAATCA	68	19	47.3	36.8	719	13			
REVERSE: GCAGCTGTTGAAATTTTC	69	19	48.1	36.8		731			
AVERAGE:			46.4		831				
NA PRIMER SETS:									
FORWARD: AAG ATG AAT CCA AAC CAA	70	18	46.1	33.3	758	1			
REVERSE: GTA TCA GCT TTT CCT GAA	71	18	46.5	38.9					758
FORWARD: GAT AGT GTT GTT TCA TGG	72	18	45.3	38.9	790	657			
REVERSE: CTAAA TTG CGA AAG CTT ATA	73	21	46.5	28.6					1429
AVERAGE:			46.1		774				
M1/M2 PRIMER SETS:									
FORWARD: ATA TTG AAA GAT GAG CCT T	74	19	46	31.6	999	1			
REVERSE: TAG TTT TTT ACT CCA ACT CTA	75	21	46	28.6					999
AVERAGE:			46		999				
NS1/NS2 PRIMER SETS:									
FORWARD: ACA AAG ACA TAA TGG ATT CT	76	20	46.5	30	803	1			
REVERSE: GGT GTT TTT TAT CAT CAA ATA AG	77	23	46.4	26.1					863
AVERAGE:			46.5		863				

FIG. 10
Continue

GACGGATTGTGCGCTCAC = pncA.P1-Fwd (SEQ ID NO: 79)
AGCCACCCTCGCAGAA = pncA.P1-Rev (SEQ ID NO: 80)

CATCGTCGACGTGCAGAA = pncA.P2-Fwd (SEQ ID NO: 81)
TGTCAGACTGGGATGGAA = pncA.P2-Rev (SEQ ID NO: 82)

ATTGCGTCAGCGGTACT = pncA.P3-Fwd (SEQ ID NO: 83)
TGGCCAAAGCCATTGCGTA = pncA.P3-Rev (SEQ ID NO: 84)

ATCATTTGTGTGCGCCAGA = pncA.P4-Fwd (SEQ ID NO: 85)
CAACAGTTCATCCCGGTT = pncA.V2.P4-Rev (SEQ ID NO: 86)

PncA.P1: 221 bp

GACGGATTGTGCGCTCACTACATCACCGGCGTGATCTATCCCGCCGGTTGCGTGGCCGCCGCTCAGCTGGTCAATGTTCCCGATCGTCCGCGGCTCATGGACCCATATCTGTGGCTGCCGCGTCCGTA
GGCAAACTGCCCGGCGAGTCGCCCGAACGTA**TGGTGGACGTA**TGGGGGGCTTGCAT**CA**TCCGTCCGACCTGCCAGAA****~~CCGACATCTGCGACCGCTGGTA~~ (SEQ ID NO: 87)

PncA.P2: 245 bp

FIG. 11B

(SEQ ID NO: 88)

CATCGTCGACGTCCAGAAAGCACTTCTCCACGCTCZCTCCCTCCGCCGTACGGCTCCGCCGCCGCCGCCGCCCATCAGGCAGCTACCTGGCCGTAACCTACCATCACCA
AGGACTTCCGATCGAACCCGGTGACGACATCTCCGGACATTCGTCTGCTGGCCACGGCATTCGGTTCACGGGTACTCCCGCGCGCGGCTCCATCCGACATGAACA

PncA.P3: 246 bp

[illegible]

PncA.P4: 184 bp

ATCATTTGTGTGGCCACAGCGGGCGGACGACCGGCTTGTGCTGGAGCCTTCACAGCGGGTGTGTGGCCGCGCTGGAGCGAGATGGCGACCGCC
AGGCGGAGTTGGTTTTCAGCTCTGATGGCACCGCGGATGCGAGCAKCTG (SEQ ID NO: 90)

Figure 11B

FIG. 11B
Continue

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2013/040302

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12Q 1/68 (2013.01)

USPC - 435/6.12

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8) - C12Q 1/68; C40B 20/00, 30/04, 40/06, 50/06; G01N 21/00, 21/64 (2013.01)

USPC - 435/6.1, 6.11, 6.12; 506/2, 9, 16, 26

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

CPC - C12N 15/10; C12Q 1/68, 1/6851; G01N 21/00, 21/64 (2013.01)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PatBase, Google Patents, PubMed

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	WO 2012/037531 A1 (POLLNER et al) 22 March 2012 (22.03.2012) entire document	1-3, 5, 8, 9, 11 ----- 4, 6, 7, 10, 17
X -- Y	US 2012/0034685 A1 (SPARKS et al) 09 February 2012 (09.02.2012) entire document	12-14, 16, 19, 21, 22 ----- 15, 17, 18, 20, 23
X -- Y	WO 2003/080870 A1 (WU et al) 02 October 2003 (02.10.2003) entire document	24-26, 28-30 ----- 27
Y	US 6,124,098 A (HEYM et al) 26 September 2000 (26.09.2000) entire document	4, 18
Y	WO 2006/110855 A2 (LEAMON et al) 19 October 2006 (19.10.2006) entire document	6, 7, 10, 15, 20, 23
Y	HOWDEN et al. "Evolution of Multidrug Resistance during Staphylococcus aureus Infection Involves Mutation of the Essential Two Component Regulator WalkR," PLoS Pathogens, 10 November 2011 (10.11.2011), Vol. 7, No. 11, e1002359, Pgs. 1-15. entire document	27
P, X	DAUM et al. "Next-Generation Ion Torrent Sequencing of Drug Resistance Mutations in Mycobacterium tuberculosis Strains," Journal of Clinical Microbiology, 12 September 2012 (12.09.2012), Vol. 50, No. 12, Pgs. 3831-3837. entire document	1-30

☐ Further documents are listed in the continuation of Box C.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

23 September 2013

Date of mailing of the international search report

04 OCT 2013

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774