Reprinted from *American Biotechnology Laboratory* December 2003

DNA Sequence-Based High-Throughput Microbial Identification Using Multiplex Sequencing

by Thuraiayah Vinayagamoorthy, Kirk Mulatz, and Roger Hodkinson

There is an odd dichotomy in the world of routine medical diagnostics. On the one hand, clinical chemistry and hematology test panels can be run cost-effectively on automated, computer-controlled instrumentation with extreme accuracy and same-day reporting, while the other major disciplines of microbiology and cytology remain relatively expensive, time-consuming manual methods with suboptimal accuracy. Routine microbiology is mainly bacteriology with very little virology, and the delayed reporting means that results are of limited value in timely clinical decision-making. This is especially so given the new importance of timely and accurate microbiological test results in preventing potential pandemics (e.g., SARS), assessing emerging zoonoses (e.g., West Nile virus), preventing large-scale outbreaks of diseases caused by contaminated food, and understanding the pathophysiology of cancer and atherosclerosis so that appropriate treatment modalities can be developed. The tests that are designed for these purposes have to demonstrate the essential criteria of accuracy, sensitivity, quick turnaround time, and delivery at a competitive price.

There are two standard methods for microbial identification: 1) the conventional culture method with biochemical and/or immunological (ELISA) confirmation, and 2) the DNA sequence method performed by the standard chain termination sequencing method.[1] Although in this postgenomic era, DNA sequence-based identification is accepted as the "gold standard," unfortunately, DNA sequencing as presently performed can detect only one pathogen per test. This is inadequate for routine microbiological testing, where ideally any one sample should be screened for multiple pathogens from the multiple phyla that commonly cause the particular presenting syndrome. Hence, the conventional culture method continues to be the general method in use for microbial identification.

Other DNA technologies that are being examined as alternatives to the two standard methods include methods based on target amplification (polymerase chain reaction [PCR, **Roche Diagnostics,** Basel, Switzerland],[2] ligase chain reaction [LCR],[3] and transcription-mediated amplification [TMA, **GeneProbe,** San Diego, CA])[4] and those based on signal amplification (i.e., b-DNA [**Bayer Diagnostics,** Tarrytown, NY],[5] hybrid capture,[6] rolling circle,[7] Invader [**Third Wave,** Madison, WI],[8] Pyrosequencing [**Pyrosequencing AB,** Uppsala, Sweden],[9] and isothermal ramification amplification[10]).

In order to adapt standard sequencing technology for routine testing, the following two approaches were developed:

1. A modification to DNA sequencing that allows the simultaneous identification of multiple pathogens in any one sample. This platform technology, called MultiGEN (**Bio-ID Diagnostic Inc.,** Saskatoon, Saskatchewan, Canada), consists of multiplex sequencing, in which specific segments from multiple genomes or multiple segments from the same genome, or a combination of the two, can be sequenced simultaneously using the standard chain termination method. *Figure 1* explains the scientific principle of the technology using three targets as an example. Initially, amplicons (three in this case) are generated, and the 3′ end of these amplicons is simultaneously sequenced. This process includes a modification of the sequencing primers in which the molecular weight of the longest truncated molecule from target 1 is less than the smallest truncated molecule from target 2. Similarly, the molecular weight of the longest truncated molecule from target 2 is less than the smallest truncated molecule from target 3. When they are separated by capillary electrophoresis, the truncated molecules generated from target 1 migrate first, followed by those of target 2, then followed by those of target 3.

2. MultiGEN technology can be automated to process a large number of samples simultaneously using the HTM (High-Throughput Module, **Bio-ID Diagnostic**) concept. An HTM utilizes off-the-shelf equipment such as automated sequencers (**Applied Biosystems,** Foster City, CA) and liquid handling systems (e.g., the Biometrix FX, **Beckman Coulter** [Fullerton, CA], and those by **Tecan** [Mannedorf, Switzerland] and **Qiagen** [Hilden, Germany]). In an automated mode, results are available within 24 hr, but single tests can be processed within 8 hr. This paper documents a typical application of the basic MultiGEN platform technology and describes how the technology may be automated for routine high-throughput applications.

MultiGEN enables the cost of testing to be reduced, while providing same-day results that are more accurate than present-day technology. More importantly, the technology offers a very effective clinical tool for the investigation of infectious disease with truly syndrome-driven test menus (e.g., food—*E. coli* 0157:H7, *Salmonella*, and *Listeria*; water—*Giardia* and *Cryptosporidium*; and sexually transmitted diseases—gonorrhea, chlamydia, and ureaplasma).

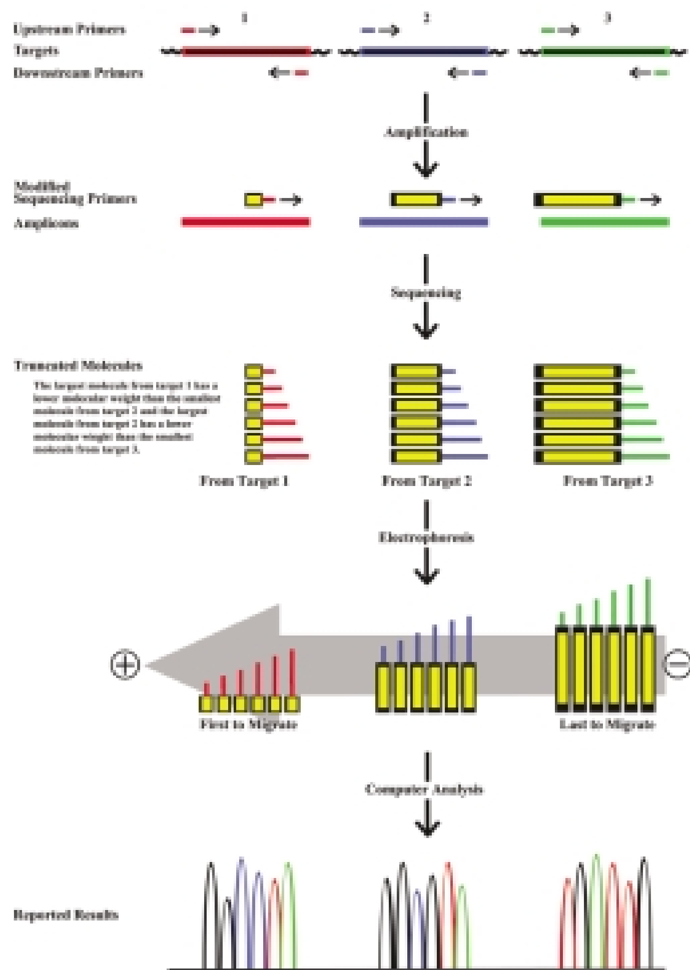

Figure 1 *Schematic representation of MultiGEN technology. Three amplicons were generated from three targets. The 3′ end of these amplicons was sequenced using modified sequencing primers (**Bio-ID Diagnostic**). The truncated molecules produced from all three target amplicons were separated by electrophoresis in a single lane.*

Materials and methods

The culture was centrifuged at 7000 RCF for 5 min to pellet the sample, and the supernatant was removed. Total DNA was extracted using the QIAamp DNA Mini Kit (**Qiagen**) according to the manufacturer's instructions. The pellet was resuspended in 180 µL of ATL buffer (**Qiagen**), and 20 µL of Proteinase K solution was added. After incubating the sample at 56 °C for 1 hr, 200 mL of AL buffer (**Qiagen**) was added, and the sample was incubated for 10 min at 70 °C. Two hundred microliters of 96% ethanol was added and the sample was transferred to the QIAamp spin column and centrifuged for 1 min at 6000 RCF. Two 500-µL washes were completed using AW1 buffer and AW2 buffer, respectively. The wash buffer was applied to the column and centrifuged at 6000 RCF for 1 min; the flowthrough was discarded each time. The sample was recovered by adding 150 µL of elution buffer, incubating at room temperature for 5 min, and then centrifuging at 6000 RCF for 1 min into a fresh 1.5-mL microfuge tube (**Rose Scientific,** Edmonton, Alberta, Canada).

PCR amplification

Specific oligonucleotide primers were designed by **Bio-ID Diagnostic** and synthesized by **Sigma-Aldrich** (Oakville, Canada). PCR amplification was performed in a 50-µL volume containing 5 µL of 10× buffer, 4 µL of 10 mM dNTP blend, 5 units of *Taq* polymerase, 10 pmol of each primer, 50 ng of target DNA, and nuclease-free water (**Invitrogen,** Ontario, Canada). The thermocycling protocol

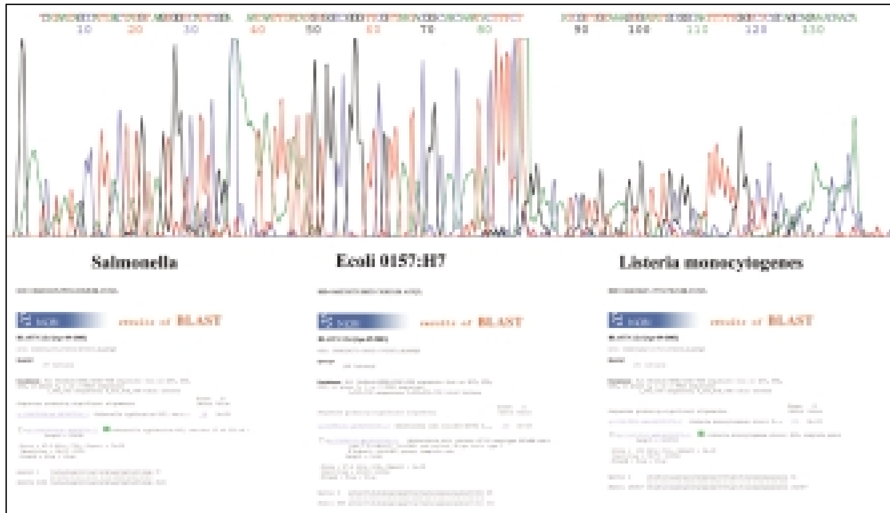

Figure 2 *Electropherogram generated from the simultaneous sequencing of* Salmonella, E. coli *0157:H7, and* Listeria, *and results from the BLAST search.*

Table 1 PCR primers and thermocycler profiles

PCR primers	
Salmonella upstream	5′-ATAACTGGCGTCAATCATAAT-3′
Salmonella downstream	5′-TCCGAATGACCACTTACATA-3′
E. coli 0157:H7 upstream	5′-TCTTCGGTATCCTATTCCC-3′
E. coli 0157:H7 downstream	5′-AGAAAGTATTTGTTGCCGTAT-3′
Listeria upstream	5′-CTAAGCGCCTCATCGACAC-3′
Listeria downstream	5′-GTTGTTTCTGCTGCGAGAC-3′
PCR thermocycler profile	
95 °C/3 min (95 °C/1 min, 56 °C/1 min, 72 °C/1 min) × 35, 4 °C/hold	
Sequencing thermocycler profile	
96 °C/10 sec (50 °C/5 sec, 60 °C/4 min) × 25, 4 °C/hold	

consisted of 35 cycles of 95 °C denaturation for 1 min, 56 °C annealing for 1 min, and 1 min of extension at 72 °C. The amplified targets were purified using PSIclone filters (**Princeton Separations,** Adelphia, NJ). Asymmetric amplification was carried out by the same method using only the downstream primer rather than both.

Cycle sequencing

Amplicons were sequenced by cycle sequencing using the ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction Kit (**Applied Biosystems**) on a GeneAmp 2400 thermocycler (**Applied Biosystems**) using the thermocycler profile 10 sec at 95 °C, 5 sec at 50 °C, and 4 min at 60 °C for 25 cycles. The modified sequencing primers were designed so that all of the amplicons could be sequenced simultaneously. Unincorporated dye terminators were removed using CentriSep chromatography columns (**Princeton Separations**). The samples were dried and resuspended in 20 µL of ABI PRISM Template Suppression Reagent (**Applied Biosystems**). Samples were analyzed by capillary electrophoresis using the ABI PRISM Genetic Analyzer 310 (**Applied Biosystems**). The 47 cm × 50 µm uncoated capillary was filled with a Performance Optimized Polymer 6 (acrylamide/urea polymer) (**Applied Biosystems**) and heated to 50 °C; 20 µL of the sequencing mixture was pipetted into a 0.2-mL microfuge tube provided by the manufacturer (**Applied Biosystems**). Samples were drawn into the capillary by an electric current (electrokinetic injection) at 2 kV for 50–200 sec. Electrophoresis was carried out at 15 kV for 20 min.

Results and discussion

Using MultiGEN technology, a number of multitarget identification assays have been developed.[11–14] A typical multitarget sequencing assay (*Salmonella*, *Listeria*, and *E. coli* 0157:H7) demonstrates use of the technology for high-throughput assays. *Figure 2* shows an electropherogram of a typical three-target assay. Three amplicons were generated from each of the

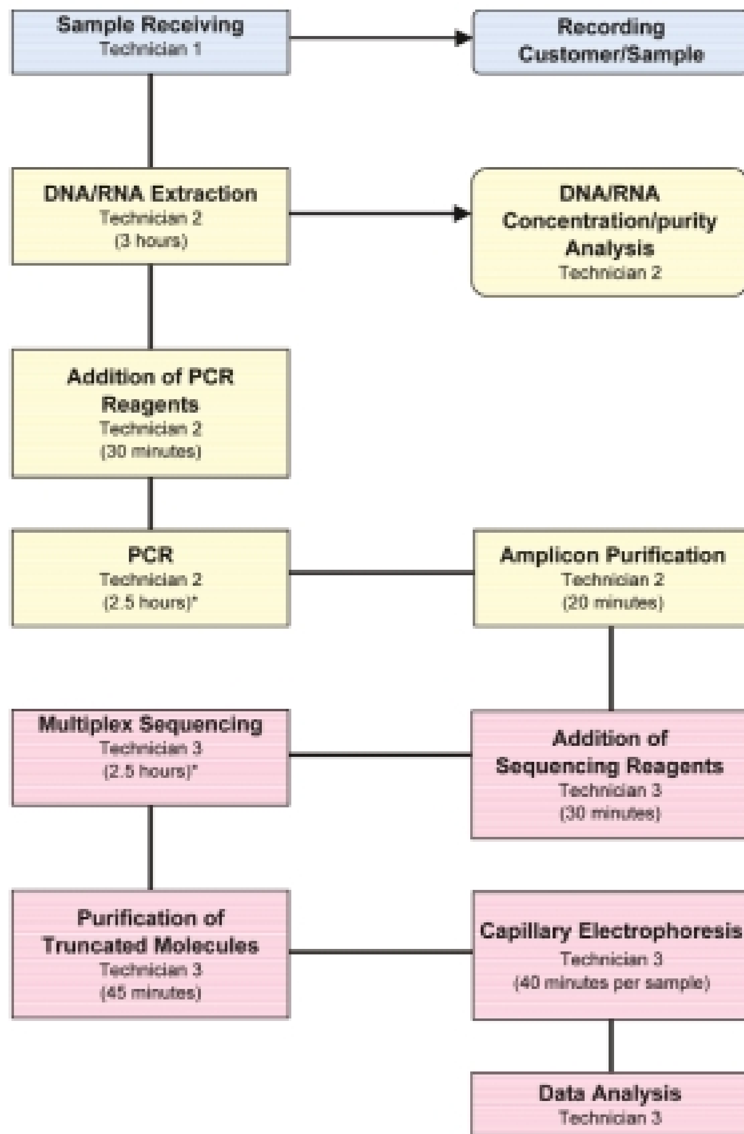

Figure 3 *Schematic representation of sample flow for a high-throughput testing module. For samples received by 10 A.M., the results will be available by 10 A.M. the following day.*

above organisms using specific primers (*Table 1*). Multiplex sequencing reactions were carried out using three modified sequencing primers. Electrophoretic separation on an automatic sequencer generated an electropherogram with 137 nucleotides. A BLAST (Basic Local Alignment Search Tool, National Institutes of Health, [NIH] Bethesda, MD) of these nucleotide sequences revealed the identity of the sequences as belonging to *Salmonella*, *E. coli* 0157:H7, and *Listeria* (Figure 2).

There are four distinct steps in a routine MultiGEN test: 1) Preparation of total DNA and/or RNA, 2) target amplification if the copy number is low, 3) DNA sequencing by chain termination, and 4) electrophoretic separation and recording of fluorescence. Steps 3 and 4 are standard methods; steps 1 and 2, which have a number of options, will be discussed.

1. Preparation of total DNA or RNA has to be sample specific, and there are a number of commercially available, ready-to-use kits. These include the QIAamp DNA Mini Kit, cat. no. 51304, and RNeasy Protect Mini Kit, cat. no. 74124 (**Qiagen**); Wizard SV 96 Genomic DNA Purification System, cat. no. A2370 (**Promega**, Madison, WI); DNA Extraction Kit, cat. no. 200600 (**Stratagene,** La Jolla, CA), and Micro Total RNA Purification System, cat. no. 12183018 (**Invitrogen**). Although purity (A260/A280) of 1.8 or greater is preferred, sequences with a purity of 1.0 have been obtained, which may not be unusual when processing real clinical and environmental samples. Using the QIAamp DNA Mini Kit, both viable and nonviable oocysts (*Giardia*) down to 10–20 oocysts per sample have been detected.

- Growing in an enrichment medium—The number of viable bacteria and fungi can be increased by growing in an enrichment medium for a short period. (The authors were able to detect a single *E. coli* 0157:H7 bacterium in 5 g of ground beef. This was carried out in two steps: 325 g of ground beef was incubated overnight at 37 °C in Luria Broth [LB], and then 5 mL of the broth was incubated at 37 °C for 2 hr.)

- Target amplification—The most common procedure to increase sensitivity is to increase the number of identical short segments of the target nucleic acid. This can be achieved by

any one of the target amplification technologies. There are commercial products for target amplification chain reactions using thermocyclers (PCR). There are also a number of linear target amplification reactions that do not require a thermocycler (e.g., transcription-mediated amplification, NASBA [Nucleic Acid Sequence Based Amplification, **Biomerieux,** Québec, Canada], LIDA [Logarithmic Isothermal DNA Amplification, **Pacific Biometrix,** Seattle, WA], and strand displacement [**BD Biosciences,** San Jose, CA]).

Tests using single-stranded and double-stranded DNA templates showed no significant changes in the relative intensities of the four nucleotides with multiplex cycle sequencing of three targets. Using single-stranded DNA template and single-step isothermal sequencing, the authors obtained relative intensities similar to double-stranded cycle sequencing. It was also observed that single-stranded DNA template and single-step sequencing gave very good repeatability when compared with double-stranded cycle sequencing.

Another advantage of MultiGEN technology is that the length of the truncated molecules from the sequencing reaction is always less than 60 nucleotides, irrespective of whether the elution position is at the beginning or toward the end of the run. Such short truncated molecules will not be subjected to the fluorescence signal drag associated with longer truncated molecules, and will therefore always generate clearly shaped signals. The sequence coming from each lane is automatically aligned against an in-house sequence data bank, and the correct identity of the target organism is then recorded. For fast, independent, third-party confirmation, some of the sample sequences are checked periodically with the GenBank (NIH) using a BLAST search with confirmation available within a few minutes.

In an automated mode, the DNA/RNA preparation, amplification, and sequencing steps can be conveniently managed with an HTM using systems that are readily available, including liquid handling systems from **Beckman Coulter** and **Tecan** and thermocyclers that handle 96–384 samples simultaneously (**Applied Biosystems**). A sample flow chart for a centralized testing facility using an HTM is shown in *Figure 3*. Typically, the final samples are loaded onto the automated sequencer at the end of the workday with sample processing overnight and results available the following morning. A typical three-target electrophoretic separation takes approx. 20 min with a total run time of only 40 min. A four-capillary automated sequencer can therefore process approx. 64 samples overnight, and a 96-capillary sequencer more than 1500 samples.

Conclusion

In this post-PCR era, MultiGEN qualifies as the new fourth wave of technology for routine diagnosis. The technology delivers the expected accurate results well within a 24-hr turnaround period.

References

1. *Sambrook J, Fritstch EF, Manniatis T. Molecular cloning laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1989:13.6–13.77.*
2. *Erlich AH. PCR technology. Principles and applications. Oxford: Oxford University Press, 1992.*
3. *Bonora S, Guitierrez MC, Di-Perri G, et al. Comparative evaluation of ligation-mediated PCR and spoligotyping as screening methods for genotyping of* Mycobacterium tuberculosis *strains. J Clin Microbiol 1999; 37(10):3118–23.*
4. *Bermann JS, Yuoh G, Fish G, Woods GL. Clinical evaluation of the enhanced Gen-Probe Amplified* Mycobacterium Tuberculosis *Direct Test for rapid diagnosis of tuberculosis in prison inmates. J Clin Microbiol 1999; 37:1419–25.*
5. *Elbeik T, Charlebois E, Nassos P, et al. Quantitative and cost comparison of ultrasensitive human immunodeficiency virus type 1 RNA viral load assays: Bayer bDNA quantiplex versions 3.0 and 2.0 and Roche PCR Amplicor monitor version 1.5. J Clin Microbiol 2000; 38(3):1113–20.*
6. *Herbart H, Gamer D, Loeffer J, et al. Evaluation of Murex CMV DNA hybrid capture assay for detection and quantitation of* Cytomegalovirus *infection in patients following allogeneic stem cell transplantation. J Clin Microbiol 1998; 36(5):1333–7.*
7. *Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. Nat Genet 1998; 19(3):225–32.*
8. *Kwiatkowski RW, Lyamichev V, De-Arruda M, Neri B. Clinical, genetic, and pharmacogenetic applications of the Invader assay. Mol Diagn 1999; 4(4):353–64.*
9. *Wasson J, Skolnick G, Love-Gregory L, Permutt MA. As-*

sessing allele frequencies of single nucleotide polymorphism in DNA pools by Pyrosequencing technology. Biotechniques 2002; 32:1144–52.

10. *Zhang W, Cohenford M, Lentricha B, et al. Detection of* Chlamydia trachomatis *by isothermal ramification amplification method: a feasibility study. J Clin Microbiol 2002; 40(1):128–32.*
11. *Vinayagamoorthy T, Mulatz K, Hodkinson R. Nucleotide sequence based multi-target identification. J Clin Microbiol 2003; 41(7):3284–92.*
12. *Vinayagamoorthy T, Mulatz K, Hodkinson R. Identification of West Nile and other Flaviviruses using multiplex sequencing (MultiGEN). Presented at the American Society of Microbiology Conference on Polymicrobial Diseases, Oct 19–23, 2003, Lake Tahoe, NV.*
13. *Hodkinson R, Mulatz K, Vinayagamoorthy T. Simultaneous genotypying of sexually transmitted bacterial pathogens using MultiGEN technology. Presented at the International Society for Sexually Transmitted Disease Research (ISSTDR) Congress, Jul 27–30, 2003, Ottawa, Canada.*
14. *Hodkinson R, Mulatz K, Vinayagamoorthy T. Identification of food pathogens using multiple signature sequences. Presented at the 12th World Food Science and Technology Congress, Jul 16–20, 2003, Chicago, IL.*

The authors are with ***Bio-ID Diagnostic Inc.,*** *#1, 410 Downey Rd., LFK Biotechnology Complex, Saskatoon, Saskatchewan S7N 4N1, Canada; tel.: 306-975-9161; fax: 306-938-0751; e-mail: moorthy@bio-id-diagnostic.com.*