Commentary

Expanding the Diagnostic Capabilities of Molecular Microbiology by Genomic Methods

Wenyong W. Zhang and James Versalovic

From the Division of Molecular Pathology, Department of Pathology, Texas Children's Hospital, Houston; and the Department of Pathology, Baylor College of Medicine, Houston, Texas

Traditional clinical microbiology practice based on phenotypic and biochemical characterization of microbes has contributed greatly to modern medicine, but such time-honored approaches have their limitations. It is estimated that between 10 and 20% of clinical isolates are novel organisms that defy phenotype-based identification, resulting in identification errors of rarely isolated or phenotypically aberrant strains.¹

Molecular techniques are particularly suitable for clinical microbiology because they do not require culture, they have rapid turnaround times, and digital genetic information can be stored in a database for epidemiological studies. Tremendous technological advances in DNA sequencing and array-based assays have propelled genomics and molecular diagnostics. For example, 16S rRNA sequence analyses are more accurate than phenotypic methods in identifying mycobacterial species and other difficult-to-identify pathogens and have supplemented phenotypic identification in some labs.^{2,3} The 16S rRNA gene sequence is widely used to identify novel pathogens, and many new species have been identified. Ribosomal RNA gene sequencing has become the current "gold standard" in microbial identification as well as the technical basis for modern bacterial taxonomy. As more microbial genomes are sequenced and gene functions elucidated, sequence-based identification methods can be further refined to identify any pathogen by using the most informative gene(s) approach.

Another exciting technology that has demonstrated clinical diagnostic utility is DNA microarray science. DNA microarrays enable simultaneous analyses of global patterns of gene expression in microorganisms or host cells. In addition, genotyping and sequencing by microarray-based hybridization have been successfully used for organism identification and molecular resistance testing. An oligonucleotide microarray containing 16S rRNA, *katG*, and *rpoB* sequences were used for simultaneous

mycobacterial species identification and molecular resistance testing.⁴ Using this array, 26 of 27 mycobacterial species, including closely related ones, were correctly identified, and 51 mutations in a 200-bp region of the *rpoB* gene that are related to rifampin-resistance were interrogated.

One particular advantage of microarray technology in pathogen detection is that the array has the capacity for simultaneous multiorganism detection in complex environmental or clinical samples. The Multi-Pathogen Identification microarray containing 53,660 oligonucleotide probes was developed to simultaneously detect 18 pathogens, including 11 bacteria, five viruses, and two eukaryotes.⁵ Species-specific primer sets were used to amplify multiple diagnostic regions unique to each pathogen. This microarray achieved high detection sensitivity and specificity by using an average of 378 probes to detect each pathogen. The sensitivity of the assay was demonstrated by documenting detection of 10 fg of purified genomic DNA or 500 fg of environmental DNA containing potential PCR inhibitors and competing targets.⁵

In addition to being used directly as a diagnostic tool, high-density microarrays have enabled comparative genomic hybridization technology to determine differences among microbial genomes. Such differences could enable discovery of virulence factors, drug or vaccine targets, unique sequences for species or strain identification, as well as assessment of phylogenetic relationships. For example, 75 diverse isolates of *Clostridium difficile* were hybridized to a microarray containing all 3688 predicted coding DNA sequences (CDSs) from the sequenced strain *C. difficile* 630.⁶ Phylogenetic analyses identified a hypervirulent clade that was characterized by shared gene loss among most of the hypervirulent strains. This study also showed that only 19.7% of genes were shared by all strains, suggesting a high degree of genetic variability or plasticity in this pathogen. Poly et al⁷

Accepted for publication August 8, 2007.

This commentary relates to Ou et al, J Mol Diagn 2007, 9:624–630, published in this issue.

Address reprint requests to James Versalovic, M.D., Ph.D., Department of Pathology, Texas Children's Hospital, 6621 Fannin MC 1-2261, Houston, TX 77030. E-mail: jxversal@texaschildrenshospital.org.

used a whole-plasmid shotgun DNA microarray approach and competitive hybridization to study genetic differences between an unsequenced invasive *Campylobacter jejuni* strain, ATCC 43431 (American Type Culture Collection, Manassas, VA), and a sequenced poorly invasive strain, NCTC 11168. They found that up to 130 genes were unique to the invasive isolate, and some of these genes might explain the invasiveness of this strain.

DNA microarrays are powerful tools in which many genes of a single pathogen or genes from different pathogens may be analyzed. In this issue of *The Journal of Molecular Diagnostics*, Ou et al⁸ beautifully illustrate an example of how to use whole-genome sequence analyses of *Salmonella enterica* Paratyphi A and existing comparative genomic hybridization data to design a highly discriminatory multiplex PCR assay that can be developed in any molecular diagnostic laboratory. In a time of overwhelmingly rapid expansion of genomic information, this article provides navigation tools and a recipe for mining the genomic databases to design species-, serovar-, or pathotype-specific PCR assays for accurate identification. Paratyphoid fever is an emerging infectious disease in developing countries, and accurate diagnostic methods are not generally available. The tasks facing Ou and his colleagues⁸ were twofold: i) how to reliably distinguish Paratyphi A from other *Salmonella enterica* serovars and two *Salmonella enterica* subspecies-*Salmonella diarizonae* and *Salmonella bongori*, and ii) how to maximize the ability of a new molecular assay to detect most or all Paratyphi A strains. The basic challenges are target specificity and universality of an assay within a taxon.

The authors' efforts to achieve their goals were enhanced by developing a web-based, freely accessible bioinformatics tool set called MobilomeFINDER, which contains several programs for strain-specific CDS identification, mobile genomic island ("mobilome") identification, primer design, and electronic PCR. All 4093 annotated CDSs of strain Paratyphi A ATCC 9150 were downloaded. Ou et al⁸ used the MobilomeFINDER's GenomeSubtractor utility to search each of the 4093 CDSs against each non-Paratyphi A *S. enterica* subspecies (complete or partial sequences), and identified 43 Paratyphi A-unique CDSs. Next, Ou et al⁸ used an array comparative genomic hybridization dataset obtained previously that included comparisons of ATCC 9150 with twelve other Paratyphi A strains. Forty-three newly identified Paratyphi A-unique CDSs were searched against the comparative genomic hybridization dataset, and 14

CDSs were present in all 13 Paratyphi A strains. These 14 CDSs were mapped on the Paratyphi A genome and were clustered at four locations. From each of four clusters, a PCR primer set was designed to amplify a representative CDS. Multiplex PCR was validated using 52 Paratyphi A strains and 75 *S. enterica* subspecies *enterica* non-Paratyphi A strains, demonstrating 100% sensitivity and 100% specificity. In short, Ou et al's elegant study⁸ describes a general method that explains how to parse DNA sequence information to design multiplex PCR assays for pathogen identification in the genomic age.

Molecular methods will be increasingly used for pathogen identification, microbial quantification, and resistance testing. The rapid evolution in microbial genomics will greatly expand the capabilities of clinical microbiology laboratories to embrace the accurate identification of novel, difficult-to-culture, or phenotypically indistinguishable pathogens. The application of genomics to assay design in molecular diagnostics, or "diagnomics," will transform the process of research and development in our field.

References

1. Clarridge III JE: Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev* 2004, 17:840-862
2. Hall L, Doerr KA, Wohlfiel SL, Roberts GD: Evaluation of the MicroSeq system for identification of Mycobacteria by 16S ribosomal DNA sequencing and its integration into a routine clinical mycobacteriology laboratory. *J Clin Microbiol* 2003, 41:1447-1453
3. Luna RA, Fasciano LR, Jones SC, Boyanton BL, Ton TT, Versalovic J: DNA pyrosequencing-based bacterial pathogen identification in a pediatric hospital setting. *J Clin Microbiol* 2007, 45:2985-2992
4. Troesch A, Nguyen H, Miyada CG, Desvarenne S, Gingeras TR, Kaplan PM, Cros P, Mabilat C: Mycobacterium species identification and rifampin resistance testing with high-density DNA probe arrays. *J Clin Microbiol* 1999, 37:49-55
5. Wilson WJ, Strout CL, DeSantis TZ, Stilwell JL, Carrano AV, Andersen GL: Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. *Mol Cell Probes* 2002, 16:119-127
6. Stabler RA, Gerding DN, Songer JG, Drudy D, Brazier JS, Trinh HT, Witney AA, Hinds J, Wren BW: Comparative phylogenomics of *Clostridium difficile* reveals clade specificity and microevolution of hyper-virulent strains. *J Bacteriol* 2006, 188:7297-7305
7. Poly F, Threadgill D, Stintzi A: Identification of *Campylobacter jejuni* ATCC 43431-specific genes by whole microbial genome comparisons. *J Bacteriol* 2004, 186:4781-4795
8. Ou H, Ju C, Thong K, Ahmad N, Deng Z, Barer M, Rajakumar K: Translational genomics to develop a *Salmonella enterica* serovar Paratyphi A multiplex PCR assay: a rapid generic strategy for pre-existing and emerging pathogens. *J Mol Diagn* 2007, 9:624-630