

16S rRNA Gene Sequencing and the Routine Clinical Microbiology Laboratory: a Perfect Marriage?

In the December 2005 issue of the *Journal of Clinical Microbiology*, Petti et al. reported three cases of endocarditis in which 16S rRNA gene sequencing revealed phenotypic misidentification of the bacterial isolates (10). The authors argued in favor of implementation of 16S rRNA gene sequencing in the clinical laboratory, stating that this genotypic method is more objective, accurate, and reliable than phenotypic methods. Regarding potential flaws, they mentioned only that 16S rRNA gene sequencing is not a perfect method. On the other hand, they clearly listed the many drawbacks of phenotypic methods. Although we do not argue against 16S rRNA gene sequencing, we believe that the opinion of the authors is too optimistic and does not always reflect the daily practice in a routine clinical microbiological laboratory. Drawbacks and pitfalls of 16S rRNA gene sequencing are well-known and have been reviewed elsewhere (6, 9). However, there has been less attention to the obstacles towards 16S rRNA gene sequencing from the routine clinical microbiology laboratory's perspective. From our daily experiences with 16S rRNA gene sequencing, we want to comment on what we believe to be the major problems regarding the implementation of 16S rRNA gene sequencing in routine practice.

First, in this era of managed care in which clinical microbiological laboratories are obliged to contain costs, the cost of 16S rRNA gene sequencing remains a major obstacle. The laboratory can save costs by purchasing directly from providers of generic primer synthesis, general molecular reagents, and sequencing services, since they operate at a competitive market. In contrast, the cost to gain access to commercial databases for 16S rRNA gene sequencing remains high and hampers its implementation in a routine clinical microbiology laboratory. To make analysis of sequencing data affordable, use of public databases such as GenBank is an alternative. This brings us to the second problem: the deposited sequences in the public databases are not peer reviewed. Since anyone can deposit and name sequence data, the quality of entries in GenBank is compromised by sequencing errors and ambiguities, incomplete sequences, and insufficient strain characterization (6, 9). On the other hand, the quality of commercial databases is also not optimal (4, 5). This is more related to the limited number of strains in the databases rather than incorrect sequence entries. Third, although sequence data are objective in nature, the interpretation of these data to establish a species designation is not. There are no universally accepted criteria for 16S rRNA gene sequence-based identification of bacterial isolates. The level of sequence homology required between the isolate under investigation and those in the databases to conclude on genus and species designation is a matter of debate (3). This introduces a level of subjectivity in sequence-based species identification and, as a consequence, potential inaccuracies in the routine laboratory setting.

In our laboratory, sequencing is used primarily as an adjunct to phenotypic methods for the identification of difficult-to-identify or rarely encountered bacteria. We analyze an amplicon of around 900 bp of the 16S rRNA gene, using primers as described by Weisburg et al. (12). For identification of streptococci, we also amplify the 16S-23S intergenic spacer region, as described by Chen et al. (1). We analyze our sequences for species identification by using the GenBank database and the BLAST algorithm with default parameters. For designation to

the species or genus level, we use the identification criteria formulated by Drancourt et al. (6) as a guideline. The genotypic results are interpreted together with the phenotypic results and clinical information to decide on the identification. It is our experience and that of others that the use of a public database as described above provides reliable identifications in the vast majority of cases (2, 7, 8). So, if cost is an issue for the laboratory, GenBank is a true alternative to commercial databases. Nevertheless, species identification is not always possible in this context. It is our experience that the frequency of problematic identifications using GenBank has been increasing over the last few years, especially due to the increase of sequence entries of ill-characterized organisms. If this trend is to continue, the usefulness of the GenBank database will be compromised further. We illustrate these issues with some of the problems we have recently experienced in our routine laboratory.

In the previous year, we have performed 16S rRNA gene sequence analysis on 285 bacterial isolates of 226 patients. The identified bacteria constitute a wide range of mostly commonly encountered bacteria but also rare ones, e.g., *Actinobaculum schaalii* (11), as expected in the laboratory of an academic hospital. However, in 14% of the cases we were able to provide identification only to the genus level. Viridans streptococci constituted 61% of these cases. Mainly the differentiation of *Streptococcus mitis*, *Streptococcus oralis*, and *Streptococcus pneumoniae* proved to be problematic. This is a well-known problem when performing gene sequencing (1), but it also illustrates a flaw in the use of the GenBank database. The following case is exemplary. Gram-positive cocci in chains were cultured from blood of a 56-year-old female. She was an oncology patient under treatment and was admitted to the hospital with neutropenic fever and mucositis. Phenotypic testing showed alpha-hemolytic streptococci of the *S. mitis/oralis* group based on Gram stain morphology, negative catalase test, optochin resistance, and negative esculin hydrolysis. The MIC of penicillin was 4.0 µg/ml. The organism was then evaluated for bile solubility, which was negative. However, when sequence data of the isolate were compared to sequences deposited in the GenBank database, *S. pneumoniae* was generated with 99% similarity for the first best matches. This identification was questioned because of the previous phenotypic results, the rarity of penicillin-resistant *S. pneumoniae* in The Netherlands, and the clinical presentation.

In 4% of the cases, we were not able to obtain any identification at first. This was due to mixed cultures and mixed or ambiguous sequences but also due to alignments with low percent identities and matching to ill-characterized strains. The latter are additional flaws which are illustrated further by the following case. An anaerobic gram-negative rod was cultured from the blood of a 61-year-old male patient admitted to the intensive care unit. Phenotypically, the bacterium was identified as belonging to the *Bacteroides fragilis* group. Analysis of the sequence data by GenBank yielded at least 100 matches, with uncultured partially sequenced microorganisms as the best matches. So, it was not possible to identify this isolate by use of the GenBank database. However, when we performed pairwise alignment with the sequences of the reference strain of *B. fragilis*, we obtained a perfect match. Align-

ment with the sequences of the other members of the *B. fragilis* group resulted in much lower percent identities and bit scores.

For now, the costs and quality of databases are major obstacles to implementation of 16S rRNA gene sequencing in the routine clinical microbiology laboratory. Use of public databases is an affordable option, but one has to be aware of their imperfections in terms of inaccurate identifications (2, 6, 9). Also, the lack of stringent interpretation criteria is an obstacle. In this context, the information obtained from phenotypic methods remains important. Nevertheless, more and more complete genome sequences of reference strains are becoming available, as is the case with data on inter- and intraspecies variability of different genera. We hope that this will improve the quality of databases and lead to clearer interpretation criteria in the future. This will guarantee the usefulness of 16S rRNA gene sequencing and hence its implementation in daily clinical laboratory practice.

REFERENCES

1. **Chen, C. C., L. J. Teng, and T. C. Chang.** 2004. Identification of clinically relevant viridans group streptococci by sequence analysis of the 16S-23S ribosomal DNA spacer region. *J. Clin. Microbiol.* **42**:2651–2657.
2. **Christensen, J. J., K. Andresen, T. Justesen, and M. Kemp.** 2005. Ribosomal DNA sequencing: experiences from use in the Danish national reference laboratory for identification of bacteria. *APMIS* **113**:621–628.
3. **Clarridge, J. E.** 2004. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin. Microbiol. Rev.* **14**:840–862.
4. **Cloud, J. L., C. Shutt, W. Aldous, and G. Woods.** 2004. Evaluation of a modified Gen-Probe Amplified Direct Test for detection of *Mycobacterium tuberculosis* complex organisms in cerebrospinal fluid. *J. Clin. Microbiol.* **42**:5341–5344.
5. **Cloud, J. L., P. S. Conville, A. Croft, D. Harmsen, F. G. Witebsky, and K. C. Carroll.** 2004. Evaluation of partial 16S ribosomal DNA sequencing for identification of *Nocardia* species by using the MicroSeq 500 system with an expanded database. *J. Clin. Microbiol.* **42**:578–584.
6. **Drancourt, M., C. Bollet, R. Carlouz, R. Martelin, J. P. Gayral, and D. Raoult.** 2000. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. *J. Clin. Microbiol.* **38**:3623–3630.
7. **Fenollar, F., V. Roux, A. Stein, M. Drancourt, and D. Raoult.** 2006. Analysis of 525 samples to determine the usefulness of PCR amplification and sequencing of the 16S rRNA gene for diagnosis of bone and joint infections. *J. Clin. Microbiol.* **44**:1018–1028.
8. **Lau, S. K., P. C. Woo, A. M. Fung, K. M. Chan, G. K. Woo, and K. Y. Yuen.** 2004. Anaerobic, non-sporulating, gram-positive bacilli bacteraemia characterized by 16S rRNA gene sequencing. *J. Med. Microbiol.* **53**:1247–1253.
9. **Patel, J. B.** 2001. 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Mol. Diagn.* **6**:313–321.
10. **Petti, C. A., C. R. Polage, and P. Schreckenberger.** 2005. The role of 16S rRNA gene sequencing in identification of microorganisms misidentified by conventional methods. *J. Clin. Microbiol.* **43**:6123–6125.
11. **Sturm, P. D., J. Van Eijk, S. Veltman, E. Meuleman, and T. Schulin.** 2006. Urosepsis with *Actinobaculum schaalii* and *Aerococcus urinae*. *J. Clin. Microbiol.* **44**:652–654.
12. **Weisburg, W. G., S. M. Barns, D. A. Pelletier, and D. J. Lane.** 1991. 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* **173**:697–703.

Michael Boudewijns
Judith M. Bakkers
Patrick D. J. Sturm
Willem J. G. Melchers*
Department of Medical Microbiology
Nijmegen University Centre for Infectious Diseases
Radboud University Nijmegen Medical Centre
P.O. Box 9101
6500 HB Nijmegen, The Netherlands

*Phone: 0031-24-2614356
 Fax: 0031-24-3540216
 E-mail: w.melchers@mmb.umcn.nl

Ed. Note: The authors of the published article declined to respond.