Utility of 16S rDNA Sequencing for Identification of Rare Pathogenic Bacteria

Shih Keng Loong,¹ Chee Sieng Khor,¹ Faizatul Lela Jafar,² and Sazaly AbuBakar^{1,2*}

¹Tropical Infectious Diseases Research & Education Centre (TIDREC), Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

²Department of Medical Microbiology, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

Background: Phenotypic identification systems are established methods for laboratory identification of bacteria causing human infections. Here, the utility of phenotypic identification systems was compared against 16S rDNA identification method on clinical isolates obtained during a 5-year study period, with special emphasis on isolates that gave unsatisfactory identification. **Methods:** One hundred and eighty-seven clinical bacteria isolates were tested with commercial phenotypic identification systems and 16S rDNA sequencing. Isolate identities determined using phenotypic identification systems and 16S rDNA sequencing were compared for similarity at genus and species level, with 16S rDNA sequencing as the reference method. **Results:** Phenotypic identification systems identified ~46% (86/187) of the isolates with identity similar to that identified using 16S rDNA

sequencing. Approximately 39% (73/187) and ~15% (28/187) of the isolates showed different genus identity and could not be identified using the phenotypic identification systems, respectively. Both methods succeeded in determining the species identities of 55 isolates; however, only ~69% (38/55) of the isolates matched at species level. 16S rDNA sequencing could not determine the species of ~20% (37/187) of the isolates. **Conclusion:** The 16S rDNA sequencing is a useful method over the phenotypic identification systems for the identification of rare and difficult to identify bacteria species. The 16S rDNA sequencing method, however, does have limitation for species-level identification of some bacteria highlighting the need for better bacterial pathogen identification tools. *J. Clin. Lab. Anal.* **30**:1056–1060, 2016. © 2016 Wiley Periodicals, Inc.

Key words: 16S rDNA; infectious disease; Malaysia; pathogenic bacteria

INTRODUCTION

Accurate identification of clinically isolated bacteria is critical for the determination of the causative agent causing an infection. It is also needed for instituting the most appropriate antimicrobial therapy and infection management strategies (1–4). Bacteria identification in a routine microbiology laboratory is often done by first culturing the bacteria in a suitable growth medium (1, 3, 5). Pure bacteria culture is obtained and following that, bacteria identification is done by various means. The most common method used largely relies on commercially available phenotypic identification systems such as the API (bioMérieux, Marcy l'Etoile, France) (1–3, 5, 6). The test is usually easy to perform, and in most cases, results can be rapidly obtained (4, 5). The phenotypic identification method, however, has relatively lower accuracy, reproducibility,

and discriminatory power often resulting in inconclusive identification (5–7). The introduction of matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) improves the accuracy of bacteria species identification; however, limitations related to cost of equipment and database coverage restricted its broad application (5, 8). 16S

Source of Support

University of Malaya High Impact Research Grant (grant no. E000013-20001).

*Correspondence to: Sazaly AbuBakar, Tropical Infectious Diseases Research & Education Centre (TIDREC), Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia. E-mail: sazaly@um.edu.my

Received 16 September 2015; Accepted 13 March 2016

DOI 10.1002/jcla.21980

Published online in Wiley Online Library (wileyonlinelibrary.com).

rDNA sequencing, on the other hand, has emerged as a widely accepted method of choice for the accurate identification of bacteria (2–4, 6, 9–12), and it is also the most well-represented gene in the public databases (GenBank, European Nucleotide Archive, DNA Data Bank of Japan and The Ribosomal Database Project) (4, 11). This method capitalizes on the presence of unique variable regions in the gene that are genus or species specific to all bacteria (1–3, 5, 7–9, 13), and comparing the sequences with those available in the public databases (12). This method is especially useful and practical in healthcare facility settings where a variety of pathogenic bacteria are encountered from patients. The method can be applied especially for identification of fastidious, rare, slow-growing bacteria and those which possess atypical phenotypic profiles (1–3, 7). Whenever the identification of bacteria isolates is not possible or doubtful by phenotypic identification systems, 16S rDNA sequencing is likely to be able to identify at least to the genus level mainly due to the availability of public databases that cover the entire extent of known taxa diversity (4). This study aimed at comparing the utility of phenotypic and 16S rDNA identification methods on clinical isolates obtained during a 5-year study period, with special emphasis on isolates that gave unsatisfactory identity using phenotypic identification systems.

MATERIALS AND METHODS

This retrospective study included 187 bacteria isolates obtained from patients admitted to University Malaya Medical Center (UMMC) from 2007 to 2011. This study received approval from the University Malaya Medical Center Medical Ethics Committee (MECID. No. 20149-575). Patient specimens were from various clinics of the UMMC, and samples were cultured on various media. All isolates were initially identified using routine Gram staining, microscopy, and biochemical tests. These tests were complimented using API kits (bioMérieux). The following API kits, API 20A, API 20E, API 20NE, API 20Strep, API 50CHB, API Coryne, API NH, Rapid 20E and Rapid ID 32A, were collectively referred to as “phenotypic identification systems” in this study. When results from the phenotypic identification systems gave weak (<95% probability) or ambiguous genus identification, 16S rDNA PCR and sequencing were performed. In cases when phenotypic identification systems yielded more than one potential species identity, the species with the highest probability was chosen and unknown identity was recorded when the system could not identify the isolate.

Bacteria DNA was extracted using NucleoSpin Tissue kit (Macherey-Nagel, Düren, Germany) according

to the manufacturer's instructions, and the extracted DNA was stored in -20°C until needed. 16S rDNA PCR was performed according to previously published protocols (9), and the amplicons were purified using NucleoSpin Gel Clean-up kit (Macherey-Nagel). The purified DNA fragments were subjected to DNA sequencing, and the resulting sequences were assembled using Sequencher 4.10.1 (Gene Codes, Ann Arbor, MI) generating approximately 1,500 nucleotides. Nucleotide sequences were compared against those available in the GenBank using BLAST (12). The entry with the highest sequence similarity was chosen for each isolate. Identities of every isolate determined using the 16S rDNA sequences and phenotypic identification systems were compared for similarity at genus and species level, with 16S rDNA sequencing as the reference method. Identification to the species level was defined as $\geq 99\%$ 16S rDNA sequence similarity to the closest GenBank entry. The isolate was assigned to the corresponding genus when its 16S rDNA sequence similarity was $<99\%$ and $\geq 95\%$ (7).

RESULTS

A total of 187 isolates were obtained over a 5-year period (2007–2011). 16S rDNA sequence analyses enabled the identification of all 187 (100%) isolates to the genus level and 150/187 (~80%) isolates to the species level. The 187 bacteria isolates belonged to the following genera: *Neisseria* ($n = 21$), *Streptococcus* ($n = 20$), *Enterococcus* ($n = 15$), *Moraxella* ($n = 12$), *Bacillus* ($n = 8$), *Campylobacter* ($n = 6$), *Burkholderia* ($n = 5$), *Corynebacterium* ($n = 5$), *Vibrio* ($n = 5$), *Clostridium* ($n = 4$), *Globicatella* ($n = 4$), *Paenibacillus* ($n = 4$), *Ralstonia* ($n = 4$), *Abiotrophia* ($n = 3$), *Brevibacterium* ($n = 3$), *Brevundimonas* ($n = 3$), *Gordonia* ($n = 3$), *Lactobacillus* ($n = 3$), *Nocardia* ($n = 3$), *Staphylococcus* ($n = 3$), *Aerococcus* ($n = 2$), *Aeromonas* ($n = 2$), *Eikenella* ($n = 2$), *Granulicatella* ($n = 2$), *Haematomobacter* ($n = 2$), *Pantoea* ($n = 2$), *Pseudomonas* ($n = 2$), *Roseomonas* ($n = 2$), *Rothia* ($n = 2$), *Serratia* ($n = 2$), and *Stenotrophomonas* ($n = 2$). Each of the following genera had only one isolate ($n = 1$), respectively: *Achromobacter*, *Actinobacillus*, *Actinobaculum*, *Aggregatibacter*, *Anaerobiospirillum*, *Arcanobacterium*, *Bacteroides*, *Brevibacillus*, *Brucella*, *Cardiobacterium*, *Citrobacter*, *Dialister*, *Escherichia*, *Haemophilus*, *Kingella*, *Microbacterium*, *Micrococcus*, *Mycobacterium*, *Mycoplasma*, *Naxibacter*, *Novosphingobium*, *Ochrobactrum*, *Oligella*, *Paracoccus*, *Peptoniphilus*, *Propionibacterium*, *Rhodococcus*, *Shewanella*, *Streptobacillus*, *Tsukamurella*, and *Weissella*.

The conventional phenotypic identification systems managed to identify ~46% (86/187) of the isolates with

TABLE 1. Performance of API Phenotypic Identification Systems in Comparison With 16S rDNA Sequencing.

	<i>n</i>	%
Bacteria genus determined using 16S rDNA sequencing	187	
Similar genus	86	45.99
Not similar genus	73	39.04
Unknown identity	28	14.97
Bacteria species determined using 16S rDNA sequencing	150	80.21
Bacteria species determined using phenotypic identification systems and 16S rDNA sequencing	55	
Similar species	38	69.09
Not similar species	17	30.91

genus identity similar to that identified using the 16S rDNA sequencing (Table 1). However, ~39% (73/187) of the isolates showed different genus identity, and ~15% (28/187) of the isolates could not be identified using the phenotypic identification system (Table 1). Both genotypic (16S rDNA sequencing) and phenotypic (API kits) methods managed to determine the species identities of 55 isolates (Table 1). However, only ~69% (38/55) of the isolates showed matching bacteria species identities using both, phenotypic identification systems and 16S rDNA sequencing (Table 1). The 16S rDNA sequencing could not determine the species of ~20% (37/187) of the isolates (Table 1), and those genus for which species identity could not be determined include *Moraxella* spp. (accession no. LN871842), *Neisseria* spp. (accession no. LN871843), and *Paenibacillus* spp. (accession no. LN871844).

DISCUSSION

In this study, 187 clinical bacteria isolates collected over a 5-year period were examined. Bacteria species identification was determined using the API phenotypic identification systems and 16S rDNA sequencing. The 16S rDNA sequencing identified 100% and ~80% of the 187 isolates to the genus and species levels, respectively. Of the 55 isolates whose species could be determined using 16S rDNA sequencing and phenotypic identification systems, only ~69% species identities matched using both methods. This suggests that at least 30% of the bacteria isolated from patient samples were not accurately identified if only the phenotypic identification systems were used. This represents a substantial percentage of the total isolates, highlighting the possibility that the patients from which the isolates were obtained, could perhaps not optimally treated. The API phenotypic identification systems designated ~46% of the total isolates into similar genus as determined by 16S rDNA sequencing with ~39% of the

isolates designated into different genus. Hence, highlighting potentially a huge discrepancy in bacteria identification when API phenotypic identification system is used alone (4, 10, 14).

The API phenotypic identification system, which identifies bacteria based on the fermentation of sugars, assimilation of certain carbon sources, and production of unique metabolites and enzymes (5, 10), had difficulties identifying bacteria with similar biochemical profiles (4, 10, 13). It is worth noting that ~15% of the clinical bacteria isolates could not be identified using the API kits, emphasizing the limitation of using the API kits. In our study, we found that the API phenotypic identification systems could not resolve *Abiotrophia defectiva*, *Clostridium sporogenes*, *Moraxella osloensis*, and *Pseudomonas oryzae*, even though their identities are present in the API and ID 32 identification databases. This could be attributed to the altered biochemical characteristics of bacteria isolated from patients who may have undergone antimicrobial therapy, unusual biochemical profiles of old cultures, phenotypic variability between different strains of the same species (10), and different results obtained from the same strain upon repeated testing (4).

The phenotypic identification systems could not accurately identify a number of these species, *Aggregatibacter aphrophilus*, *Cardiobacterium hominis*, and *Globicatella sulfidifaciens* (Table 2), despite them being members of the clinically significant classes, *Gammaproteobacteria* and *Bacilli*, which are well represented in the API identification database (6). Our findings also demonstrated compelling evidence of the utility of 16S rDNA sequencing for the identification of rare and difficult to isolate bacteria, which were not found in the API and ID 32 identification databases; these include *Brevibacterium paucivorans*, *Corynebacterium sundsvallense*, *Gordonia terrae*, *Nocardia cyriacigeorgica*, *Novosphingobium panipatense*, *Paracoccus yeei*, *Peptoniphilus harei*, and *Tsukamurella tyrosinosolvens* (Table 2). This finding is congruent with the earlier report that members of the classes *Actinobacteria* and *Clostridia* are the least represented in the API database (6). In addition, our study also found two *Alphaproteobacteria* members (*N. panipatense* and *P. yeei*) (Table 2), which could not be identified using phenotypic identification systems. In the case of *Rothia mucilaginosa*, however, the API kits we employed were not suitable (15); hence, an unknown identification was scored. The latter exemplified the additional cost that would be necessary to procure additional API kits (3) for the identification of rare and uncommon bacteria species.

The 16S rDNA sequencing method, however, is not without its own disadvantages. As shown here, the 16S rDNA sequencing could not provide adequate

TABLE 2. Identification of Rare Clinical Bacteria Isolates by API Phenotypic Identification Systems Compared With 16S rDNA Sequencing.

Phenotypic identification systems	16S rDNA sequencing	GenBank accession no.	Reference
Identity	Identity		
<i>Haemophilus paraphrophilus</i> / <i>Haemophilus aphrophilus</i>	<i>Aggregatibacter aphrophilus</i>	–	(9)
<i>Rhodococcus</i> spp./ <i>Nocardia</i> spp./ <i>Gordona</i> spp./ <i>Mycobacterium</i> spp.	<i>Tsukamurella tyrosinosolvens</i>	AY254699	(10)
Unknown	<i>Novosphingobium panipatense</i>	LN871837	This study
<i>Chryseomonas luteola</i>	<i>Pseudomonas oryzihabitans</i>	LN871840	This study
<i>Corynebacterium propinquum</i>	<i>Brevibacterium paucivorans</i>	LN871829	This study
<i>Alloiooccus otitis</i>	<i>Abiotrophia defectiva</i>	LN871828	This study
<i>Eubacterium lentum</i>	<i>Clostridium sporogenes</i>	LN871831	This study
<i>Streptococcus</i> spp.	<i>Corynebacterium sundsvallense</i>	LN871832	This study
<i>Peptostreptococcus</i> spp.	<i>Peptoniphilus harei</i>	LN871839	This study
<i>Rhodococcus</i> spp.	<i>Gordonia terrae</i>	LN871834	This study
<i>Pasteurella multocida</i>	<i>Cardiobacterium hominis</i>	LN871830	This study
Unknown	<i>Rothia mucilaginosa</i>	LN871841	This study
<i>Aerococcus viridans</i>	<i>Globicatella sulfidifaciens</i>	LN871833	This study
<i>Haemophilus parainfluenzae</i>	<i>Paracoccus yeei</i>	LN871838	This study
<i>Sphingomonas paucimobilis</i>	<i>Moraxella osloensis</i>	LN871835	This study
<i>Propionibacterium avidum</i>	<i>Nocardia cyriacigeorgica</i>	LN871836	This study

discrimination for the identification to species level for some bacteria (1, 4–6, 8) and these include *Moraxella*, *Neisseria*, and *Paenibacillus*. One plausible explanation is that the 16S rDNA gene evolved slowly in these groups of bacteria; hence, it lacks the required resolution to differentiate between the different species (16). *Moraxella* and *Neisseria*, for example, once belonged to the same family, *Neisseriaceae* before a new family, and *Moraxellaceae* was proposed in 1991 (17), suggesting insufficient sequence variations for characterization to the species level. It was also reported that the presence of multiple copies of the 16S rDNA gene restricted *Neisseria* (13) and *Paenibacillus* (18) species identification. Another limitation of the 16S rDNA sequencing method is that it depends on the quality of the public databases, as sequences are deposited regardless of the correct assignment, length, and the number of ambiguous nucleotides (4).

Changes in the taxonomic description of bacteria could also lead researchers into misidentification of isolates (4), as illustrated by the classification of genus *Moraxella* into the new family *Moraxellaceae*, from *Neisseriaceae* (17). Furthermore, the 16S rDNA sequencing method essentially uses polymerase chain amplification to amplify DNA fragments for sequencing, and it is subject to various sources of errors. Errors in PCR polymerases and sequencing can significantly impact downstream sequence data by introducing false nucleotides (11).

Currently, there is no single quintessential method to identify all bacteria species. Findings from this study suggest that 16S rDNA sequencing should

complement existing phenotypic identification systems to accurately identify clinically relevant, fastidious, and rare bacteria isolates. A full bacterial genome sequencing (19) could also be feasible if the cost becomes less prohibitive. In conclusion, we showed here nonetheless that the 16S rDNA sequencing is a useful method for the identification of rare and difficult to identify bacteria species. This method, however, does have its limitation for species-level identification of some bacteria.

REFERENCES

1. Woo PCY, Lau SKP, Teng JLL, Tse H, Yuen KY. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clin Microbiol Infect* 2008;14:908–934.
2. Jenkins C, Ling CL, Ciesielczuk HL, et al. Detection and identification of bacteria in clinical samples by 16S rRNA gene sequencing: comparison of two different approaches in clinical practice. *J Med Microbiol* 2012;61:483–488.
3. Millar BC, Xu J, Moore JE. Molecular diagnostics of medically important bacterial infections. *Curr Issues Mol Biol* 2007;9:21–39.
4. Bosshard PP, Zbinden R, Abels S, Böddinghaus B, Altwegg M, Böttger EC. 16S rRNA gene sequencing versus the API 20 NE system and the VITEK 2 ID-GNB card for identification of nonfermenting gram-negative bacteria in the clinical laboratory. *J Clin Microb* 2006;44:1359–1366.
5. Vithanage NR, Yeager TR, Jadhav SR, Palombo EA, Datta N. Comparison of identification systems for psychrotrophic bacteria isolated from raw bovine milk. *Int J Food Microbiol* 2014; 189:26–38.
6. França L, Simões C, Taborda M, Diogo C, da Costa MS. Microbial contaminants of cord blood units identified by 16S rRNA sequencing and by API test system, and antimicrobial sensitivity profiling. *PLoS One* 2015;10:e0141152.

7. de Melo Oliveira MG, Abels S, Zbinden R, Bloemberg GV, Zbinden A. Accurate identification of fastidious Gram-negative rods: integration of both conventional phenotypic methods and 16S rRNA gene analysis. *BMC Microbiol* 2013;13:162.
8. Wragg P, Randall L, Whatmore AM. Comparison of Biolog GEN III MicroStation semi-automated bacterial identification system with matrix-assisted laser desorption ionization-time of flight mass spectrometry and 16S ribosomal RNA gene sequencing for the identification of bacteria of veterinary interest. *J Microbiol Meth* 2014;105:16–21.
9. Misbah S, Hassan H, Yusof MY, Hanifah YA, AbuBakar S. Genomic species identification of *Acinetobacter* of clinical isolates by 16S rDNA sequencing. *Singapore Med J* 2005;46:461–464.
10. Kim M, Heo SR, Choi SH, et al. Comparison of the MicroScan, VITEK 2, and Crystal GP with 16S rRNA sequencing and MicroSeq 500 v2.0 analysis for coagulase-negative Staphylococci. *BMC Microbiol* 2008;8:233.
11. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One* 2011;6:e27310.
12. Loong SK, Mahfodz NH, Mohamad Wali HA, et al. Molecular and antimicrobial analyses of non-classical *Bordetella* isolated from a laboratory mouse. *J Vet Med Sci* 2016; e-pub ahead of print 17 January 2016; doi:10.1292/jvms.15-0472.
13. Mechergui A, Achour W, Hassen AB. Comparison of 16S rRNA sequencing with biochemical testing for species-level identification of clinical isolates of *Neisseria* spp. *World J Microbiol Biotechnol* 2014;30:2181–2188.
14. Moraes PM, Perin LM, Júnior AS, Nero LA. Comparison of phenotypic and molecular tests to identify lactic acid bacteria. *Braz J Microbiol* 2013;44:109–112.
15. Cho EJ, Sung H, Park SJ, Kim MN, Lee SO. *Rothia mucilaginosa* pneumonia diagnosed by quantitative cultures and intracellular organisms of bronchoalveolar lavage in a lymphoma patient. *Ann Lab Med* 2013;33:145–149.
16. Fox GE, Wisotzkey JD, Jurtshuk P Jr. How close is close, 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* 1992;42:166–170.
17. Rossau R, van Landschoot A, Gillis M, De Ley J. Taxonomy of *Moraxellaceae* fam. nov., a new bacterial family to accommodate the genera *Moraxella*, *Acinetobacter*, and *Psychrobacter* and related organisms. *Int J Syst Bacteriol* 1991;41:310–319.
18. da Mota FF, Gomes EA, Paiva E, Rosado AS, Seldin L. Use of *rpoB* gene analysis for identification of nitrogen-fixing *Paenibacillus* species as an alternative to the 16S rRNA gene. *Lett Appl Microbiol* 2004;39:34–40.
19. Tan KK, Tan YC, Chang LY, et al. Full genome SNP-based phylogenetic analysis reveals the origin and global spread of *Brucella melitensis*. *BMC Genom* 2015;16:93.