In silico analysis of 16S rRNA gene sequencing based methods for identification of medically important aerobic Gram-negative bacteria

Jade L. L. Teng,¹ Ming-Yiu Yeung,¹ Geoffrey Yue,¹ Rex K. H. Au-Yeung,¹ Eugene Y. H. Yeung,¹ Ami M. Y. Fung,¹ Herman Tse,^{1,2,3,4} Kwok-Yung Yuen,^{1,2,3,4} Susanna K. P. Lau^{1,2,3,4} and Patrick C. Y. Woo^{1,2,3,4}

Correspondence

Patrick C. Y. Woo
pcywoo@hkucc.hku.hk
Susanna K. P. Lau
skplau@hkucc.hku.hk

¹Department of Microbiology, The University of Hong Kong, Hong Kong SAR

²State Key Laboratory of Emerging Infectious Diseases, The University of Hong Kong, Hong Kong SAR

³Research Centre of Infection and Immunology, The University of Hong Kong, Hong Kong SAR

⁴Carol Yu Centre for Infection, The University of Hong Kong, Hong Kong SAR

This study provides guidelines on the usefulness of full and 527 bp 16S rRNA gene sequencing and Microseq databases for identifying medically important aerobic Gram-negative bacteria. Overall, full and 527 bp 16S rRNA gene sequencing can identify 26.1 % and 32.6 %, respectively, of medically important aerobic Gram-negative bacteria confidently to the species level, whereas the full-MicroSeq and 500-MicroSeq databases can identify 15.2 % and 26.1 %, respectively, of medically important aerobic Gram-negative bacteria confidently to the species level. Among the major groups of aerobic Gram-negative bacteria, the methods and databases are least useful for identification of *Aeromonas*, *Bordetella* and *Bartonella* species. None of the *Aeromonas* species can be confidently or doubtfully identified, whereas only 0 % and 0–33.3 % of *Bordetella* species and 0–10 % and 0–10 % of *Bartonella* species can be confidently and doubtfully identified, respectively. On the other hand, these methods and databases are most useful for identification of members of the families *Pasteurellaceae* and *Legionellaceae* and *Campylobacter* species: 29.6–59.3 % and 7.4–18.5 % of members of *Pasteurellaceae*, 36–52 % and 12–24 % of members of *Legionellaceae*, and 26.7–60 % and 0–13.3 % of *Campylobacter* species can be confidently and doubtfully identified, respectively. Thirty-nine medically important aerobic Gram-negative bacteria that should be confidently identified by full 16S rRNA gene sequencing are not included in the full-MicroSeq database. Twenty-three medically important aerobic Gram-negative bacteria that should be confidently identified by 527 bp 16S rRNA gene sequencing are not included in the 500-MicroSeq database. Compared with results of our previous studies on anaerobic and Gram-positive bacteria, full and 527 bp 16S rRNA gene sequencing are able to confidently identify significantly more anaerobic Gram-positive and Gram-negative bacteria than aerobic Gram-positive and Gram-negative bacteria.

Received 7 November 2010

Accepted 10 April 2011

INTRODUCTION

Since its first use 30 years ago, 16S rRNA gene sequencing has played a pivotal role in bacterial classification and discovery of novel bacteria (Woo *et al.*, 2008). In the last decade, as a result of the widespread use of PCR and DNA sequencing, this technology has been increasingly applied for identification of medically important bacteria (Lau *et al.*, 2002, 2004; Teng *et al.*, 2003; Woo *et al.*, 2001a, b, 2002, 2003a, 2004, 2005; Yuen *et al.*, 2001). Although 16S

rRNA gene sequencing has been widely used in reference laboratories, one of the major limitations to its wider use in routine clinical microbiology laboratories is the difficulty associated with interpretation of 16S rRNA gene sequence results. The use of 16S rRNA gene sequencing for bacterial identification depends on significant inter-species differences and small intra-species differences in 16S rRNA gene sequences. As a result, one of the major limitations is that when two bacterial species share a similar 16S rRNA gene sequence, this technology alone would not be useful for distinguishing them confidently. This limitation has also generated the main obstacle for clinical microbiologists and

Supplementary material is available with the online version of this paper.

technicians when they interpret BLAST results generated from their 16S rRNA gene sequence data. Misinterpretation of such data will lead to incorrect identification of bacteria, which defeats the original purpose of 16S rRNA gene sequencing.

With the aim of facilitating 16S rRNA gene sequence interpretation, the MicroSeq 500 16S rRNA bacterial identification system (version 2.1) (MicroSeq, Applied Biosystems) has been designed. The 500-MicroSeq database can be used to analyse a 527 bp fragment amplified and sequenced from the 5' end of the 16S rRNA gene (Fontana *et al.*, 2005; Lau *et al.*, 2006; Tang *et al.*, 1998; Woo *et al.*, 2003b). A full 16S rRNA gene sequence (full-MicroSeq) database (version 2.0) has also been generated. Nevertheless, the MicroSeq databases do not take account of bacterial species sharing similar 16S rRNA gene sequences that cannot be confidently identified by 16S rRNA gene sequence analysis. Another drawback for identification of medically important bacteria is that the MicroSeq databases do not include a significant number of medically important bacteria that 16S rRNA gene sequencing is able to identify.

In view of these limitations, 5 years ago we started to systematically evaluate the usefulness of full and 527 bp 16S rRNA gene sequencing and the MicroSeq databases for identification of various groups of medically important bacteria. In the last 4 years, we have published the results for medically important anaerobes and aerobic Gram-positive bacteria, respectively (Woo *et al.*, 2007, 2009). We found that 53.3–67.1% and 22.7–37.3% of medically important anaerobic and aerobic Gram-positive bacteria that should be confidently identified by 16S rRNA gene sequencing are not included in the full- and 500-MicroSeq databases, respectively (Woo *et al.*, 2007, 2009). In this article, we describe the results of the evaluation of the usefulness of full and 527 bp 16S rRNA gene sequencing and the MicroSeq databases for identification of 328 medically important aerobic Gram-negative bacteria. The usefulness of full and 527 bp 16S rRNA gene sequencing and the MicroSeq databases for identification of aerobic Gram-positive bacteria, aerobic Gram-negative bacteria, anaerobic Gram-positive bacteria and anaerobic Gram-negative bacteria was also compared.

METHODS

16S rRNA gene sequences of medically important aerobic Gram-negative bacteria. The species of medically important aerobic Gram-negative bacteria included in this study include all 32 species of the 13 genera of *Alphaproteobacteria*, 67 species of the 19 genera of *Betaproteobacteria*, 195 species of the 55 genera of *Gamma-proteobacteria*, two species of the one genus of *Deltaproteobacteria* and 32 species of the three genera of *Epsilonproteobacteria* listed in the most recent edition of the *Manual of Clinical Microbiology* (Murray *et al.*, 2007). For each bacterial species, a list of the 16S rRNA gene sequences was retrieved from the GenBank database (release 179.0). In the list, the most representative 16S rRNA gene sequence for each species was chosen for analysis according to the following criteria: (a) strains with good phenotypic and/or genotypic characterization

(e.g. type strains, strains with complete genomes sequenced) were preferred; (b) strains isolated from humans were preferred to those isolated from animals or the environment; (c) sequences with fewer undetermined bases were preferred to those with more undetermined bases; and (d) longer sequences, especially those with better coverage of the 5' end, were preferred to shorter sequences and those with poorer coverage of the 5' end.

Comparison of full 16S rRNA gene sequences of medically important aerobic Gram-negative bacteria. The percentage differences of the 16S rRNA gene sequences among the different species in the same group/genus of medically important aerobic Gram-negative bacteria were determined by pairwise alignment using CLUSTAL W.

Comparison of 527 bp 16S rRNA gene sequences of medically important aerobic Gram-negative bacteria. The 527 bp sequences that should be amplified by the primers of MicroSeq and analysed in the 500-MicroSeq database were extracted from the full 16S rRNA gene sequences. The percentage differences of the resultant partial 16S rRNA gene sequences among the different species in the same group/genus of medically important aerobic Gram-negative bacteria were determined by pairwise alignment using CLUSTAL W.

MicroSeq databases. All medically important aerobic Gram-negative bacteria listed in the most recent edition of the *Manual of Clinical Microbiology* (Murray *et al.*, 2007) were searched in the 500-MicroSeq database (version 2.1) and full-MicroSeq databases (version 2.0) (MicroSeq, Applied Biosystems).

Definitions. In this study, a bacterial species is defined as 'can be confidently identified by 16S rRNA gene sequencing' if there is >3% difference between the 16S rRNA gene sequence of the species and those of other medically important bacterial species. A bacterial species is defined as 'cannot be confidently identified by 16S rRNA gene sequencing' if there is <2% difference between the 16S rRNA gene sequence of the species and that of one or more medically important aerobic Gram-negative bacterial species. A bacterial species is defined as 'can only be doubtfully identified by 16S rRNA gene sequencing' if there is 2–3% difference between the 16S rRNA gene sequence of the species and that of one or more medically important aerobic Gram-negative bacterial species.

RESULTS

Full 16S rRNA gene sequencing should be useful for the identification of 8 of 32 *Alphaproteobacteria* (none of 10 *Bartonella* species), 18 of 67 *Betaproteobacteria* (none of 9 *Bordetella* species, 1 of 13 *Burkholderia* species and 3 of 12 *Neisseria* species), 50 members of 195 *Gammaproteobacteria* (none of the members of 10 *Aeromonadaceae*, 6 members of 83 *Enterobacteriaceae*, 10 members of 25 *Legionellaceae*, 16 members of 27 *Pasteurellaceae*, 3 members of 17 *Moraxellaceae*, 4 members of 11 *Pseudomonadaceae* and 4 members of 12 *Vibrionaceae*), none of 2 *Deltaproteobacteria* (none of 2 *Desulfovibrio* species) and 10 members of 32 *Epsilonproteobacteria* (7 of 15 *Campylobacter* species and 2 of 11 *Helicobacter* species). For the existing full-MicroSeq database, it should be useful for the identification of 4 members of 32 *Alphaproteobacteria* (none of 10 *Bartonella* species), 9 members of 67 *Betaproteobacteria* (none of 9 *Bordetella* species, none of 13 *Burkholderia* species and 3 of 12 *Neisseria* species), 34 members of 195 *Gammaproteobacteria* (none of

the members of 10 *Aeromonadaceae*, 6 members of 83 *Enterobacteriaceae*, 9 members of 25 *Legionellaceae*, 8 members of 27 *Pasteurellaceae*, 1 member of 17 *Moraxellaceae*, 3 members of 11 *Pseudomonadaceae* and 2 members of 12 *Vibrionaceae*), none of 2 *Deltaproteobacteria* (none of 2 *Desulfovibrio* species) and 6 members of 32 *Epsilonproteobacteria* (4 of 15 *Campylobacter* species and 1 of 11 *Helicobacter* species).

The 527 bp 16S rRNA gene sequencing should be useful for the identification of 8 members of 32 *Alphaproteobacteria* (1 of 10 *Bartonella* species), 25 members of 67 *Betaproteobacteria* (none of 9 *Bordetella* species, 2 of 13 *Burkholderia* species and 5 of 12 *Neisseria* species), 64 members of 195 *Gammaproteobacteria* (none of the members of 10 *Aeromonadaceae*, 13 members of 83 *Enterobacteriaceae*, 13 members of 25 *Legionellaceae*, 16 members of 27 *Pasteurellaceae*, 5 members of 17 *Moraxellaceae*, 5 members of 11 *Pseudomonadaceae* and 5 members of 12 *Vibrionaceae*), 2 of 2 *Deltaproteobacteria* (2 of 2 *Desulfovibrio* species) and 10 members of 32 *Epsilonproteobacteria* (9 of 15 *Campylobacter* species and 1 of 11 *Helicobacter* species). For the existing 500-MicroSeq database, it should be useful for the identification of 5 members of 32 *Alphaproteobacteria* (1 of 10 *Bartonella* species), 19 members of 67 *Betaproteobacteria* (none of 9 *Bordetella* species, 2 of 13 *Burkholderia* species and 5 of 12 *Neisseria* species), 60 members of 195 *Gammaproteobacteria* (none of the members of 10 *Aeromonadaceae*, 13 members of 83 *Enterobacteriaceae*, 12 members of 25 *Legionellaceae*, 15 members of 27 *Pasteurellaceae*, 5 members of 17 *Moraxellaceae*, 5 members of 11 *Pseudomonadaceae*

and 5 members of 12 *Vibrionaceae*), none of 2 *Deltaproteobacteria* (none of 2 *Desulfovibrio* species) and 7 members of 32 *Epsilonproteobacteria* (6 of 15 *Campylobacter* species and 1 of 11 *Helicobacter* species) (Table 1 and Supplementary Tables S1–S8, available with the online version of this paper).

DISCUSSION

In this and two previous studies, we systematically analysed the usefulness and degree of confidence in using 16S rRNA gene sequence for identification of medically important bacterial species (Woo *et al.*, 2007, 2009). Since different bacterial species are likely to evolve at different rates, there has been no agreement on a universal cut-off for bacterial species delineation when 16S rRNA gene sequencing is used for their identification. In the literature, different cut-offs, ranging from 97 % to 99.5 % nucleotide identity, have been arbitrarily used by various groups. Stackebrandt & Goebel (1994) proposed 97 % nucleotide identity in the 16S rRNA gene as the cut-off for bacterial speciation. Drancourt *et al.* (2000) employed >99 % and >97 % nucleotide identity as the cut-off for species and genus identification respectively. However, these sharp cut-offs were only made for practical purposes for interpretation of their large sequence dataset. On the other hand, Janda & Abbott (2007) suggested that a minimum of >99 %, and ideally >99.5 %, nucleotide identity be used for bacterial species identification. In the present study and two previous ones (Woo *et al.*, 2007, 2009), we used >3 %,

Table 1. Number and percentage of major groups of medically important Gram-negative bacteria confidently identified by full 16S rRNA gene sequence, 527 bp 16S rRNA gene sequence and the MicroSeq databases

Bacterial taxon	Total no. of species	No. (%) of species confidently identified			
		Full 16S rRNA gene sequencing	Full-MicroSeq database	527 bp 16S rRNA gene sequencing	500-MicroSeq database
<i>Alphaproteobacteria</i>	32	8 (25.0)	4 (12.5)	8 (25.0)	5 (15.6)
<i>Bartonella</i>	10	0 (0)	0 (0)	1 (10)	1 (10)
<i>Betaproteobacteria</i>	67	18 (26.9)	9 (13.4)	25 (37.3)	19 (28.4)
<i>Bordetella</i>	9	0 (0)	0 (0)	0 (0)	0 (0)
<i>Burkholderia</i>	13	1 (7.7)	0 (0)	2 (15.4)	2 (15.4)
<i>Neisseria</i>	12	3 (25)	3 (25)	5 (41.7)	5 (41.7)
<i>Gammaproteobacteria</i>	195	50 (25.6)	34 (17.4)	64 (32.8)	60 (30.8)
<i>Aeromonadaceae</i>	10	0 (0)	0 (0)	0 (0)	0 (0)
<i>Enterobacteriaceae</i>	83	6 (7.2)	6 (7.2)	13 (15.7)	13 (15.7)
<i>Legionellaceae</i>	25	10 (40)	9 (36)	13 (52)	12 (48)
<i>Pasteurellaceae</i>	27	16 (59.3)	8 (29.6)	16 (59.3)	15 (55.6)
<i>Moraxellaceae</i>	17	3 (17.6)	1 (5.9)	5 (29.4)	5 (29.4)
<i>Pseudomonadaceae</i>	11	4 (36.4)	3 (27.3)	5 (45.5)	5 (45.5)
<i>Vibrionaceae</i>	12	4 (33.3)	2 (16.7)	5 (41.7)	5 (41.7)
<i>Deltaproteobacteria</i>	2	0 (0)	0 (0)	2 (100)	0 (0)
<i>Epsilonproteobacteria</i>	32	10 (31.2)	6 (18.8)	10 (31.2)	7 (21.9)
<i>Campylobacter</i>	15	7 (46.7)	4 (26.7)	9 (60)	6 (40)
<i>Helicobacter</i>	11	2 (18.2)	1 (9.1)	1 (9.1)	1 (9.1)

2–3 % and <2 % difference between the most representative 16S rRNA gene sequence of bacterial species obtained from GenBank and those of other medically important bacteria to indicate the degree of confidence when using 16S rRNA gene sequence for identification of medically important bacterial species. For example, for aerobic Gram-negative bacteria, the percentage of most groups of bacteria, including *Bordetella*, *Neisseria*, *Enterobacteriaceae*, *Legionellaceae*, *Pasteurellaceae*, *Moraxellaceae*, *Pseudomonadaceae*, *Vibrionaceae*, *Deltaproteobacteria*, *Campylobacter* and *Helicobacter*, that can be identified would increase if a cut-off of 2 % instead of 3 % was used (Supplementary Fig. S1). Similarly the percentage of most groups of bacteria, including *Burkholderia*, *Neisseria*, *Enterobacteriaceae*, *Legionellaceae*, *Pasteurellaceae*, *Moraxellaceae*, *Pseudomonadaceae*, *Vibrionaceae*, *Campylobacter* and *Helicobacter*, that can be identified would decrease if a cut-off of 4 % instead of 3 % was used (Fig. S1). When there is 2–3 % difference between the 16S rRNA gene sequence of a bacterial species and a closely related one, the degree of 16S rRNA gene sequence homogeneity in different strains of the species would be very important in determining whether 16S rRNA gene sequencing is able to identify the bacterium to the species level. The only group of bacteria for which the percentage of identification is not affected by either an increase or a decrease in the cut-off is *Aeromonadaceae* (Fig. S1). This is because all *Aeromonas* species shared >99 % identities in their 16S rRNA gene sequences. Such guidelines for clinical microbiologists and technicians are important because if the first hit by BLASTN search against the GenBank database is used for deciding the identity of a 16S rRNA gene sequence, a significant amount of the searches will give wrong or ambiguous results. For example, the first hit

of 15 (18.1 %) of the 83 *Enterobacteriaceae* members by BLASTN search using our representative 16S rRNA gene sequences was of a wrong species (five were of a wrong genus) and another 15 (18.1 %) gave ambiguous results, even after filtering off all 'uncultured/environmental sample sequences'.

16S rRNA gene sequencing is more useful for identification of medically important anaerobic bacteria than aerobic bacteria. Comparing the results of the present study and our two previous ones on medically important anaerobes and aerobic Gram-positive bacteria (Woo *et al.*, 2007, 2009), 16S rRNA gene sequencing is found to be able to confidently identify significantly more anaerobic Gram-positive (62.4 %) and Gram-negative (75 %) bacteria than aerobic Gram-positive (26.8 %) and Gram-negative (26.1 %) bacteria (Table 2). A similar phenomenon was also observed when the first 527 bp of the 16S rRNA genes were analysed (Table 2). On the other hand, for both the full- and 500-MicroSeq databases, their usefulness for identification of anaerobic Gram-positive and Gram-negative bacteria was only slightly better than for aerobic Gram-positive and Gram-negative bacteria (Table 2). This is because there are more anaerobic Gram-positive and Gram-negative bacteria that should be identifiable by full or 527 bp 16S rRNA gene sequencing but are not included in the full- and 500-MicroSeq databases (Tables 2 and 3). Notably, 527 bp 16S rRNA gene sequencing can always identify more bacteria than full 16S rRNA gene sequencing, and the 500-MicroSeq database can always identify more bacteria than the full-MicroSeq database for all groups of bacteria. This is because the 5' 527 bp of the 16S rRNA gene is more variable than the other parts of the gene.

Table 2. Comparison of full 16S rRNA gene sequence, 527 bp 16S rRNA gene sequence and the existing MicroSeq databases for identification of medically important aerobic Gram-negative, aerobic Gram-positive, anaerobic Gram-negative and anaerobic Gram-positive bacteria

Sequence analysis and database	No. (%) of species confidently identified‡			
	Aerobic Gram-negative bacteria (n=356)	Aerobic Gram-positive bacteria (n=380)	Anaerobic Gram-negative bacteria (n=84)	Anaerobic Gram-positive bacteria (n=157)
Full 16S rRNA gene sequence analysis	93 (26.1) ^{c,d}	102 (26.8) ^{e,f}	63 (75.0) ^{c,e}	98 (62.4) ^{d,f}
527 bp 16S rRNA gene sequence analysis	116 (32.6) ^{g,h}	171 (45.2) ^{i,j}	71 (84.5) ^{g,i}	113 (72.0) ^{h,j}
Full-MicroSeq database*	54 (15.2) ^k	64 (16.9) ^l	23 (27.4) ^{k,l}	30 (19.1)
500-MicroSeq database†	93 (26.1) ^m	132 (34.9)	30 (35.7)	56 (35.7) ^m
Should be confidently identified by full 16S rRNA gene sequence analysis but not included in full-MicroSeq database	39 (41.9) ^{n,o}	38 (37.3) ^{p,q}	40 (63.5) ^{n,p}	68 (69.4) ^{o,q}
Should be confidently identified by 527 bp 16S rRNA gene sequence analysis but not included in 500-MicroSeq database	23 (19.8) ^{r,s}	39 (22.7) ^{t,u}	41 (57.7) ^{r,t}	57 (50.4) ^{s,u}

*Calculated using the full-MicroSeq database version 2.0.

†Calculated using the 500-MicroSeq database version 2.1.

‡Significance by chi-squared test: ^{c, d, e, f, g, h, i, j, o, p, q, s, t, u}, $P < 0.001$; ^{k, n, r}, $P < 0.01$; ^{l, m}, $P < 0.05$.

Table 3. Medically important aerobic Gram-negative bacteria that can be confidently identified by full or 527 bp 16S rRNA gene sequencing but not included in the MicroSeq databases

Bacterial taxon	Bacterial species not included in:	
	Full-MicroSeq database	500-MicroSeq database
Alphaproteobacteria		
<i>Agrobacterium</i>	<i>A. tumefaciens</i>	<i>A. tumefaciens</i>
<i>Asaia</i>	<i>A. bogorensis</i>	<i>A. bogorensis</i>
<i>Azospirillum</i>	<i>A. brasilense</i>	<i>A. brasilense</i>
<i>Paracoccus</i>	<i>P. yeii</i>	
Betaproteobacteria		
<i>Advenella</i>	<i>A. incenata</i>	<i>A. incenata</i>
<i>Alcaligenes</i>	<i>A. faecalis</i>	
<i>Burkholderia</i>	<i>B. fungorum</i>	
<i>Kerstersia</i>	<i>K. gyiorum</i>	<i>K. gyiorum</i>
<i>Kingella</i>	<i>K. kingae</i>	<i>K. kingae</i>
	<i>K. oralis</i>	<i>K. oralis</i>
	<i>K. potus</i>	<i>K. potus</i>
<i>Laribacter</i>	<i>L. hongkongensis</i>	<i>L. hongkongensis</i>
<i>Massilia</i>	<i>M. timonae</i>	
<i>Ralstonia</i>		<i>R. mannitolilytica</i>
Gammaproteobacteria		
<i>Acinetobacter</i>	<i>A. ursingii</i>	
<i>Aggregatibacter</i>	<i>A. actinomycetemcomitans</i>	
	<i>A. aphrophilus</i>	
	<i>A. segnis</i>	
<i>Avibacterium</i>	<i>A. gallinarum</i>	
<i>Balneatrix</i>	<i>B. alpica</i>	<i>B. alpica</i>
<i>Cardiobacterium</i>	<i>C. valvarum</i>	<i>C. valvarum</i>
<i>Coxiella</i>	<i>C. burnetii</i>	<i>C. burnetii</i>
<i>Haemophilus</i>	<i>H. ducreyi</i>	
	<i>H. haemoglobinophilus</i>	<i>H. haemoglobinophilus</i>
	<i>H. pittmaniae</i>	
<i>Legionella</i>	<i>L. lytica</i>	<i>L. lytica</i>
<i>Moraxella</i>	<i>M. atlantae</i>	
	<i>M. osloensis</i>	
<i>Pseudomonas</i>	<i>P. luteola</i>	
<i>Vibrio</i>	<i>V. cincinnatiensis</i>	
	<i>V. metschnikovii</i>	
Deltaproteobacteria		
<i>Desulfovibrio</i>		<i>D. desulfuricans</i>
		<i>D. fairfieldensis</i>
Epsilonproteobacteria		
<i>Campylobacter</i>	<i>C. concisus</i>	<i>C. concisus</i>
	<i>C. hominis</i>	<i>C. hominis</i>
	<i>C. showae</i>	<i>C. showae</i>
	<i>C. sputorum</i>	
<i>Helicobacter</i>	<i>H. winthamensis</i>	

Similar to medically important anaerobic bacteria and aerobic Gram-positive bacteria, 16S rRNA gene sequencing is particularly not useful for identification of certain groups of medically important aerobic Gram-negative bacteria. For anaerobes, 16S rRNA gene sequencing is able to identify only one out of 11 medically important *Fusobacterium* species (Woo *et al.*, 2007). For aerobic Gram-positive bacteria, 16S rRNA gene sequencing is not useful for identification of staphylococci, enterococci, mycobacteria and nocardia (Woo

et al., 2009). As for aerobic Gram-negative bacteria, 16S rRNA gene sequencing is able to identify none of the *Bartonella* species, *Bordetella* species or members of the *Aeromonadaceae*, and less than 10 % of *Burkholderia* species and members of the *Enterobacteriaceae*. It should be noted that similar to the situation in Gram-positive bacteria, the percentage of Gram-negative bacteria estimated to be confidently identified by the MicroSeq databases in this study is low compared to the accuracies reported for Gram-negative bacteria in the

literature (Tang *et al.*, 1998; Simmon *et al.*, 2006). This is mainly because the MicroSeq databases do not take into account bacteria with similar 16S rRNA gene sequences, which can confuse the identity of the bacterium. In fact, it is well known that the 16S rRNA gene is probably not the best target for identifying *Aeromonas* and *Bartonella* species and members of the *Enterobacteriaceae*, and alternative gene targets, such as *gyrB*, *rpoB*, *gltA* and *dnaJ*, have been used for their identification.

ACKNOWLEDGEMENTS

This work is partly supported by the HKSAR Research Fund for the Control of Infectious Diseases of the Health, Welfare and Food Bureau, Committee of Research and Conference Grants and University Development Fund, The University of Hong Kong.

REFERENCES

- Drancourt, M., Bollet, C., Carlioz, A., Martelin, R., Gayral, J. P. & Raoult, D. (2000). 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. *J Clin Microbiol* **38**, 3623–3630.
- Fontana, C., Favaro, M., Pelliccioni, M., Pistoia, E. S. & Favalli, C. (2005). Use of the MicroSeq 500 16S rRNA gene-based sequencing for identification of bacterial isolates that commercial automated systems failed to identify correctly. *J Clin Microbiol* **43**, 615–619.
- Janda, J. M. & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol* **45**, 2761–2764.
- Lau, S. K., Woo, P. C., Teng, J. L., Leung, K. W. & Yuen, K. Y. (2002). Identification by 16S ribosomal RNA gene sequencing of *Arcobacter butzleri* bacteraemia in a patient with acute gangrenous appendicitis. *Mol Pathol* **55**, 182–185.
- Lau, S. K., Woo, P. C., Mok, M. Y., Teng, J. L., Tam, V. K., Chan, K. K. & Yuen, K. Y. (2004). Characterization of *Haemophilus segnis*, an important cause of bacteremia, by 16S rRNA gene sequencing. *J Clin Microbiol* **42**, 877–880.
- Lau, S. K., Ng, K. H., Woo, P. C., Yip, K. T., Fung, A. M., Woo, G. K., Chan, K. M., Que, T. L. & Yuen, K. Y. (2006). Usefulness of the MicroSeq 500 16S rDNA bacterial identification system for identification of anaerobic Gram positive bacilli isolated from blood cultures. *J Clin Pathol* **59**, 219–222.
- Murray, P. R., Baron, E. J., Jorgensen, J. H., Landry, M. L. & Pfaller, M. A. (2007). *Manual of Clinical Microbiology*, 9th edn. Washington, DC: American Society for Microbiology.
- Simmon, K. E., Croft, A. C. & Petti, C. A. (2006). Application of SmartGene IDNS software to partial 16S rRNA gene sequences for a diverse group of bacteria in a clinical laboratory. *J Clin Microbiol* **44**, 4400–4406.
- Stackebrandt, E. & Goebel, B. M. (1994). Taxonomic note: a place for DNA–DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* **44**, 846–849.
- Tang, Y. W., Ellis, N. M., Hopkins, M. K., Smith, D. H., Dodge, D. E. & Persing, D. H. (1998). Comparison of phenotypic and genotypic techniques for identification of unusual aerobic pathogenic gram-negative bacilli. *J Clin Microbiol* **36**, 3674–3679.
- Teng, J. L., Woo, P. C., Leung, K. W., Lau, S. K., Wong, M. K. & Yuen, K. Y. (2003). Pseudobacteraemia in a patient with neutropenic fever caused by a novel paenibacillus species: *Paenibacillus hongkongensis* sp. nov. *Mol Pathol* **56**, 29–35.
- Woo, P. C., Cheung, E. Y., Leung, K. W. & Yuen, K. Y. (2001a). Identification by 16S ribosomal RNA gene sequencing of an *Enterobacteriaceae* species with ambiguous biochemical profile from a renal transplant recipient. *Diagn Microbiol Infect Dis* **39**, 85–93.
- Woo, P. C., Chong, K. T., Leung, K. W., Que, T. L. & Yuen, K. Y. (2001b). Identification of *Arcobacter cryaerophilus* isolated from a traffic accident victim with bacteremia by 16S ribosomal RNA gene sequencing. *Diagn Microbiol Infect Dis* **40**, 125–127.
- Woo, P. C., Leung, K. W., Tsoi, H. W., Wong, S. S., Teng, J. L. & Yuen, K. Y. (2002). Thermo-tolerant *Campylobacter fetus* bacteraemia identified by 16S ribosomal RNA gene sequencing: an emerging pathogen in immunocompromised patients. *J Med Microbiol* **51**, 740–746.
- Woo, P. C., Lau, S. K., Woo, G. K., Fung, A. M., Ngan, A. H., Hui, W. T. & Yuen, K. Y. (2003a). Seronegative bacteremic melioidosis caused by *Burkholderia pseudomallei* with ambiguous biochemical profile: clinical importance of accurate identification by 16S rRNA gene and *groEL* gene sequencing. *J Clin Microbiol* **41**, 3973–3977.
- Woo, P. C., Ng, K. H., Lau, S. K., Yip, K. T., Fung, A. M., Leung, K. W., Tam, D. M., Que, T. L. & Yuen, K. Y. (2003b). Usefulness of the MicroSeq 500 16S ribosomal DNA-based bacterial identification system for identification of clinically significant bacterial isolates with ambiguous biochemical profiles. *J Clin Microbiol* **41**, 1996–2001.
- Woo, P. C., Teng, J. L., Leung, K. W., Lau, S. K., Wong, M. K. & Yuen, K. Y. (2004). Bacteremia in a patient with colonic carcinoma caused by a novel *Sedimentibacter* species: *Sedimentibacter hongkongensis* sp. nov. *Diagn Microbiol Infect Dis* **50**, 81–87.
- Woo, P. C., Tse, H., Lau, S. K., Leung, K. W., Woo, G. K., Wong, M. K., Ho, C. M. & Yuen, K. Y. (2005). *Alkanindiges hongkongensis* sp. nov. A novel *Alkanindiges* species isolated from a patient with parotid abscess. *Syst Appl Microbiol* **28**, 316–322.
- Woo, P. C., Chung, L. M., Teng, J. L., Tse, H., Pang, S. S., Lau, V. Y., Wong, V. W., Kam, K. L., Lau, S. K. & Yuen, K. Y. (2007). *In silico* analysis of 16S ribosomal RNA gene sequencing-based methods for identification of medically important anaerobic bacteria. *J Clin Pathol* **60**, 576–579.
- Woo, P. C., Lau, S. K., Teng, J. L., Tse, H. & Yuen, K. Y. (2008). Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clin Microbiol Infect* **14**, 908–934.
- Woo, P. C., Teng, J. L., Wu, J. K., Leung, F. P., Tse, H., Fung, A. M., Lau, S. K. & Yuen, K. Y. (2009). Guidelines for interpretation of 16S rRNA gene sequence-based results for identification of medically important aerobic Gram-positive bacteria. *J Med Microbiol* **58**, 1030–1036.
- Yuen, K. Y., Woo, P. C., Teng, J. L., Leung, K. W., Wong, M. K. & Lau, S. K. (2001). *Laribacter hongkongensis* gen. nov., sp. nov., a novel gram-negative bacterium isolated from a cirrhotic patient with bacteremia and empyema. *J Clin Microbiol* **39**, 4227–4232.