Title	In silico analysis of 16S ribosomal RNA gene sequencing-based methods for identification of medically important anaerobic bacteria
Author(s)	Woo, PCY; Chung, LMW; Teng, JLL; Tse, H; Pang, SSY; Lau, VYT; Wong, VWK; Kam, KL; Lau, SKP; Yuen, KY
Citation	Journal Of Clinical Pathology, 2007, v. 60 n. 5, p. 576-579
Issued Date	2007
URL	http://hdl.handle.net/10722/57298
Rights	Journal of Clinical Pathology. Copyright © B M J Publishing Group.

In silico analysis of 16S ribosomal RNA gene sequencing-based methods for identification of medically important anaerobic bacteria

Patrick C Y Woo, Liliane M W Chung, Jade L L Teng, Herman Tse, Sherby S Y Pang, Veronica Y T Lau, Vanessa W K Wong, Kwok-ling Kam, Susanna K P Lau and Kwok-Yung Yuen

J. Clin. Pathol. 2007;60;576-579; originally published online 17 Oct 2006; doi:10.1136/jcp.2006.038653

In silico analysis of 16S ribosomal RNA gene sequencing-based methods for identification of medically important anaerobic bacteria

Patrick C Y Woo, Liliane M W Chung, Jade L L Teng, Herman Tse, Sherby S Y Pang, Veronica Y T Lau, Vanessa W K Wong, Kwok-ling Kam, Susanna K P Lau, Kwok-Yung Yuen

J Clin Pathol 2007;**60**:576–579. doi: 10.1136/jcp.2006.038653

This study is the first study that provides useful guidelines to clinical microbiologists and technicians on the usefulness of full 16S rRNA sequencing, 5′-end 527-bp 16S rRNA sequencing and the existing MicroSeq full and 500 16S rDNA bacterial identification system (MicroSeq, Perkin-Elmer Applied Biosystems Division, Foster City, California, USA) databases for the identification of all existing medically important anaerobic bacteria. Full and 527-bp 16S rRNA sequencing are able to identify 52–63% of 130 Gram-positive anaerobic rods, 72–73% of 86 Gram-negative anaerobic rods and 78% of 23 anaerobic cocci. The existing MicroSeq databases are able to identify only 19–25% of 130 Gram-positive anaerobic rods, 38% of 86 Gram-negative anaerobic rods and 39% of 23 anaerobic cocci. These represent only 45–46% of those that should be confidently identified by full and 527-bp 16S rRNA sequencing. To improve the usefulness of MicroSeq, bacterial species that should be confidently identified by full and/or 527-bp 16S rRNA sequencing but not included in the existing MicroSeq databases should be included.

Comparison of bacterial gene sequences has shown that 16S rRNA gene sequencing can be used as a working standard for the classification and identification of bacteria.[1] The MicroSeq 500 16S rDNA bacterial identification system (MicroSeq, Perkin-Elmer Applied Biosystems Division, Foster City, California) has been designed for rapid and accurate identification of bacterial pathogens, using the 5′-end 527-bp of the 16S rRNA gene.[2–7] Recently, the company has also included a full 16S rRNA gene sequence (full-MicroSeq) database (http://docs.appliedbiosystems.com/pebiodocs/00113462.pdf). Because identification of medically important anaerobic bacteria is notoriously difficult, 16S rRNA sequencing would be particularly useful for the identification of this group of bacteria.[8–12]

Problems exist in both 16S rRNA sequencing and MicroSeq. When two different bacterial species share almost the same 16S rRNA sequence, this technique would not be useful for distinguishing them. Moreover, MicroSeq is further limited by the database of the system.[7 13] In this study, we systematically evaluated the potential usefulness of full and 527-bp 16S rRNA sequencing and the existing MicroSeq databases for identification of all known medically important anaerobic bacterial species.

MATERIALS AND METHODS

16S rRNA sequences of medically important anaerobic bacteria

The medically important anaerobic bacterial species included in this study comprise all anaerobic bacterial species listed in the most recent edition of the *Manual of clinical microbiology*.[14] For each bacterial species, a list of the 16S rRNA sequence was retrieved from the GenBank database. In the list, the most representative 16S rRNA sequence for each species was chosen for analysis according to the following criteria: (1) strains with good phenotypic characterisation (eg, type strains); (2) strains isolated from humans; (3) sequences with fewer undetermined bases;and (4) longer sequences, especially those with better coverage of the 5′ end.

Comparison of full 16S rRNA sequences of medically important anaerobic bacteria

The percentage differences of the 16S rRNA sequences between the different species of medically important anaerobic bacteria were determined by pairwise alignment.[15] For sequences with undetermined bases, other 16S rRNA sequences of the same species were retrieved and the undetermined bases manually amended. If there was no other 16S rRNA sequence for the same species, the positions of the undetermined bases were deleted in the analysis.

Comparison of 527-bp 16S rRNA sequences of medically important anaerobic bacteria

The 527-bp 16S rRNA sequence that should be amplified by the primers of MicroSeq were extracted from the full 16S rRNA sequence. The percentage differences of the resultant partial 16S rRNA sequences between the different species were determined by pairwise alignment.[15]

RESULTS

Supplementary table 1 (available at http://jcp.bmj.com/supplemental) shows the percentage differences of the full and 527-bp 16S rRNA sequences between the different groups of medically important anaerobic bacteria. Full 16S rRNA sequencing should be useful for the identification of 21 of 42 *Clostridium* species, 47 of 88 non-sporulating Gram-positive rods, 13 of 15 *Bacteroides* species, 49 of 71 Gram-negative rods and 18 of 23 anaerobic cocci (supplementary tables 2–4, available at http://jcp.bmj.com/supplemental). For the existing full-MicroSeq database, it should be useful for the identification of 13 of 42 *Clostridium* species, 12 of 88 non-sporulating Gram-positive rods, 11 of 15 *Bacteroides* species, 22 of 71 Gram-negative rods and 9 of 23 anaerobic cocci. The 527-bp 16S rRNA sequencing should be useful for the identification of 23 of 42 *Clostridium* species, 59 of 88 non-sporulating Gram-positive rods, 13 of 15 *Bacteroides* species, 50 of 71 Gram-negative rods and 18 of 23 anaerobic cocci. For the existing MicroSeq database, it should be useful for the identification of 14 of 42 *Clostridium* species, 19 of 88 non-sporulating Gram-positive rods, 11 of 15 *Bacteroides* species, 22 of 71 Gram-negative rods and 9 of 23 anaerobic cocci.

Abbreviation: MicroSeq, MicroSeq 500 16S rDNA bacterial identification system

Table 1 Number and percentage of major groups of medically important anaerobic bacteria confidently identified by full 16S rRNA gene sequence, 5′-end 527-bp 16S rRNA gene sequence and the existing MicroSeq 16S rDNA bacterial identification system databases

Bacterial groups	Total no of species	Species confidently identified, n(%)			
		Full 16S rRNA gene sequencing	Existing MicroSeq full 16S rDNA bacterial identification system database	5′-end 527-bp 16S rRNA gene sequencing	Existing MicroSeq 500 16S rDNA bacterial identification system database
Anaerobic Gram-positive rods	130	68 (52)	25 (19)	82 (63)	33 (25)
Actinomyces	24	13 (54)	4 (17)	16 (67)	5 (21)
Bifidobacterium	8	3 (38)	0 (0)	6 (75)	2 (25)
Clostridium	42	21 (50)	13 (31)	23 (55)	14 (33)
Eubacterium	17	8 (47)	2 (12)	9 (53)	2 (12)
Lactobacillus	5	3 (60)	1 (20)	3 (60)	1 (20)
Anaerobic Gram-negative rods	86	62 (72)	33 (38)	63 (73)	33 (38)
Bacteroides	15	13 (87)	11 (73)	13 (87)	11 (73)
Fusobacterium	11	1 (9)	1 (9)	1 (9)	1 (9)
Porphyromonas	11	9 (82)	2 (18)	11 (100)	2 (18)
Prevotella	20	20 (100)	10 (50)	18 (90)	9 (45)
Selenomonas	5	2 (40)	1 (20)	3 (60)	1 (20)
Anaerobic cocci	23	18 (78)	9 (39)	18 (78)	9 (39)
Anaerococcus	6	6 (100)	4 (67)	6 (100)	4 (67)
Peptoniphilus	5	3 (60)	1 (20)	3 (60)	1 (20)

DISCUSSION

This study is the first to provide useful guidelines to clinical microbiologists and technicians on the usefulness of 16S rRNA sequencing for the identification of medically important anaerobic bacteria. Interpretation of 16S rRNA gene sequence results is often difficult for those with limited experience in the use of this technique for the identification of pathogenic bacteria. Owing to the large number of unvalidated 16S rRNA sequences in GenBank, inexperienced users often find it difficult to decide whether the ''first hit'' or ''closest match'' is the one that corresponds to the real identity of a bacterium. As for commercially available databases such as MicroSeq, their usefulness is largely limited by (a) the limited database and (b) the fact that the database also includes those bacterial species that obviously cannot be identified confidently by 16S rRNA sequencing, but no guideline is given on the limited usefulness of 16S rRNA sequencing for identification of such species. Therefore, we undertook this study, by using ''the most representative'' 16S rRNA sequences for each medically important anaerobic bacterium in GenBank, and analysing the usefulness of different forms of such a technique for the identification of medically important anaerobic bacteria.

Overall, both full and 527-bp 16S rRNA sequencing are very useful for the identification of medically important anaerobic bacteria to the genus level, but are only able to identify 62% and 68% of these bacteria confidently to the species level. In general, 16S rRNA sequencing is more useful for the identification of medically for important anaerobic Gram-negative rods and cocci than for Gram-positive rods (table 1). However, it is not particularly useful for speciation of *Fusobacterium* species, with only 1 of 11 medically important *Fusobacterium* species being identified confidently to the species level. This is of major clinical relevance because *F necrophorum*, a virulent bacterium that causes peritonsillar abscess, is associated with serious complications, including jugular vein septic thrombophlebitis (Lemierre syndrome), lung abscess and empyema. Of note is that the 16S rRNA sequences is not able to confidently speciate *Clostridium botulinum*, *C septicum*, *C tertium* and *C tetani*, which are associated with important clinical syndromes.

The existing MicroSeq databases need to be markedly expanded. Overall, the existing MicroSeq databases are able to confidently identify only 19–25% Gram-positive anaerobic rods, 38% Gram-negative anaerobic rods and 39% anaerobic cocci (table 1). These represent only 45–46% of those that should be confidently identified by full and 527-bp rRNA sequencing. When compared with the manual interpretation of 16S rRNA sequencing results, the MicroSeq databases are particularly not good for some major genera, such as Actinomyces, Bifidobacterium, Eubacterium, Porphyromonas and Prevotella (table 1). To improve the usefulness of the MicroSeq databases, bacterial species that should be confidently identified by 16S rRNA sequencing but are not included in the existing MicroSeq databases should be included (table 2).

ACKNOWLEDGEMENTS

This work was partly supported by the Committee of Research and Conference Grants and University Development Fund, The University of Hong Kong, Hong Kong.

Supplementary tables are available at http://jcp.bmj.com/supplemental

Take-home messages

- Full and 5′-end 527-bp 16S rRNA sequencing are able to identify 52–63%, 72–73% and 78% of medically important Gram-positive anaerobic rods, Gram-negative anaerobic rods and anaerobic cocci, respectively.
- The existing MicroSeq databases are able to identify only 19–25%, 38% and 39% of medically important Gram-positive anaerobic rods, Gram-negative anaerobic rods and anaerobic cocci, respectively, representing only 45–46% of those that should be confidently identified by full and 5′-end 527-bp 16S rRNA sequencing.
- To improve the usefulness of MicroSeq, bacterial species that should be confidently identified by full and 5′-end 527-bp 16S rRNA sequencing but not currently in the existing MicroSeq databases should be included.

Table 2 Medically important anaerobic bacteria that should be confidently identified by full or 5′-end 527-bp 16S rRNA gene sequencing but are not included in the existing MicroSeq 16S rDNA bacterial identification system databases

Bacterial groups	Bacterial species: Not included in existing MicroSeq full 16S rDNA bacterial identification system database	Bacterial species: Not included in existing MicroSeq 500 16S rDNA bacterial identification system database
Anaerobic Gram-positive rods		
Actinobaculum	*A schaalii*	*A schaalii*
	A suis	*A suis*
Actinomyces	*A canis*	*A canis*
	A catuli	*A catuli*
	A denticolens	*A denticolens*
	A europaeus	*A europaeus*
	A gerencseriae	*A funkei*
	A graevenitzii	*A georgiae*
	A hordeovulneris	*A gerencseriae*
	A israelii	*A graevenitzii*
	A radicidentis	*A hordeovulneris*
		A hyovaginalis
		A radicidentis
Arcanobacterium	*A pluranimalium*	*A pluranimalium*
Atopobium	*A parvulum*	*A parvulum*
	A vaginae	*A vaginae*
Bifidobacterium	*B adolescentis*	*B adolescentis*
	B bifidum	*B breve*
	B globosum	*B dentium*
		B globosum
Bulleidia	*B extructa*	*B extructa*
Catenibacterium	*C mitsuokai*	*C mitsuokai*
Clostridium	*C aminovalericum*	*C aminovalericum*
	C coccoides	*C coccoides*
	C glycolicum	*C disporicum*
	C hiranonis	*C glycolicum*
	C hylemonae	*C hiranonis*
	C indolis	*C hylemonae*
	C spiroforme	*C indolis*
	C symbiosum	*C spiroforme*
		C symbiosum
Collinsella	*C aerofaciens*	*C aerofaciens*
Cryptobacterium	*C curtum*	*C curtum*
Eubacterium	*E brachy*	*E brachy*
	E saburreum	*E saburreum*
	E combesi	*E combesii*
	E minutum	*E minutum*
	E nodatum	*E nodatum*
	E saphenum	*E saphenum*
		E yurii subsp. *schtitka*
Holdemania	*H filiformis*	*H filiformis*
Lactobacillus	*L catenaformis*	*L catenaformis*
	L vitulinus	*L vitulinus*
Mogibacterium		*M timidum*
Olsenella	*O uli*	*O uli*
Propionibacterium	*P granulosum*	*P granulosum*
Pseudoramibacter	*P alactolyticus*	*P alactolyticus*
Slackia	*S exigua*	*S exigua*
	S heliotrinireducens	*S heliotrinireducens*
Anaerobic Gram-negative rods		
Alistipes	*A putredinis*	*A putredinis*
Anaerobiospirillum	*A succiniciproducens*	*A succiniciproducens*
	A thomasii	*A thomasii*
Bacteroides	*B capillosus*	*B capillosus*
	B splanchnicus	*B splanchnicus*
Bilophila	*B wadsworthia*	*B wadsworthia*
Butyrivibrio	*B fibrisolvens*	*B fibrisolvens*
Desulfovibrio	*D piger*	*D piger*
Dialister	*D pneumosintes*	*D pneumosintes*
Faecalibacterium	*F prausnitzii*	*F prausnitzii*
Filifactor	*F alocis*	
Johnsonella	*J ignava*	*J ignava*
Leptotrichia	*L buccalis*	*L buccalis*
Megamonas	*M hypermegale*	*M hypermegale*
Porphyromonas	*P asaccharolytica*	*P asaccharolytica*
	P cangingivalis	*P cangingivalis*
	P canoris	*P canoris*
	P cansulci	*P cansulci*
	P endodontalis	*P endodontalis*
	P levii	*P gingivalis*
	P macacae	*P gulae*
		P levii
		P macacae
Prevotella	*P bivia*	*P bivia*
	P dentalis	*P dentalis*
	P enoeca	*P enoeca*
	P intermedia	*P intermedia*
	P loescheii	*P loescheii*
	P melaninogenica	*P melaninogenica*
	P nigrescens	*P pallens*
	P pallens	*P tannerae*
	P tannerae	*P veroralis*
	P veroralis	
Selenomonas	*S noxia*	*S flueggei*
		S noxia
Sneathia	*S sanguinegens*	*S sanguinegens*
Succinivibrio	*S dextrinosolvens*	*S dextrinosolvens*
Sutterella	*S wadsworthensis*	*S wadsworthensis*
Tannerella	*T forsythensis*	*T forsythensis*
Anaerobic cocci		
Acidaminococcus	*A fermentans*	*A fermentans*
Anaerococcus	*A lactolyticus*	*A lactolyticus*
	A octavius	*A octavius*
Centipeda	*C periodontii*	*C periodontii*
Finegoldia	*F magna*	*F magna*
Gallicola	*G barnesae*	*G barnesae*
Megasphaera	*M elsdenii*	*M elsdenii*
Peptoniphilus	*P harei*	*P harei*
	P ivorii	*P ivorii*

Authors' affiliations

Patrick C Y Woo, Liliane M W Chung, Jade L L Teng, Herman Tse, Sherby S Y Pang, Veronica Y T Lau, Vanessa W K Wong, Kwok-ling Kam, Susanna K P Lau, Kwok-Yung Yuen, Department of Microbiology, The University of Hong Kong, Hong Kong
Patrick C Y Woo, Herman Tse, Susanna K P Lau, Kwok-Yung Yuen, Research Centre of Infection and Immunology, The University of Hong Kong, Hong Kong; State Key Laboratory of Emerging Infectious Diseases, The University of Hong Kong, Hong Kong

Competing interests: None.

Correspondence to: Dr K-Y Yuen, Department of Microbiology, The University of Hong Kong, University Pathology Building, Queen Mary Hospital Compound, Pokfulam Road, Hong Kong; hkumicro@hkucc.hku.hk

Accepted 18 September 2006
Published Online First 17 October 2006

REFERENCES

1 **Olsen GJ**, Woese CR. Ribosomal RNA: a key to phylogeny. *FASEB J* 1993;**7**:113–23.
2 **Cloud JL**, Conville PS, Croft A, *et al*. 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* 2004;**42**:578–84.
3 **Fontana C**, Favaro M, Pelliccioni M, *et al*. 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* 2005;**43**:615–19.
4 **Patel JB**, Leonard DG, Pan X, *et al*. Sequence-based identification of Mycobacterium species using the MicroSeq 500 16S rDNA bacterial identification system. *J Clin Microbiol* 2000;**38**:246–51.

5 **Tang YW**, Ellis NM, Hopkins MK, *et al*. Comparison of phenotypic and genotypic technique for identification of unusual aerobic pathogenic Gram-negative bacilli. *J Clin Microbiol* 1998;**36**:3674–9.
6 **Tang YW**, Von Graevenitz A, Waddington MG, *et al*. Identification of coryneform bacterial isolates by ribosomal DNA sequence analysis. *J Clin Microbiol* 2000;**38**:1676–8.
7 **Woo PCY**, Ng KHL, Lau SKP, *et al*. 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* 2003;**41**:1996–2001.
8 **Lau SKP**, Woo PCY, Woo GKS, *et al*. Eggerthella hongkongensis sp nov and *Eggerthella sinensis* sp nov, two novel *Eggerthella* species, account for half of the cases of *Eggerthella* bacteremia. *Diagn Microbiol Infect Dis* 2004;**49**:255–63.
9 **Woo PCY**, Fung AMY, Lau SKP, *et al*. Identification by 16S ribosomal RNA gene sequencing of *Lactobacillus salivarius* bacteremic cholecystitis. *J Clin Microbiol* 2002;**40**:265–7.
10 **Woo PCY**, Fung AMY, Lau SKP, *et al*. *Actinomyces hongkongensis* sp nov. A novel *Actinomyces* species isolated from a patient with pelvic actinomycosis. *Syst Appl Microbiol* 2003;**26**:518–22.
11 **Woo PCY**, Teng JLL, Leung KW, *et al*. Bacteremia in a patient with colonic carcinoma caused by a novel *Sedimentibacter* species: *Sedimentibacter hongkongensis* sp nov *Diagn Microbiol Infect Dis* 2004;**50**:81–7.
12 **Woo PCY**, Lau SKP, Chan KM, *et al*. *Clostridium* bacteraemia characterised by 16S ribosomal RNA gene sequencing. *J Clin Pathol* 2005;**58**:301–7.
13 **Lau SKP**, Ng KHL, Woo PCY, *et al*. Usefulness of MicroSeq 500 16S rDNA bacterial identification system for identification of anaerobic Gram-positive bacilli isolated from blood cultures. *J Clin Pathol* 2006;**59**:219–22.
14 **Murray PR**, Baro EJ, Jorgensen JH, *et al*. *Manual of clinical microbiology*, 8th edn. Washington, DC: American Society for Microbiology, 2003.
15 **Thompson JD**, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;**22**:4673–80.