 10.1111/j.1469-0691.2008.02070.x

Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories

P. C. Y. Woo[1,2,3], S. K. P. Lau[1,2,3*], J. L. L. Teng[3], H. Tse[1,2,3] and K.-Y. Yuen[1,2,3]*

[1]State Key Laboratory of Emerging Infectious Diseases, [2]Research Centre of Infection and Immunology and [3]Department of Microbiology, The University of Hong Kong, Hong Kong, China

ABSTRACT

In the last decade, as a result of the widespread use of PCR and DNA sequencing, 16S rDNA sequencing has played a pivotal role in the accurate identification of bacterial isolates and the discovery of novel bacteria in clinical microbiology laboratories. For bacterial identification, 16S rDNA sequencing is particularly important in the case of bacteria with unusual phenotypic profiles, rare bacteria, slow-growing bacteria, uncultivable bacteria and culture-negative infections. Not only has it provided insights into aetiologies of infectious disease, but it also helps clinicians in choosing antibiotics and in determining the duration of treatment and infection control procedures. With the use of 16S rDNA sequencing, 215 novel bacterial species, 29 of which belong to novel genera, have been discovered from human specimens in the past 7 years of the 21st century (2001–2007). One hundred of the 215 novel species, 15 belonging to novel genera, have been found in four or more subjects. The largest number of novel species discovered were of the genera *Mycobacterium* (n = 12) and *Nocardia* (n = 6). The oral cavity/dental-related specimens (n = 19) and the gastrointestinal tract (n = 26) were the most important sites for discovery and/or reservoirs of novel species. Among the 100 novel species, *Streptococcus sinensis*, *Laribacter hongkongensis*, *Clostridium hathewayi* and *Borrelia spielmanii* have been most thoroughly characterized, with the reservoirs and routes of transmission documented, and *S. sinensis*, *L. hongkongensis* and *C. hathewayi* have been found globally. One of the greatest hurdles in putting 16S rDNA sequencing into routine use in clinical microbiology laboratories is automation of the technology. The only step that can be automated at the moment is input of the 16S rDNA sequence of the bacterial isolate for identification into one of the software packages that will generate the result of the identity of the isolate on the basis of its sequence database. However, studies on the accuracy of the software packages have given highly varied results, and interpretation of results remains difficult for most technicians, and even for clinical microbiologists. To fully utilize 16S rDNA sequencing in clinical microbiology, better guidelines are needed for interpretation of the identification results, and additional/supplementary methods are necessary for bacterial species that cannot be identified confidently by 16S rDNA sequencing alone.

Keywords 16S rDNA gene, bacteria, discovery, identification, review, sequencing

Accepted: 11 June 2008

Clin Microbiol Infect 2008; **14:** 908–934

INTRODUCTION

One and a half centuries after the 22-year-old Charles Darwin's 5-year voyage on HMS *Beagle*, the analysis of small rDNA gene sequences is another important landmark in the study of the evolution and classification of living organisms. Traditionally, living organisms were classified, according to similarities and differences in their phenotypic characteristics, into prokaryotes and

Corresponding author and reprint requests: K. Y. Yuen, State Key Laboratory of Emerging Infectious Diseases, Department of Microbiology, The University of Hong Kong, University Pathology Building, Queen Mary Hospital, Hong Kong, China
E-mail: hkumicro@hkucc.hku.hk

*P. C. Y. Woo and S. K. P. Lau contributed equally to the article.

eukaryotes, and these were in turn further classified into various kingdoms, phyla, classes, orders, families, genera and species. However, objective taxonomic classification by these methods can be difficult because of variations in phenotypic characteristics. Three decades ago, Carl Woese and others started to analyse and sequence the 16S rDNA genes of various bacteria, using DNA sequencing, a state-of-the-art technology at that time, and used the sequences for phylogenetic studies [1,2]. The invention of PCR and automated DNA sequencing two decades ago, and subsequent work on 16S rDNA sequencing of bacteria, as well as 18S rDNA sequencing of eukaryotes, has led to the accumulation of a vast amount of sequence data on the rDNA genes of the smaller subunit of the ribosomes in a large number of living organisms. Comparison of these sequences has shown that the rDNA gene sequences are highly conserved within living organisms of the same genus and species, but that they differ between organisms of other genera and species. Using these rDNA gene sequences for phylogenetic studies, three domains of life, Archaea, Bacteria and Eukarya, as opposed to the traditional classification of living organisms into prokaryotes and eukaryotes only, were described [3].

Among the three domains of life, the largest amount of rDNA gene sequencing work concerns bacteria. Using 16S rDNA sequences, numerous bacterial genera and species have been reclassified and renamed, classification of uncultivable bacteria has been made possible, phylogenetic relationships have been determined, and the discovery and classification of novel bacterial species has been facilitated. In the last decade, sequencing of various bacterial genomes and comparison between genome and 16S rDNA gene phylogeny has confirmed the representativeness of the 16S rDNA gene in bacterial phylogeny [4].

As a result of the increasing availability of PCR and DNA sequencing facilities, the use of 16S rDNA sequencing has not been limited to research purposes, but has also been exploited in clinical microbiology laboratories. Accurate identification of bacterial isolates is one of the most important functions of clinical microbiology laboratories. On a patient-to-patient basis, accurate identification is crucial in determining whether the isolate is causing genuine infection or is simply a colonizer or contaminant, the choice and duration of antibiotic treatment, and the appropriate infection control procedures. On a population scale, accurate identification is important in analysing the epidemiology, antibiotic resistance patterns, treatment plans and outcomes of infections associated with a particular bacterium. Traditionally, identification of bacteria in clinical microbiology laboratories was performed using phenotypic tests, including Gram smear and biochemical tests, taking into account culture requirements and growth characteristics. However, these methods of bacterial identification have major limitations. First, organisms with biochemical characteristics that do not fit into the patterns of any known genus and species are occasionally encountered. Second, they cannot be used for uncultivable organisms. Third, identification of some particular groups of bacteria, such as anaerobes and mycobacteria, would require additional equipment and expertise, which are not available in most clinical microbiology laboratories. Using 16S rDNA sequencing, these problems can be overcome by a single technology, which also facilitates the discovery of novel genera and species.

In this article, the use of 16S rDNA sequencing in clinical microbiology laboratories for bacterial identification, the discovery of novel bacterial genera and species, the detection of uncultivable bacteria and the diagnosis of culture-negative infections are reviewed. Automation of 16S rDNA sequencing and the usefulness and limitations of 16S rDNA sequencing in clinical microbiology laboratories are also discussed. The technological aspects of 16S rDNA sequencing are not included in this review.

METHODS

In the section on 'Novel bacterial genus and species discovery', novel species discovered from human specimens that were reported during the first 7 years of the 21st century (2001–2007) in the English literature are reviewed. For initial screening, 'gen. nov.' and 'sp. nov.' were used as the key words for a Medline search. The results were manually screened for novel bacterial species isolated from human specimens, regardless of the criteria used for their definition. Bacteria that were discovered only in non-human specimens were excluded. Bacteria that were discovered, characterized and named before 2001, but were renamed after 2001 (e.g. *Streptococcus lutetiensis* and *S. pasteurianus* [5]), and those that were further subclassified (e.g. *S. pseudopneumoniae* [6]) because of acquisition of new information such as 16S rDNA sequences or phenotypic test results, were also not included.

Additional reports on infections associated with each of the novel species isolated from human specimens, as well as those concerning their isolation from non-human specimens, were searched. All novel species that have been found in four or more subjects are presented.

USE OF 16S RDNA SEQUENCING IN CLINICAL MICROBIOLOGY LABORATORIES

Bacterial identification

Rapid and accurate identification of bacterial isolates is a fundamental task in clinical microbiology, and provides insights into aetiologies of infectious disease and appropriate antibiotic treatment. Although conventional phenotypic methods are relatively inexpensive and allow identification of most commonly encountered bacteria, certain groups of bacteria are difficult to identify, and special equipment and expertise may be required, e.g. gas chromatography–mass spectrometry for anaerobes. These methods also fail in cases of rare bacteria or bacteria with ambiguous profiles. Moreover, phenotypic methods rely on the availability of pure culture and are dependent on subsequent growth characteristics and biochemical profiling. Therefore, considerable time is required for slow-growing bacteria to be identified. 16S rDNA sequencing represents a universal technology that, theoretically, provides solutions to these problems, yielding unambiguous data, even for unusual and slow-growing isolates, often within 48 h, which are reproducible among laboratories.

Identification of rare bacteria and bacteria with unusual phenotypic profiles. 16S rDNA sequencing is particularly useful in identifying unusual bacteria that are difficult to identify by conventional methods, providing genus identification in >90% of cases, and identification of 65–83% of these at the species level [7,8]. The MicroSeq 500 16S rDNA-based identification system was also able to identify 81% of clinically significant bacterial isolates with ambiguous biochemical profiles and 89.2% of unusual aerobic Gram-negative bacilli to the species level [9,10].

In many situations, 16S rDNA sequencing is the ultimate solution to identification of aetiological agents of infectious diseases caused by rarely encountered bacteria [11–22]. This not only allows correct identification and selection of appropriate treatment, but also contributes to a better understanding of the epidemiology and pathogenic role of these bacteria, which has not been possible in the past. For example, using 16S rDNA sequencing, cases of invasive *Streptococcus iniae* infections, which had previously been reported only in North America, have been recognized in Asia [23–27].

Unlike phenotypic identification, which can be affected by the presence or absence of non-housekeeping genes or by variability in expression of characters, 16S rDNA sequencing provides accurate identification of isolates with atypical phenotypic characteristics. Using the technique, it has been possible to identify thermotolerant *Campylobacter fetus* strains as important causes of bacteraemia in immunocompromised patients [28]. Similar applications have also been shown to have significant impacts on the decision of whether to prescribe antibiotic treatment [29–31] and on the choice of specific antibiotic regimen [32–34], which could lead to improved clinical outcomes [10].

Mistakes in identifying rarely encountered or phenotypically aberrant isolates are probably quite common in clinical microbiology laboratories. Sometimes, it is even difficult to know whether a bacterium has been incorrectly identified. For example, using phenotypic methods, *Francisella tularensis* subsp. *novicida* was consistently misidentified twice, as *Neisseria meningitidis* or *Actinobacillus actinomycetemcomitans* [35]. 16S rDNA sequencing will continue to play a major role in the identification of rare bacteria and bacteria with ambiguous characteristics in clinical microbiology laboratories.

Identification of slow-growing bacteria. 16S rDNA sequencing and similar molecular identification methods have the additional advantage of reducing the time required to identify slow-growing bacteria such as mycobacteria, which may take 6–8 weeks to grow in culture sufficiently for phenotypic tests to be performed. Even for rapidly growing mycobacteria, some biochemical reactions may take up to 28 days to complete. As for whole cell fatty acid analysis by gas chromatography, this requires special equipment and expertise that are often not available in clinical microbiology laboratories. Therefore, 16S rDNA sequencing has been used for identification of *Mycobacterium* species isolated from

clinical specimens, making clinical diagnosis more rapid and guiding prompt antibiotic treatment [29,36]. Using this technique, we were able to describe a novel clinical syndrome, acupuncture mycobacteriosis, caused by relatively alcohol-resistant mycobacteria in patients receiving acupuncture [37,38]. However, a limited number of mycobacterial species could not be differentiated from one another by 16S rDNA sequencing, e.g. *Mycobacterium avium intracellulare* and *M. paratuberculosis*, *M. chelonae* and *M. abscessus*, and members of the *M. tuberculosis* complex. Therefore, other gene targets, sometimes supplemented by phenotypic results, have to be used for differentiation of these species, e.g. *hsp65* and *rpoB* for those that grow rapidly, and 16S–23S rRNA internal transcribed spacer or *gyrB* for the mycobacteria that grow slowly [39–42].

Routine identification. Several studies have been conducted to compare the usefulness of 16S rDNA sequencing with conventional or commercial methods for the identification of various groups of medically important bacteria. In general, 16S rDNA sequencing results in a higher percentage of species identification than conventional or commercial methods. The success rate of species identification by 16S rDNA sequencing ranged from 62% to 92%, depending on the group of bacteria and the criteria used for species definition [43–49]. However, there are 'blind spots' within some major genera, in which 16S rDNA sequences are not sufficiently discriminative for the identification of certain species. In these circumstances, alternative targets have to be investigated. For example, *groEL* is a commonly used essential gene other than 16S rDNA which is useful for classification and identification of many groups of bacteria, e.g. staphylococci and *Burkholderia* species [50,51]. In particular, it has been shown to be useful in differentiating *Burkholderia pseudomallei* from *B. thailandensis*, the 16S rDNA sequences of which are indistinguishable [52,53]. Similarly, 16S rDNA sequencing has limited discriminatory power for closely related *Staphylococcus* species. Therefore, sequencing of *groEL* and *tuf* has been proposed as being more reliable for identification of coagulase-negative staphylococci [47,54].

As simple conventional methods are often available for commonly encountered bacteria, 16S rDNA sequencing for identification of 'routine' bacterial strains is most useful in the context of bacterial species that are often difficult to identify with phenotypic tests. This has led to a better understanding of the epidemiology and pathogenicity of these clinically 'unidentifiable' bacteria. One remarkable example is provided by anaerobic Gram-positive rods, where conventional methods are simply not reliable, even for genus identification; with the use of 16S rDNA sequencing, many previously undescribed or ignored anaerobic bacterial species were found to contribute to cases of bacteraemia [55–59]. With regard to streptococci, cases of β-haemolytic Lancefield group G streptococcus bacteraemia have been found to be almost exclusively caused by *S. dysgalactiae* subsp. *equisimilis* using 16S rDNA sequencing, although *S. canis* infections in dog owners have been reported rarely [60,61]. In another study, six cases of '*S. milleri*' endocarditis were attributed to *S. anginosus*, suggesting that *S. anginosus* had the highest propensity to cause infective endocarditis among the three species of the '*S. milleri* group' [62]. Using this 'state-of-the-art' technique, it was also found that *Haemophilus segnus* is an important cause of non-*H. influenzae* bacteraemia [19,63,64]. In addition, the use of 16S rDNA sequencing for clinically 'unidentifiable' bacteria could have clinical significance regarding the choice of antibiotic regimen as well as the duration of treatment. For example, we have previously reported a case of empyema thoracis caused by *Enterococcus cecorum* that was unidentified by conventional methods and three commercial identification systems. Unlike other *Enterococcus* species, the organism is known to be susceptible to cefotaxime and ceftriaxone. In this case, the use of 16S rDNA sequencing not only allowed continuation of cefotaxime as treatment for the patient, who responded very well, but also suggested that the bacterium's cephalosporin susceptibility could well be explained by its unique phylogenetic position, it being the ancestor of other *Enterococcus* species and more closely related to *Streptococcus* species [65]. In other instances, 16S rDNA sequencing has been particularly useful in differentiating between *Actinomyces* and non-*Actinomyces* anaerobic Gram-positive bacilli, which is often difficult in clinical microbiology laboratories [66–68]. A definitive diagnosis of actinomycosis is important, as a prolonged course of appropriate antibiotic

treatment (weeks to months) is indicated to prevent relapse.

Novel bacterial genus and species discovery

During the process of bacterial identification, potential novel species will be encountered when there is a significant difference between the phenotypic characteristics and/or 16S rDNA sequences of the unknown bacterium and those of the most closely related ones. As no single test or gene sequence is ideal for the definition of new species in all groups of bacteria, a polyphasic approach is usually used when a novel species is defined. Depending on the group of bacteria, this involves various combinations of phenotypic characteristics, 16S rDNA sequences, sequences of other housekeeping genes, and DNA–DNA hybridization results. In this section, we review the novel species discovered from human specimens in the first 7 years of the 21st century (2001–2007) in the English literature.

According to the inclusion and exclusion criteria used in the present article, 215 novel species, 29 of which belonged to novel genera, have been discovered from human specimens in the past 7 years. Among these 215 novel species, 100 (15 of novel genera) have been found in four or more subjects (Table 1). Among these 100 novel species, there were four (4%) aerobic Gram-positive cocci, 30 (30%) aerobic Gram-positive rods, three (3%) aerobic Gram-negative cocci, 24 (24%) aerobic Gram-negative rods, four (4%) anaerobic Gram-positive cocci, ten (10%) anaerobic Gram-positive rods, one (1%) anaerobic Gram-negative coccus, 22 (22%) anaerobic Gram-negative rods, and two (2%) spirochaetes. The largest numbers of novel species discovered and isolated from four or more subjects were of the genera *Mycobacterium* ($n = 12$) and *Nocardia* ($n = 6$). The oral cavity/dental-related specimens ($n = 19$) and the gastrointestinal tract specimens ($n = 26$) were the most important for discovery and/or the most important reservoirs of novel species. This is in line with the huge diversity of potential novel bacterial species in the human oral cavity and gastrointestinal tract [69,70]. Among the 100 novel species, *S. sinensis*, *Laribacter hongkongensis*, *Clostridium hathewayi* and *Borrelia spielmanii* have been most thoroughly characterized, with the reservoirs and routes of transmission documented, and *S. sinensis*, *L. hongkongensis* and *C. hathewayi* have been found globally [71–89], although most *Nocardia* and *Mycobacterium* species were probably from the environment, and most anaerobes were probably from the oral cavity and/or gastrointestinal tract.

In our opinion, novel bacterial species recovered from human specimens should be reported, even if only one well-characterized strain is available. Despite extensive investigations, a microbiological cause cannot be determined in approximately half of the patients with infectious disease. For some clinical syndromes, such as neutropenic fever, no microbiological cause can be found in up to 80% of patients who are believed to be suffering from infective causes [90,91]. For some other syndromes, e.g. acute gastroenteritis and community-acquired pneumonia, the cause was undetermined in *c.* 40% of patients [92,93]. Over the years, tremendous efforts have been made to determine the microorganisms associated with these 'unexplained infectious disease syndromes'. The discovery of novel aetiological agents responsible for 'unexplained infectious disease syndromes' relies quite heavily on the description of novel microbes. Although standard or unusual forms of known microbes are sometimes considered to be novel causes of 'unexplained infectious disease syndromes' [94,95], most of the novel causes are indeed previously undescribed microbes. There has been much debate on the number of strains of a particular bacterium required for description of a novel bacterial species, but our experience with the discovery and characterization of *L. hongkongensis* and *S. sinensis* suggests that any well-characterized strain of a novel bacterial species recovered from human specimens should be reported. Concerning *L. hongkongensis*, when the bacterium was first discovered, only one strain was obtained [76]. The description of the novel bacterium and the wide availability of molecular techniques, sophisticated databases and bioinformatics tools have made discovery of additional strains from other countries and rapid sharing of information possible [77]. This has led to rapid confirmation of its association with gastroenteritis and the determination of its reservoir [79]. In retrospect, if the first strain of *L. hongkongensis* had not been described, the demonstration of its association with

Table 1. Novel bacterial species discovered in four or more human subjects in the 21st century

Bacteria	Year of first publication	References	Associated diseases	Source of isolation: Human: Patients	Source of isolation: Human: Asymptomatic individuals	Source of isolation: Non-human	Reservoir/ source of infection
Aerobic and facultative anaerobic Gram-positive cocci							
Kytococcus schroeteri	2002	[159–162]	Infective endocarditis, pneumonia	Blood, BAL, prosthetic valve, abscess pus	–	–	–
Streptococcus oligofermentans	2003	[163]	–	–	Dental plaque, saliva	–	–
S. pseudoporcinus	2006	[164]	–	Female genitourinary tract	–	–	–
S. sinensis	2002	[71–75]	Infective endocarditis	Blood	Saliva	–	Oral cavity
Aerobic and facultative anaerobic Gram-positive rods							
Actinomyces cardiffensis	2002	[165]	Postmastoidectomy brain and ear abscesses, jaw abscess, pericolic abscess, sinusitis, pelvic actinomycosis	Abscess pus, pleural fluid, antral washout, IUCD	–	–	–
A. funkei	2002	[166,167]	–	Blood, sternum, abdominal incision, neck–face, thorax, pelvis, IUCD, vagina/penis, brain/CSF, superficial soft tissue lesion	–	–	–
Bifidobacterium scardovii	2002	[168]	–	Blood, urine, hip	–	–	–
Brevibacterium lutescens	2003	[169]	Peritonitis	Peritoneal fluid, infected ear discharge, peritoneal dialysate fluid	–	Peptone preparation	–
Brevibacterium paucivorans	2001	[170]	–	CSF, groin abscess, blood, ear discharge, intravascular catheter, wound, groin swab	–	–	–
Corynebacterium aurimucosum (= *Corynebacterium nigricans*)	2002	[171–173]	Meningitis, uterine infection, UTI, vaginitis, placental infection, vulval ulcer, female urogenital tract infection	Cervix, bartholin gland, vagina, CSF, endometrium, placenta, urine, vulva, amniotic fluid, blood, vaginal swab, vaginal ulcer, vaginal–rectal swab	–	–	–
Corynebacterium freneyi	2001	[174,175]	Bacteraemia	Pus from toe, varicose ulcer, subcutaneous abscess fistula, blood	–	–	–
Corynebacterium resistens	2005	[176]	Bacteraemia, nosocomial pneumonia, cellulitis	Blood, bronchial aspirate, cellulitis aspirate	–	–	–
Lactobacillus coleohominis	2001	[177]	–	–	Vagina, urine, cervix	–	–
Microbacterium paraoxydans	2003	[178]	Catheter-related bacteraemia	Blood, bronchial aspirate	–	Vegetable	–
Mycobacterium arupense	2006	[179]	–	Respiratory specimen, gastrointestinal tract, various sterile sites, e.g. tendon, lymph node, lung biopsy, pleural fluid, surgical tissue, urine	–	–	Possibly environmental
M. bolletii	2006	[42]	Chronic pneumonia	Sputum, stomach aspirate	–	–	–
M. canariasense	2004	[180]	Catheter-related bacteraemia	Blood	–	–	–
M. colombiense	2006	[181]	–	Sputum, blood	–	–	–
M. florentinum	2005	[182]	Lymphadenopathy, pneumonia	Sputum, stool, lymph node, gastric aspirate, BAL	–	–	–
M. holsaticum	2002	[183,184]	–	Sputum, urine, gastric fluid	–	Lymph node of cattle with evidence of bovine tuberculosis	–
M. immunogenum	2001	[185–190]	Disseminated cutaneous infection, keratitis, catheter-related infection, septic arthritis, chronic pneumonia, pacemaker-related sepsis, hypersensitivity pneumonitis, chronic leg ulcer, peritonitis	Skin, cornea, urine, intravenous catheter site, joint fluid, BAL, blood, broviac site, leg biopsy, corneal scraping, peritoneal fluid of patients receiving IPD	–	Hospital environment, metalworking fluid, environment, tap water, water in IPD machines	Probably tap water
M. massiliense	2004	[191–193]	Pneumonia, post-intramuscular injection abscess, pacemaker pocket infection	Sputum, BAL, pus, tissue debris, specimen from surgical debridement, blood	–	–	–

Table 1. (continued)

Bacteria	Year of first publication	References	Associated diseases	Source of isolation: Human: Patients	Source of isolation: Human: Asymptomatic individuals	Source of isolation: Non-human	Reservoir/ source of infection
M. monacense	2006	[194]	Post-traumatic wound infection	Biopsy, sputum, BAL	–	–	Possibly environmental
M. nebraskense	2004	[195,196]	Respiratory infection	Sputum, BAL	–	–	–
M. palustre	2002	[184,197]	Lymphadenitis	Submandibular lymph node biopsy	–	Submandibular lymph node of pig, stream water, lymph node of cattle with evidence of bovine tuberculosis	–
M. saskatchewanense	2004	[198]	Pneumonia in bronchiectasis patients	Sputum, pleural fluid, respiratory sample	–	–	–
Nocardia africana	2001	[199,200]	Pulmonary nocardiosis	Sputum	–	Subcutaneous nodule of cat	–
N. aobensis	2004	[201]	–	–	–	–	–
N. arthritidis	2004	[202,203]	Thigh abscess, pulmonary nocardiosis, keratitis	Sputum, thigh inflammatory swelling discharge, cornea	–	–	–
N. asiatica	2004	[203–205]	Pulmonary nocardiosis, skin infection, keratitis, ocular nocardiosis	Sputum, granuloma, transtracheal aspirate, cutaneous ulcer, cornea, vitreous body	–	–	Possibly environmental
N. ignorata	2001	[206–208]	Pulmonary nocardiosis	Sputum, BAL, blood, tracheal aspirate	–	Soil	Probably soil
N. veterana	2001	[209–215]	Pulmonary nocardiosis, mycetoma, bacteraemia, peritonitis	BAL, subcutaneous nodule biopsy, sputum, open lung biopsy tissue, fine-needle transthoracic biopsy and aspirate, blood, peritoneal fluid	–	–	–
Olsenella profusa[a]	2001	[216–218]	Periodontitis, acute periradicular abscess	Subgingival plaque, carious dental lesion, root canal, pus	–	–	–
Weissella cibaria (= *Weissella kimchii*)	2002	[219–225]	–	Gall	Stool	Ear of dog with otitis, cheese whey, sugar cane, canary liver, chili bo, tapai, wheat sourdough, blood sausage, faecal swab of healthy dog, plaa-som (fermented fish product from Thailand), fermented kimchii, whole crop paddy rice silage	–
Aerobic and facultative anaerobic Gram-negative cocci							
Achromobacter insolitus	2003	[226]	–	Urine, wound	–	Laboratory sink	–
Achromobacter spanius	2003	[226]	–	Blood	–	–	–
Kerstersia gyiorum[a]	2003	[227]	–	Stool, leg and ankle wounds, sputum	–	–	–
Aerobic and facultative anaerobic Gram-negative rods							
Acinetobacter parvus	2003	[228]	Catheter-related bacteraemia	Ear, eye, forehead, skin, blood	–	Ear of dog with refractory otitis media	–
Acinetobacter schindleri	2001	[229,230]	Cystitis, bacteraemia	Urine, vagina, throat, ear, nasal swab, conjunctiva, skin, pleural effusion, liquor, blood, central catheter	–	–	–
Acinetobacter ursingii	2001	[229–231]	Infective endocarditis, bacteraemia, UTI, postneurosurgical meningitis, cholangitis, cervical adenopathy	Blood, intravenous line, pus, ulcer, eye, wound, urine, hairy skin, toe, abscess, central catheter, ganglion biopsy, CSF	–	–	–
Advenella incenata[a]	2005	[232]	–	Sputum, blood	–	Horse blood	–

Table 1. (continued)

Bacteria	Year of first publication	References	Associated diseases	Source of isolation: Human: Patients	Source of isolation: Human: Asymptomatic individuals	Source of isolation: Non-human	Reservoir/ source of infection
Averyella dalhousiensis[a]	2005	[233]	Catheter-related bacteraemia, cellulitis, wound infection	Wound, stool, blood, intestinal fluid	–	–	–
Burkholderia ambifaria	2001	[234,235]	Keratitis	Sputum and respiratory tract of CF patients, corneal culture	–	Biocontrol strains; corn roots; corn, pea, snap bean and sedge rhizosphere; forest soil, leaves of *Sesbania exaltata*, commercial soil	–
B. fungorum	2001	[236,237]	Septic arthritis	Vaginal secretion of pregnant woman, CSF, blood	–	White-rot fungus *Phanerochaete chrysosporium*, mouse nose, haemoglobin solution	–
Campylobacter hominis	2001	[88,238,239]	Bacteraemia	Blood	Stool	–	Gastrointestinal tract
Cardiobacterium valvarum	2004	[240–244]	Infective endocarditis, prosthetic valve endocarditis	Blood, noma lesion	–	–	Probably oral cavity
Cupriavidus respiraculi (= *Wautersia respiraculi* = *Ralstonia respiraculi*)	2003	[245–247]	–	Sputum and respiratory tract of CF patients	–	–	–
Enterobacter ludwigii	2005	[248]	Nosocomial UTI	Urine, trachea, fat tissue of left thigh, venous line, sputum, blood, stool, BAL, biopsy, throat, skin, swab	–	–	–
Escherichia albertii	2003	[249]	Diarrhoea	Stool	–	–	Probably gastrointestinal tract
Granulibacter bethesdensis[a]	2006	[250,251]	Lymphadenitis in CGD patients	Lymph node	–	–	–
Haematobacter massiliensis[a] (= *Rhodobacter massiliensis*)	2003	[252,253]	Bacteraemia, wound infection	Nose, blood, wound	–	–	–
Halomonas phocaeensis	2007	[254]	Fresh frozen plasma transfusion-associated bacteraemia/pseudobacteraemia	Blood	–	–	Probably environmental, e.g. water
Helicobacter winghamensis	2001	[255,256]	–	Stool of patients with gastroenteritis	–	Stool of gerbils	Probably gastrointestinal tract
Inquilinus limosus[a]	2002	[257–261]	Acute respiratory exacerbation and progressive loss of pulmonary function in CF patients?	Respiratory secretion and sputum of CF patients	–	Roots of grass	Possibly environmental
Laribacter hongkongensis[a]	2001	[76–84, 262–265]	Bacteraemia, empyema thoracis, community-acquired gastroenteritis, travellers' diarrhoea	Blood, empyema pus, stool	–	Freshwater fish, drinking water reservoir	Freshwater fish
Neisseria bacilliformis	2006	[266]	Wound infection, bronchitis, lung abscess	Ear, blood, oral fistula, submandibular wound, sputum, lung abscess	–	–	Possibly oral cavity

gastroenteritis would have been hampered. Therefore, in the context of pathogenic microbes, we think that even one strain of a novel species should be described, so that global, concerted efforts can be made to identify more cases associated with that pathogen.

Table 1. (continued)

Bacteria	Year of first publication	References	Associated diseases	Source of isolation: Human: Patients	Source of isolation: Human: Asymptomatic individuals	Source of isolation: Non-human	Reservoir/ source of infection
Paracoccus yeei	2003	[267]	CAPD peritonitis, wound infection, biliary tract infection, bacteraemia	Abdominal dialysate, wound, bile, CSF, blood	–	–	–
Pseudomonas mosselii	2002	[268,269]	–	Tracheal aspirate, bronchial aspirate, blood, venous catheter, stool, drainage liquid	–	Mineral water	–
P. otitidis	2006	[270]	Acute otitis externa, acute otitis media, chronic suppurative otitis media	Ear of patients with acute otitis externa, acute otitis media, chronic suppurative otitis media	–	–	–
Ralstonia taiwanensis	2001	[271]	–	Sputum of CF patients	–	Root nodule of *Mimosa* species	–
Wautersiella falsenii[a]	2006	[272]	–	Blood, wound, pus, respiratory tract, ear discharge, vaginal swab, pleural fluid, oral cavity	–	–	–
Anaerobic Gram-positive cocci							
Anaerococcus murdochii	2007	[273]	Infected foot ulcer, infected sternal wound, soft tissue neck infection, post-traumatic thumb abscess	Wound, abscess	–	–	–
Peptoniphilus gorbachii	2007	[273]	Infected dry gangrene, cellulitis, infected diabetic foot ulcer	Wound	–	–	–
Peptoniphilus olsenii	2007	[273]	Osteomyelitis, infected dry gangrene, infected diabetic foot ulcer, toe infection	Bone, wound	–	–	–
Peptostreptococcus stomatis	2006	[274,275]	Dento-alveolar abscess, endodontic infection, pericoronal infection, bacteraemia	Specimen from infected structure of oral cavity, oropharyngeal specimen, appendix, stool, blood	–	–	Possibly oral cavity and gastrointestinal tract
Anaerobic Gram-positive rods							
Alloscardovia omnicolens[a]	2007	[276]	–	Urine, blood, urethra, oral cavity, tonsil, abscesses of lung and aortic valve	–	–	–
Anaerostipes caccae[a]	2002	[277,278]	–	–	Stool	–	Probably gastrointestinal tract
Catabacter hongkongensis[a]	2007	[59]	Bacteraemia	Blood	–	–	Probably gastrointestinal tract
Clostridium bolteae	2003	[279–281]	Bacteraemia, intra-abdominal abscess, necrotizing fasciitis, pelvic abscess, wound infection	Blood, intra-abdominal abscess, abscess, wound	Stool	–	Probably gastrointestinal tract
C. hiranonis	2001	[282,283]	–	–	Stool	–	Probably gastrointestinal tract
Eggerthella hongkongensis	2004	[55]	Perianal abscess, infected rectal tumour, liver abscess, appendicitis	Blood	–	–	Probably gastrointestinal tract
Lactobacillus ultunensis	2005	[284–286]	–	Carious dentine	Gastric mucosa biopsy	Stool of pigs	–
Roseburia intestinalis	2002	[287]	–	–	Stool	–	Probably gastrointestinal tract
Shuttleworthia satelles[a]	2002	[288]	Periodontitis	Periodontal pocket, subgingival plaque	–	–	Probably oral cavity
Varibaculum cambriensis[a]	2003	[289]	Facial abscess, otogenic cerebral abscess	Abscess, IUCD	–	–	Possibly oral cavity
Anaerobic Gram-negative cocci							
Veillonella montpellierensis	2004	[290,291]	Infective endocarditis	Gastric fluid, amniotic fluid, blood	–	–	–
Anaerobic Gram-negative rods							
Alistipes onderdonkii	2006	[292]	Intra-abdominal infection	Appendix, abdominal abscess, stool, urine	Stool	–	Probably gastrointestinal tract
Alistipes shahii	2006	[292]	Intra-abdominal infection	Intra-abdominal fluid and appendix	Stool	–	Probably gastrointestinal tract
Bacteroides intestinalis	2006	[293]	–	–	Stool	–	Probably gastrointestinal tract

Table 1. (continued)

Bacteria	Year of first publication	References	Associated diseases	Source of isolation: Human: Patients	Source of isolation: Human: Asymptomatic individuals	Source of isolation: Non-human	Reservoir/ source of infection
Bacteroides plebeius	2005	[294]	–	–	Stool	–	Probably gastrointestinal tract
Cetobacterium somerae	2003	[295]	–	–	Stool	–	Probably gastrointestinal tract
Clostridium hathewayi[b]	2001	[85–88]	Bacteraemic acute cholecystitis and liver abscess, bacteraemic acute gangrenous appendicitis, bacteraemia	Blood, liver abscess	Stool	–	Gastrointestina tract
Desulfomicrobium orale	2001	[296]	Periodontitis	Subgingival plaque	–	–	Probably oral cavity
Dialister invisus	2003	[128, 297–299]	Endodontic and periodontal infections, UTI, acute periradicular abscess, acute periradicular periodontitis, acute apical abscess	Dental root canal, deep periodontal pocket, urine of renal transplant recipients, abscess pus, abscess aspirate	–	–	–
Dialister micraerophilus	2005	[300,301]	Anal, buttock, breast, perinephric abscess, bacteraemia, wound infection	Amniotic fluid, blood, pilonidal cyst, abscess pus, bone, bartholin gland, wound, phlegmon, skin and soft tissues, vagina	–	–	–
Dialister propionicifaciens	2005	[300]	Groin abscess, wound infection	Pressure ulcer, semen, abscess pus, wound	–	–	–
Dysgonomonas mossii	2002	[302,303]	–	Biliary drainage, abdominal drain, intestinal juice	–	–	–
Leptotrichia amnionii	2002	[304–307]	Septic arthritis, chorioamnionitis, renal abscess	Amniotic fluid, joint fluid, blood, renal abscess pus	–	–	Possibly female genital tract
Parabacteroides goldsteinii (= *Bacteroides goldsteinii*)	2005	[308,309]	Peritonitis, appendicitis, intra-abdominal abscess	Peritoneal fluid, appendix tissue, intra-abdominal abscess	–	–	Probably gastrointestinal tract
Porphyromonas somerae	2005	[310]	Chronic skin and soft tissue infections, osteomyelitis	Skin and soft tissues, bone	–	–	–
Porphyromonas uenonis	2004	[311]	Appendicitis, peritonitis, pilonidal abscess, infected sacral decubitus ulcer	Stool	–	–	Probably gastrointestinal tract
Prevotella baroniae	2005	[312]	Endodontic and periodontal infections, dentoalveolar abscess	Oral cavity	Dental plaque	–	–
Prevotella bergensis	2006	[313]	Skin and soft tissue abscesses, wound infection	Wound, decubitus ulcer	–	–	–
Prevotella copri	2007	[314]	–	–	Stool	–	Probably gastrointestinal tract
Prevotella marshii	2005	[312]	Endodontic and periodontal infections	Oral cavity	Subgingival dental plaque	–	–
Prevotella multiformis	2005	[315]	Chronic periodontitis	Subgingival plaque	–	–	–
Prevotella multisaccharivorax	2005	[316]	Chronic periodontitis	Subgingival plaque	–	–	–
Sneathia sanguinegens[a]	2001	[317,318]	Postpartum bacteraemia	Blood, amniotic fluid	–	–	Probably female genital tract
Spirochetes							
Borrelia spielmanii (= *Borrelia spielmani*)	2004	[89]	Lyme disease	Erythema migrans lesion	–	Ticks	Garden dormice
Treponema parvum	2001	[319]	Periodontitis, acute necrotizing ulcerative gingivitis, acute and chronic apical periodontitis	Subgingival plaque, acute necrotizing ulcerative gingivitis lesion, root canal	–	–	–

[a]Also a novel genus.
[b]The isolates discovered by Steer *et al*. were Gram-positive, but those found in the subsequent studies were Gram-negative.
BAL, bronchoalveolar lavage; IUCD, intrauterine contraceptive device; CSF, cerebrospinal fluid; UTI, urinary tract infection; CGD, chronic granulomatous disease; CF, cystic fibrosis; CAPD, continuous ambulatory peritoneal dialysis; IPD, intermittent peritoneal dialysis.

Detection of uncultivable bacteria and diagnosis of culture-negative infections

Bacterial culture had been the most important technique for diagnosing bacterial infections since the first days of microbiology. Indeed, Koch's original postulates explicitly required that the pathogen be grown in a pure culture. This specific requirement has since been updated, and the clinical importance of 'uncultivable' bacteria, e.g. *Treponema pallidum*, is now well recognized. Nonetheless, given the central role of bacterial culture in most clinical laboratories, the diagnosis of culture-negative infections and those caused by uncultivable bacteria remains a somewhat arduous task. Direct microscopy and immunology-based assays have been the culture-independent methods traditionally used. The sensitivities and specificities of these methods can vary greatly, depending on the implementation and the organisms concerned. Direct microscopic examination demands a significant microbial presence in the specimen for positive identification, and serological tests suffer from cross-reactivities and may be affected by a patient's immunocompromised status. The introduction of molecular diagnostics significantly enhanced the ability to diagnose culture-negative infections. Among the various molecular assays available, 16S rDNA sequencing stands out as a useful technique for detecting uncultivable bacteria. The universal presence and sequence conservation of the gene allows broad-range PCR primers to be readily designed. This characteristic is of great importance, as it would be impractical to perform multiple specific assays to rule out individual bacteria associated with the disease.

An important use of 16S rDNA sequencing is the identification of clinical syndromes caused by uncultivable bacteria. These relate to infections caused by bacteria that could not be cultured reliably by known methods at the time of discovery. For example, Whipple's disease was known only as a systemic disorder of unknown aetiology when first described in 1907. Since then, gradual progress towards better diagnosis and treatment of the disease has been made, although the classic method of periodic acid–Schiff staining for diagnosis lacks both sensitivity and specificity when used alone. The first breakthrough in the understanding of Whipple's disease arrived with the identification of *Tropheryma whippeli* as the causative agent. This was accomplished by the PCR amplification and sequencing of its 16S rDNA gene [96,97]. The 16S rDNA sequence allowed the phylogenetic position of *T. whippeli* to be defined immediately, and enabled the later development of molecular diagnostic tests [97–100]. The definitive identification of the aetiological agent greatly accelerated research into the pathophysiology of the disease. In 2003, the genome sequences of two *T. whipplei* strains were completed [101,102], marking another milestone in medical history. Similar successes were seen in the context of bacillary angiomatosis (caused by *Bartonella henselae* and *Bartonella quintana*) [103,104] and human ehrlichiosis (caused by bacteria of the genera *Ehrlichia* and *Anaplasma*) [105–107].

16S rDNA sequencing has also been used for diagnosing culture-negative infections, the best example being culture-negative endocarditis. Up to one-third of all cases of infective endocarditis are culture-negative [108], and diagnosis has relied mainly upon clinical and ultrasonographic findings. Goldenberger *et al.* reported one of the earliest situations in which 16S rDNA PCR amplification and sequencing were performed on DNA extracted from infected valves [109]. Many subsequent studies confirmed the usefulness of the method [110–119]. In general, the sensitivity was found to be comparable to that of serological testing and histopathological examination. An interesting and important result of these studies is the recognition that many culture-negative cases were caused by cultivable Gram-positive bacteria, but the blood culture results might have been negative due to previous antibiotic usage. Other factors that may contribute to the false-negative blood culture results include inadequate microbiological techniques, e.g. use of an insufficient volume of blood, or infection by a fastidious organism [120].

Diagnosis of other culture-negative infections, including meningitis [121–125], brain abscess [126], keratitis [127], urinary tract infections [128], empyema [129,130], septic arthritis [131,132] and septicaemia [120,133,134], also benefits from the use of 16S rDNA sequencing. The encouraging results from these studies reinforced the idea that 16S rDNA sequencing is broadly applicable as a diagnostic technique in the context of clinical microbiology. Future developments should involve further optimization of universal primer sets and attempts to discriminate between

positive results and contamination or true pathogens. Recent progress in these areas has been made through the use of a broad-range real-time PCR design [124,135,136]. With continued improvements in performance and cost profile, 16S rDNA sequencing is expected to play an increasingly significant role in the diagnosis of culture-negative bacterial infections.

Automation of 16S rDNA sequencing

One of the greatest hurdles in putting 16S rDNA sequencing into routine use in clinical microbiology laboratories is automation of the technology. At present, complete automation of 16S rDNA sequencing, i.e. input of the bacterial isolate/clinical sample for identification or its extracted DNA into a machine that will generate the result of the identity of the isolate, is impossible. Most of the steps, including DNA extraction, PCR amplification, purification of PCR products, DNA sequencing and sequence editing, have to be performed manually, although user-friendly commercial kits for some of these steps are available. With the advances in high-throughput technologies, these steps may be incorporated into a robotic system, making complete automation of 16S rDNA sequencing possible in the future.

At present, the only steps that can be automated are the input of the finalized 16S rDNA sequence concerned into one of the software packages that contains a database with 16S rDNA sequences of selected bacterial species and the output of the result of the identity of the isolate. The most well known of these software packages include Ribosomal Database Project (RDP-II) [137–141], MicroSeq [9,48,142,143], Ribosomal Differentiation of Medical Microorganisms (RIDOM) [144–146] and the SmartGene Intergrated Database Network System (SmartGene IDNS) [147] (Table 2). The databases of RDP-II and SmartGene IDNS contain sequences downloaded from GenBank, whereas all sequences in the databases of RIDOM and MicroSeq were obtained by sequencing the 16S rDNA genes of bacterial strains of culture collections.

The major studies that have evaluated the usefulness of the various software packages for different groups of bacteria are summarized in Table 3. However, it is not possible to compare accuracy among the different studies, because

Table 2. Commonly used software packages for bacterial identification using 16S rDNA sequencing

Software	Year of first description	Company/Organization	Website	Partial/full 16S rDNA sequence included	Database size	Source of sequences	Cost	Quality control	Updates
RDP-II	1991	Michigan State University, USA	http://rdp.cme.msu.edu/	Partial and full	>450 000 (release 9.56)	GenBank	No	Partial	Monthly
MicroSeq microbial identification system[a]	1998	Applied Biosystems, Foster City, CA, USA	http://www.microseq.com	Partial and full[a]	1716 and 1261 in the two databases (version 2) respectively[a]	Sequence of 16S rDNA gene of one strain from each species	Yes	All type strains from culture collections	Periodically
RIDOM	1999	Ridom GmbH, Würzburg, Germany	http://rdna2.ridom.de/	Partial	240[b]	Sequence of 16S rDNA gene of bacteria in Neisseriaceae, Moraxellaceae and *Mycobacterium*[c]	No	All strains from culture collections	Periodically
SmartGene IDNS	2006	SmartGene GmbH, Switzerland	http://www.smartgene.ch/index.shtml	Partial and full	~112 000[d]	GenBank	Yes	Partial	Weekly

[a]It contains two databases, MicroSeq ID 16S rDNA 500 Library v2.0, which contains sequences from the 5′- end 527 bp of 16S rDNA genes, and MicroSeq ID 16S rDNA Full Gene Library v2.0, which contains full 16S rDNA sequences.
[b]Number counted from http://rdna2.ridom.de/ridom2/servlet/link?page=list (18 January 2008).
[c]*Staphylococcus*, *Nocardia* and *Bacillus* species were also studied [149,151,320] but were not found in the software database.
[d]Number obtained from reference [147].

Table 3. Studies on the usefulness of commonly used software pachages for bacterial identification using 16S rDNA sequencing

References	Bacterial isolates tested	Methods for determination of real identity of the isolates	Accuracy of software
Tang *et al*. [10]	65 unusual aerobic Gram-negative bacilli from clinical specimens	Conventional phenotypic methods	MicroSeq 500: 89.2% (72% of 25 fermenters and 100% of 40 non-fermenters; 97.2% to genus level)
Tang *et al*. [143]	52 coryneform Gram-positive bacilli (42 *Corynebacterium* species and 10 *Corynebacterium*-related species) from clinical specimens	Conventional phenotypic methods[a]	MicroSeq 500: 66.7%[a] (100% to genus level)
Patel *et al*. [142]	113 *Mycobacterium* clinical isolates (18 different species)	Combination of phenotypic methods, 16S rDNA sequencing using the MicroSeq database, and other methods, e.g. standard biochemical assays, *hsp65* gene sequencing, AccuProbe rRNA hybridization, HPLC of mycolic acids. Identifications were counted as correct if two methods provided the same answer	MicroSeq 500 (version 1.36): 92.8%[b]
Turenne *et al*. [321]	79 non-tuberculous mycobacterial isolates (ATCC type strains)	Not described, but all were ATCC strains	BLAST (GenBank and EMBL databases, dated 30 March 2001): 38%[c] RDP-II (dated 30 March 2001): 41%[c] RIDOM (dated 8 March 2001): 100%[c]
Cloud *et al*. [148]	119 *Mycobacterium* isolates (94 clinical isolates and 25 ATCC strains)	Combination of conventional phenotypic methods, 16S rDNA sequencing using MicroSeq and RIDOM databases, and additional tests on problematic isolates	MicroSeq 500 (version 1.36) in conjunction with RIDOM database: 96.6%[d] (100% of 25 ATCC strains and 95.7% of 94 clinical isolates)
Mellmann *et al*. [149]	81 *Nocardia* clinical isolates	Combination of conventional phenotypic methods, 16S rDNA sequencing using MicroSeq and RIDOM databases	RIDOM (version 1.1): 100%[e] MicroSeq 500 (version 1.4.3, library version 500-0125): 44.4%[e]
Woo *et al*. [9]	37 (15 aerobic or facultative anaerobic Gram-positive; 11 aerobic, micro-aerophilic or facultative anaerobic Gram-negative, seven anaerobic; three mycobacterial; and one mycoplasma) clinical isolates with ambiguous biochemical profiles	Combination of phenotypic methods and full 16S rDNA sequencing	MicroSeq 500 (version 1.0): 81.1% (86.5% to genus level)
Hall *et al*. [43]	387 *Mycobacterium* isolates (59 ATCC strains and 328 clinical isolates)	Combination of conventional phenotypic methods, HPLC, PCR restriction analysis of 65-kDa heat shock protein region, 16S rDNA sequencing using MicroSeq 500 database	MicroSeq 500 (version 1.4.2): 68% (98.3% of 59 ATCC strains and 62.5% of 328 clinical isolates)
Cloud *et al*. [150]	94 *Nocardia* clinical isolates (ten separate species)	Combination of conventional phenotypic methods, 16S rDNA sequencing using expanded MicroSeq 500 database, restriction endonuclease assay for portions of 16S rDNA and *hsp65* genes, sequencing of a 999-bp fragment of the 16S rDNA gene	Expanded MicroSeq 500 (version 1.4.3, library version 500-0125)[f]: 82%
Becker *et al*. [151]	55 clinical *Staphylococcus* isolates	Combination of phenotypic methods, 16S rDNA sequencing using RIDOM and NCBI databases and additional tests (chemotaxonomy and riboprinting) on isolates with ambiguous or below the-threshold RIDOM results	RIDOM (version 1.2): 90.9%[g] GenBank (dated 12 January 2004): 65.5%[g]
Lau *et al*. [48]	20 anaerobic Gram-positive bacilli isolated from blood cultures	Preliminary phenotypic methods and full 16S rDNA sequencing	MicroSeq 500 (version 1.0): 65% (80% to genus level)
Patel *et al*. [322]	99 aerobic actinomycetes (28 reference strains and 71 clinical isolates, including members of the genera *Streptomyces*, *Gordonia*, and *Tsukamurella*, and ten taxa of *Nocardia*)	Phenotyping, PRA, drug susceptibility testing	Microseq 500: 70.3%[h] BLAST (GenBank): 70.3%[i]
Fontana *et al*. [152]	83 (25 Gram-positive and 58 Gram-negative) clinical isolates not identifiable by conventional systems	Not mentioned	Not determined
Simmon *et al*. [147]	300 (158 Gram-positive and 142 Gram-negative) clinical isolates	Combination of conventional phenotypic methods and 16S rDNA sequencing using SmartGene IDNS and MicroSeq 500 databases	SmartGene IDNS (version 3.2.3r8): 87% (98% to genus level)[j] MicroSeq 500 (version 1.4.3): 74% (90% to genus level)[j]

hsp65, heat shock protein 65; ATCC, American Type Culture Collection; BLAST, Basic Local Alignment Search Tool; EMBL, European Molecular Biology Laboratory; JCM, Japan Collection of Microorganisms; PRA, PCR-restriction endonuclease analysis.

[a]Four of the 52 isolates not identified by conventional phenotypic methods were not included.

[b]Two of the 113 isolates not identified by phenotypic methods were not included.

[c]Identification was considered correct if it was the best matched species, although not a perfect match. Over 1400 bp of the 16S gene was aligned in BLAST and RDP-II, whereas *Escherchia coli* bp 54–510 was used in RIDOM.

[d]Sequences resulting in distance score of <0.8%.

[e]Similarity ≥99.12% reported by either RIDOM or MicroSeq.

[f]An expanded portion of the database was developed from partial 5′ 16S rDNA sequences derived from 28 reference strains (from ATCC and JCM); samples with 99–100% similarity to a species in the expanded MicroSeq 500 library were assigned that species designation, provided that the colony morphology and growth characteristics were consistent.

[g]Partial 16S rDNA fragment (corresponding to *E. coli* positions 54–510); classification is based on percentage similarity ≥98.5 to the 16S rDNA sequence of the type and reference strain.

[h]Eight of the 99 isolates not identified by phenotypic methods were not included. Eighty per cent of 30 *Actinomadura*, *Gordona*, *Rhodococcus*, *Streptomyces* and *Tsukamurella* species, and 65.6% of 61 *Nocardia* species.

[i]Eight of the 99 isolates not identified by phenotypic methods were not included. Fifty per cent of 30 *Actinomadura*, *Gordona*, *Rhodococcus*, *Streptomyces* and *Tsukamurella* species, and 80.3% of 61 *Nocardia* species.

[j]Genus-level and species-level identifications were assigned using the following criteria: ≥99% identity to a reference entry identified a bacterium to the species level, 97.0–98.9% identity identified a bacterium to the genus level, and <97% identity to any sequence was considered to be unable to provide a definite identification.

bacterial isolates from different sources (reference strains vs. clinical strains, identifiable strains vs. unidentified strains), different reference standards and different criteria for 'correct' identification were used in the different studies. In some studies, the identities of the isolates were determined by a combination of phenotypic and genotypic methods [43,142,147–151]. In others, such as a recent one on a heterogeneous group of clinical isolates, no reference standard was mentioned, except whether the isolates were 'identified' by MicroSeq [152]. As there are intrinsic problems with MicroSeq, e.g. inclusion of bacterial species with 16S rDNA sequences with minimal differences from those of some other species, it would be difficult for readers to assess the usefulness of MicroSeq from this study. As the intrinsic problems of the software packages may not be fully addressed in the publications, some of the reported accuracies of the software packages described may be overestimated.

USEFULNESS AND LIMITATIONS OF USING 16S RDNA SEQUENCING IN CLINICAL MICROBIOLOGY LABORATORIES

Although 16S rDNA sequencing is being used increasingly for bacterial identification in clinical microbiology laboratories, there are no widely accepted guidelines for using the technique or for the interpretation of sequence data. Given the limitations of the technique for some taxa, the ever-expanding sequence databases and taxonomic complexity, and the inaccuracies in some databases, some recommendations have been suggested for the use of 16S rDNA sequencing for bacterial identification [7,153]. For indications for the use of 16S rDNA sequencing, the strains most often chosen are those that cannot be accurately identified with phenotypic tests in clinical microbiology laboratories. However, as certain groups of bacteria are known to present difficulties in identification by 16S rDNA sequencing, these bacteria should be excluded, and other housekeeping gene targets, e.g. *rpoB*, if available, are required [39–42,47,50–54,153]. Regarding the interpretation of sequence data, this involves requirements concerning the length and quality of sequences, the choice of appropriate programs for analysis, and the final species assignment based on similarity search results. Although a minimum of 500–525 bp that includes the more variable 5′-region may be adequate for identification of some groups of bacteria, Drancourt *et al.*, according to their several recommendations concerning the criteria for 16S rDNA sequencing as a reference method for bacterial identification, proposed that full 16S rDNA sequences with <1% ambiguities should be used [7]. The comparison should be made using at least 1500 positions in all sequences, of the same length, included in the similarity search, and an ungapped program should be used. In a recent review by Janda and Abbott, the use of full 16S rDNA sequences, whenever possible, and in particular for groups such as *Campylobacter* species [153], was confirmed as a viable technique.

The major difficulties and controversies in the interpretation of sequence data concern the assignment of bacterial species according to similarity search results, as no threshold values are available, as in the case of DNA–DNA hybridization [154]. Although a 97% similarity level has been proposed for bacterial speciation using 16S rDNA sequences [155], a >0.5% difference may be indicative of a new species [156]. In fact, as different bacterial species are likely to evolve at different rates, it is impossible to determine a universal cut-off for bacterial genus and species delineation. Although, in the study by Drancourt *et al.*, >99% and >97% sequence similarities were used as the cut-offs for species and genus identification, respectively, the authors indicated that these sharp values were set mainly for practical purposes, for interpretation of their large sequence dataset, and it may be necessary to use different cut-off values, depending on the bacterial genus under investigation [7]. On the other hand, Janda and Abbott, in their recommended guidelines, suggested that a minimum of >99%, and ideally >99.5%, sequence similarity be used as the criteria for species identification [153]. They also proposed that for matches with distance scores <0.5% to the next closest species, other properties, e.g. phenotype, should be considered in final species identification. However, using these criteria, it is difficult to determine the species identity with sequences that have very similar distances to the closest and next closest matches, especially those within 0.5–1%. Although different studies have identified groups of bacteria for which 16S rDNA sequences are not sufficiently discriminative and for which other

gene targets have to be used, very few studies have addressed and attempted to solve this problem in a systematic way.

The use of 16S rDNA sequencing for bacterial identification depends on significant interspecies differences and small intraspecies differences in 16S rDNA sequences. Therefore, one of the major limitations is that when two different bacterial species share almost the same 16S rDNA sequence, this technique would not be useful for distinguishing between them. In our experience, clinical microbiologists and technicians with limited experience in 16S rDNA sequencing often find interpretation of 16S rDNA similarity search results difficult. As a result of the large number of unvalidated 16S rDNA sequences in GenBank, it is not a straightforward task for inexperienced users to decide whether the 'first hit' or 'closest match' represents the actual identity of a bacterial isolate.

Regarding the software packages described above, the usefulness is further limited by the choice of bacterial species in the database. If a bacterial species is not included in the database, it would never be given as the identity of an isolate, and if the database also includes bacterial species with minimal differences in their 16S rDNA sequences and which, therefore, cannot be identified confidently by 16S rDNA sequencing, this may also give rise to incorrect identification.

In view of these limitations, we initiated a systematic evaluation of the potential usefulness of full and 527-bp 16S rDNA sequencing and the existing MicroSeq databases for identification of medically important bacteria 2 years ago. The species of medically important bacteria included in this analysis comprise all species of medically important bacteria listed in the *Manual of Clinical Microbiology* [157]. The most representative 16S rDNA sequence for each species was chosen from the GenBank database. The percentage differences of the 16S rDNA sequences among the different species in the same group/genus were determined by pairwise alignment.

The study first involved medically important anaerobic bacteria, and the results were published in 2007 [158]. Each medically important bacterial species is classified as one of the following: (i) the bacterium can be confidently identified by 16S rDNA sequencing (√ in Table 4 and Supplementary Tables 2–4 of reference [158]), meaning that there is a >3% difference between the 16S rDNA sequence of the species and those of other medically important bacteria; (ii) the bacterium cannot be confidently identified by 16S rDNA sequencing (× in Table 4 and Supplementary Tables 2–4 of reference [158]), meaning that there is a <2% difference between the 16S rDNA sequence of the species and that of a closely related medically important bacterium; and (iii) the bacterium can only be doubtfully identified by 16S rDNA sequencing (? in Table 4 and Supplementary Tables 2–4 of reference [158]), meaning that there is a 2–3% difference between the 16S

Table 4. An example of guidelines for interpretation of usefulness of full 16S rDNA sequencing and MicroSeq full 16S rDNA bacterial identification system database for identification of medically important *Actinomyces* species (adapted from Supplementary Table 2 of reference [158] with permission)

Actinomyces species	Usefulness for species identification	Medically important bacteria with similar (<3% difference) full 16S rDNA gene sequences	Usefulness of existing MicroSeq full 16S rDNA bacterial identification system database
A. bovis	?	*A. urogenitalis*	?[c]
A. bowdeni	?	*A. viscosus*	×[a]
A. canis	√		×[a]
A. catuli	√		×[a]
A. denticolens	√		×[a]
A. europaeu	√		×[a]
A. funkei	?	*A. hyovaginalis*	×[a]
A. georgiae	?	*A. meyeri*	×[a]
A. gerencseriae	√		×[a]
A. graevenitzii			×[a]
A. hordeovulneris	√		×[a]
A. howellii	√		√
A. hyovaginalis	?	*A. funkei*	×[a]
A. israelii	√		×[a]
A. meyeri	?	*A. georgiae*, *A. odontolyticus*, *A. turicensis*	?[c]
A. naeslundii	×	*A. viscosus*	×[a]
A. neuii	√		√
A. odontolyticus	×	*A. meyeri*, *A. turicensis*	×[a]
A. radicidentis	√		×[a]
A. radingae	√		√
A. slackii	√		√
A. turicensis	×	*A. meyeri*, *A. odontolyticus*	×[a]
A. urogenitalis	?	*A. bovis*	×[a]
A. viscosus	×	*A. bowdenii*, *A. naeslundii*	×[b]

√, >3% difference between the 16S rDNA gene sequence of the species and those of other medically important bacteria; ?, 2–3% difference between the 16S rDNA gene sequence of the species and that of a closely related medically important bacterium; ×, <2% difference between the 16S rDNA gene sequence of the species and that of a closely related medically important bacterium.

[a]The species is not included in the existing MicroSeq full 16S rDNA bacterial identification system database.

[b]Although the species is included in the existing MicroSeq full 16S rDNA bacterial identification system database, the high 16S rDNA gene sequence similarity (>98% nucleotide identity) between the species and a closely related species does not allow them to be distinguished confidently.

[c]Although the species is included in the existing MicroSeq full 16S rDNA bacterial identification system database, a 2–3% difference is observed between the 16S rDNA gene of the species and a closely related species.

rDNA sequence of the species and that of a closely related medically important bacterium. According to our guidelines for 16S rDNA sequence analysis, if the bacterium belongs to the '?' or '×' category (e.g. *Actinomyces meyeri*; Table 4 and Supplementary Tables 2–4 of reference [158]), the bacterial species with similar 16S rDNA sequences will also be known (i.e. *A. georgiae*, *A. odontolyticus* and *A. turicensis*; Table 4 and Supplementary Tables 2–4 of reference [158]). If further species differentiation is necessary, one can look for additional/supplementary methods, which may be key phenotypic tests or sequencing of additional gene loci, to distinguish among these species with similar 16S rDNA sequences. It should be noted that the cut-offs mentioned are not meant to be used as criteria for defining new species, but to provide a clearer meaning for the search results.

Concerning MicroSeq database analysis, a supplementary note is given to indicate whether the reason for the inability of the database to identify the bacterium is due to the species being not included in the existing database (x^a in Table 4 and Supplementary Tables 2–4 of reference [158]), or high 16S rDNA sequence similarity (>98% nucleotide identity) to a closely related species (x^b in Table 4 and Supplementary Tables 2–4 of reference [158]). In the study on anaerobes, full and 527-bp 16S rDNA sequencing were capable of identifying 52–63% of 130 anaerobic Gram-positive rods, 72–73% of 86 anaerobic Gram-negative rods, and 78% of 23 anaerobic cocci. Surprisingly, the MicroSeq databases (version 1.0) were 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. This represents only 45–46% of those that should be confidently identified by full-length and 527-bp 16S rDNA sequencing, indicating that, in order to improve the usefulness of MicroSeq, bacterial species that should be confidently identified by full-length and 527-bp 16S rDNA sequencing but are not included in the MicroSeq databases (version 1.0) should be included. Recently, there has been an expansion of the MicroSeq databases (version 2.0) to include more bacterial species, which has led to some improvement. At present, we are performing further studies of aerobic and facultative anaerobic Gram-positive and Gram-negative bacteria, and the results will be available in the near future for the development of similar guidelines.

Despite the extensive analysis, there are still limitations in these guidelines. First, intraspecies variation in 16S rDNA sequences was not taken into account in the analysis. In the evaluation, only whether there was a significant difference between the 16S rDNA sequence of a particular species and those of other species was determined, and a 3% difference was chosen as the cut-off. Whether there is a high similarity among the 16S rDNA sequences in different strains of the same species is difficult to study, because: (i) 16S rDNA sequence information is available for multiple strains of the same species in only a minority of species; and (ii) even when 16S rDNA sequence information is available for multiple strains of the same species, the strains are often not well characterized phenotypically, so the reliability of the sequence information is difficult to judge. However, despite this limitation, a high degree of conservation of 16S rDNA sequences for the same species is often assumed, because most existing results do not indicate otherwise. On the other hand, for some genera, the intraspecies variation in 16S rDNA sequences may be so small that a difference of 3% in the 16S rDNA sequences between two species may not be necessary for confident identification. Second, the present study included only the bacterial species that are known to be associated with infections. This was deliberate, because, if those bacterial species that have never been reported to cause infections had also been included, some species may have been classified as 'not confidently identified by 16S rDNA sequencing' because of the sequence similarity between a bacterium and another that has never been reported to cause infection. One must bear in mind that those species that have never been reported to be associated with infections may still have the potential to do so. Owing to these limitations, we suggest that, in order to increase the accuracy of 16S rDNA sequencing for identification of pathogenic bacteria, it would be necessary to interpret the results of 16S rDNA sequencing with preliminary phenotypic test results. Nevertheless, our data not only facilitate interpretation of sequence data by inexperienced users in the clinical microbiology laboratories, but also provide clues to the potential usefulness of 16S rDNA