Detection and identification of bacteria in clinical samples by 16S rRNA gene sequencing: comparison of two different approaches in clinical practice

Claire Jenkins,1,2 Clare L. Ling,1,2 Holly L. Ciesielczuk,1,3 Julianne Lockwood,1 Susan Hopkins,1,3 Timothy D. McHugh,3 Stephen H. Gillespie3,4 and Christopher C. Kibbler1

Amplification and sequence analysis of the 16S rRNA gene can be applied to detect and identify bacteria in clinical samples. We examined 75 clinical samples (17 culture-positive, 58 culture-negative) prospectively by two different PCR protocols, amplifying either a single fragment (1343 bp) or two fragments (762/598 bp) of the 16S rRNA gene. The 1343 bp PCR and 762/598 bp PCRs detected and identified the bacterial 16S rRNA gene in 23 (31%) and 38 (51%) of the 75 samples, respectively. The 1343 bp PCR identified 19 of 23 (83%) PCR-positive samples to species level while the 762/598 bp PCR identified 14 of 38 (37%) bacterial 16S rRNA gene fragments to species level and 24 to the genus level only. Amplification of shorter fragments of the bacterial 16S rRNA gene (762 and 598 bp) resulted in a more sensitive assay; however, analysis of a large fragment (1343 bp) improved species discrimination. Although not statistically significant, the 762/598 bp PCR detected the bacterial 16S rRNA gene in more samples than the 1343 bp PCR, making it more likely to be a more suitable method for the primary detection of the bacterial 16S rRNA gene in the clinical setting. The 1343 bp PCR may be used in combination with the 762/598 bp PCR when identification of the bacterial rRNA gene to species level is required.

INTRODUCTION

Use of broad-range 16S rRNA gene PCR as a tool for identification of bacteria is possible because the 16S rRNA gene is present in all bacteria (Woese, 1987). The 16S rRNA gene consists of highly conserved nucleotide sequences, interspersed with variable regions that are genus- or species-specific. PCR primers targeting the conserved regions of rRNA amplify variable sequences of the rRNA gene (Relman, 1999). Bacteria can be identified by nucleotide sequence analysis of the PCR product followed by comparison of this sequence with known sequences stored in a database (Clarridge, 2004).

The method is appropriate for use where the range of pathogens likely to present is wide and where organism-specific PCRs are inappropriate. It can also be used to detect bacteria that are difficult to grow or be applied to samples from post-antibiotic treatment (Relman & Falkow, 1992; Brouqui & Raoult, 2001; Harris et al., 2002). Others have used this method on cerebrospinal fluid, pus, joint fluids and tissue (Harris & Hartley, 2003).

Published protocols for the assay vary; it is thought that analysis of large fragments (>1000 bp) may improve species discrimination but that amplification of shorter fragments (<1000 bp) may be more sensitive and produce better quality sequence data. In this study, we compared a single fragment PCR (1343 bp PCR) using primers published previously (Woo et al., 2001), which amplifies approximately 1343 bp of the 16S rRNA gene, with a two-fragment PCR assay (762/598 bp PCR) (Xu et al., 2004), which uses two primer sets to amplify two smaller overlapping fragments of the 16S rRNA gene to produce a composite sequence of 1068 bp (Table 1; Fig. 1).
**METHODS**

**Conventional microbiological methods.** Pleural fluid and pus samples were processed by a range of routine microbiological methods as laid out in the Departmental Standard Operating Procedures (Department of Medical Microbiology, Royal Free NHS Trust, London, UK). In brief, samples were cultured and isolated were identified by routine biochemical tests and an API kit (bioMérieux) or using the BD Phoenix Automated Microbiology System (USA).

**DNA extraction.** DNA was extracted from fluid samples (200 μl) using the QiAmp DNA Minikit (Qiagen) according to the manufacturer’s instructions, with a final elution volume of 100 μl. Extracted DNA was stored at 4°C until required for PCR.

**16S rRNA gene PCR.** The 10 μl bacterial DNA extract and controls were amplified with 0.5 μM primers (Table 1), 200 μM of each dNTP (Promega), 10 mM KCl PCR buffer, 2 mM MgCl2, and 1.0 U Taq polymerase (Bioline). Amplification conditions for both PCRs were as follows: 5 min at 94°C to denature the DNA, followed by 40 cycles of denaturation at 94°C for 1 min, primer annealing at 55°C for 1 min and strand extension at 72°C for 2 min on a Rotorgene thermal cycler. PCR products were separated on a 1.5% agarose gel and DNA bands were visualized with ethidium bromide. Primers and excess nucleotides were removed from the amplified DNA using a PCR clean-up kit (Qiagen). The amount of DNA in the cleaned-up product was quantified by comparing the intensity of the band to bands of known intensity in a HyperLadder marker (Bioline). Spiked samples were included with each PCR to determine whether or not the sample was inhibited. Spiked samples contained 8 μl clinical sample and 2 μl known positive 16S rRNA.

**Sequence analysis.** Purified PCR products were sequenced in the forward and reverse direction in separate reactions and in duplicate. Each reaction contained 40 μl template DNA, 2 μl of the appropriate PCR primer (Table 1), 10 μl water and 2 μl BigDye Terminator v3.1 Ready Reaction Mix (Applied Biosystems). Each reaction was heated to 96°C for 1 min, followed by 25 cycles at 96°C for 10 s, 50°C for 5 s and 60°C for 4 s. The sequencing products were purified using an ethanol precipitation method to remove unincorporated reagents and ensure a neutral charge. Briefly, sequencing products were washed in 80 μl ethanol precipitation mix (3 μl NaAc, 62.5 μl 95% ethanol and 14.5 μl water) and the DNA was pelleted by centrifugation (13 000 g, 15 min). The pellet was again washed in 200 μl 75% ethanol and centrifuged (13 000 g, 5 min). The pelleted DNA was air-dried and rehydrated in 15 μl formamide and then loaded onto a 3130 Genetic Analyzer Capillary Array for detection (Applied Biosystems). Two forward and two reverse sequences for each sample were aligned using Bionumerics v3.5 (Applied Maths) to obtain a composite sequence. The quality of each sequence trace was visually assessed and poor quality sequence was edited and removed. Organisms were identified for each assay by comparing consensus sequences to a database library of known 16S rRNA gene sequences in GenBank (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) by multiple sequence alignment. The bacterial source of the sequence was identified by matching it with a sequence with the highest maximum identity score from the GenBank database. Where more than one bacterial species had the same highest score, all species were recorded in the results. Sequences with <97% similarity to hits from the GenBank database were considered to be of poor quality and were excluded from this study (Welinder-Olsson et al., 2007). With respect to the 762/598 bp PCR, the individual 762 bp and 598 bp sequences as well as the 1068 bp composite sequence were all analysed using the GenBank database to ensure that the two fragments were derived from the same bacterial species.

**RESULTS**

Seventy-five clinical samples were analysed between October and November 2007, including the prospective analysis of 13 pus samples and 10 joint fluids, and the retrospective analysis of 51 pleural fluids. A total of 17 of the 75 samples (23%) were culture-positive. Bacterial 16S rDNA was amplified in 23 out of 75 (31%) of the samples analysed using the 1343 bp PCR protocol and in 38 out of 75 (51%) using 762/598 bp PCR (Tables 2 and 3). Thirty-three samples were negative by both culture and PCR.

![Fig. 1. Relative position of the amplified fragments to the whole 16S rRNA gene (A). Fragment amplified by the 1343 bp PCR (B), and the fragments amplified by 762/598 bp PCR-1 (C) and 762/598 bp PCR-2 (D) are indicated. The 762/598 bp PCR primer sets amplify two overlapping fragments of the 16S rRNA gene to produce a composite sequence of 1068 bp.](image-url)
methods. Insufficient clinical details were available to determine whether the patients’ symptoms were likely to be due to a fungal or viral infection, associated with a non-microbiological cause, or resulting from a bacterial infection that could not be detected by culture or PCR.

Using the 1343 bp PCR, bacterial 16S rDNA was identified in eight of the 17 culture-positive samples (Table 2). Samples containing *Staphylococcus aureus* (sample 11), *Streptococcus pneumoniae* (sample 9), viridans streptococci (sample 12), *Prevotella oulorum* (sample 6) and *Prevotella pleuritidis* (sample 8) gave concordant results, although there was variation between the phenotypic and genotypic bacterial identification in the remaining three culture-positive/1343 bp PCR-positive samples (samples 3, 7 and 13) (Table 2). Anaerobes were identified in culture-positive samples by 16S rRNA gene sequencing (samples 3, 6, 8 and 13) but in only two samples using conventional culture (samples 6 and 8). Sample 3 grew *Escherichia coli* and *Enterococcus* species but *Prevotella denticola* was the only bacterium identified by sequencing.

The 762/598 bp PCR and sequencing identified a significant pathogen in 13 of the 17 culture-positive samples and the phenotypic and genotypic identifications were concordant in 10 (Table 2). Two of the mismatched identifications made using the 762/598 bp PCR protocol showed similar results as those using the 1343 bp PCR protocol (Table 2).

Of the four culture-positive/762/598 bp-PCR negative samples, three contained Gram-negative bacteria [E. coli (2), *Klebsiella oxytoca*] and one grew *Staphylococcus aureus*. The 1343 bp PCR failed to detect bacterial 16S rRNA in samples that were culture-positive for *E. coli* (5), *K. oxytoca*, viridans type streptococci (2) and *Staphylococcus aureus*.

### Table 2. Comparison of culture-positive samples with the detection and identification of the bacterial 16S rRNA gene using 1343 bp PCR and 762/598 bp PCR assays

With respect to the PCR results, the bacterial source of the sequence was identified by matching it with a sequence with the highest maximum identity score from the GenBank database.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sample type</th>
<th>Conventional bacteriology*</th>
<th>1343 bp PCR assay</th>
<th>762/598 bp PCR assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pleural fluid</td>
<td><em>Escherichia coli</em></td>
<td>Bacterial 16S rRNA gene not detected</td>
<td><em>Escherichia coli</em>, <em>Shigella</em> sp., <em>Streptococcus anginosus</em>, <em>Streptococcus intermedius</em></td>
</tr>
<tr>
<td>2</td>
<td>Pus</td>
<td><em>Streptococcus</em> sp.</td>
<td>Bacterial 16S rRNA gene not detected</td>
<td><em>Streptococcus anginosus</em>, <em>Streptococcus intermedius</em></td>
</tr>
<tr>
<td>3</td>
<td>Pleural fluid</td>
<td><em>Escherichia coli</em>, <em>Enterococcus</em> sp.</td>
<td><em>Prevotella denticola</em></td>
<td><em>Prevotella denticola</em></td>
</tr>
<tr>
<td>4</td>
<td>Pus</td>
<td><em>Staphylococcus aureus</em></td>
<td>Bacterial 16S rRNA gene not detected</td>
<td><em>Bacterial 16S rRNA gene not detected</em></td>
</tr>
<tr>
<td>5</td>
<td>Pleural fluid</td>
<td>VTS</td>
<td>Bacterial 16S rRNA gene not detected</td>
<td><em>Bacterial 16S rRNA gene not detected</em></td>
</tr>
<tr>
<td>6</td>
<td>Pleural fluid</td>
<td><em>Prevotella sp.</em>, <em>Neisseria cinerea</em>, <em>Haemophilus influenzae</em></td>
<td><em>Prevotella oulorum</em></td>
<td><em>Prevotella oulorum</em></td>
</tr>
<tr>
<td>7</td>
<td>Pleural fluid</td>
<td><em>Pseudomonas aeruginosa</em></td>
<td><em>Streptococcus intermedius</em>, <em>Streptococcus anginosus</em></td>
<td><em>Streptococcus constellatus</em>, <em>Streptococcus intermedius</em></td>
</tr>
<tr>
<td>8</td>
<td>Pleural fluid</td>
<td>Mixed anaerobes</td>
<td><em>Prevotella pleuritidis</em></td>
<td><em>Prevotella pleuritidis</em></td>
</tr>
<tr>
<td>9</td>
<td>Pleural fluid</td>
<td><em>Streptococcus pneumoniae</em></td>
<td><em>Streptococcus pneumoniae</em>†</td>
<td><em>Streptococcus pneumoniae</em>†</td>
</tr>
<tr>
<td>10</td>
<td>Pleural fluid</td>
<td><em>Escherichia coli</em>, VTS</td>
<td>Bacterial 16S rRNA gene not detected</td>
<td><em>Bacterial 16S rRNA gene not detected</em></td>
</tr>
<tr>
<td>11</td>
<td>Pleural fluid</td>
<td>MRSA</td>
<td><em>Staphylococcus aureus</em></td>
<td><em>Staphylococcus aureus</em></td>
</tr>
<tr>
<td>12</td>
<td>Pleural fluid</td>
<td>GPC, VTS</td>
<td><em>Streptococcus intermedius</em>, <em>Streptococcus anginosus</em></td>
<td><em>Streptococcus intermedius</em>, <em>Streptococcus anginosus</em></td>
</tr>
<tr>
<td>13</td>
<td>Pleural fluid</td>
<td>NOS, <em>Acinetobacter</em> sp.</td>
<td><em>Fusobacterium nucleatum</em></td>
<td><em>Fusobacterium nucleatum</em>, <em>Fusobacterium naviforme</em></td>
</tr>
<tr>
<td>14</td>
<td>Pleural fluid</td>
<td><em>Klebsiella oxytoca</em></td>
<td>Bacterial 16S rRNA gene not detected</td>
<td><em>Bacterial 16S rRNA gene not detected</em></td>
</tr>
<tr>
<td>15</td>
<td>Pleural fluid</td>
<td><em>Escherichia coli</em></td>
<td>Bacterial 16S rRNA gene not detected</td>
<td><em>Bacterial 16S rRNA gene not detected</em></td>
</tr>
<tr>
<td>16</td>
<td>Pleural fluid</td>
<td><em>Escherichia coli</em></td>
<td>Bacterial 16S rRNA gene not detected</td>
<td><em>Bacterial 16S rRNA gene not detected</em></td>
</tr>
<tr>
<td>17</td>
<td>Pleural fluid</td>
<td><em>Escherichia coli</em></td>
<td>Bacterial 16S rRNA gene not detected</td>
<td><em>Bacterial 16S rRNA gene not detected</em></td>
</tr>
</tbody>
</table>

*VTS, Viridans type streptococcus; MRSA, meticillin-resistant *Staphylococcus aureus*; NOS, no organisms seen in the Gram stain; GPC, Gram-positive cocci seen in the Gram stain.

†Confirmed by *Streptococcus pneumoniae*-specific PCR targeting the pneumolysin and autolysin genes.
Of the 58 samples that were culture-negative, 1343 bp PCR identified clinically relevant pathogens in 15 samples, while the 762/598 bp PCR identified 25 clinically relevant pathogens (Table 3). Of the PCR-positive samples from both culture-positive and culture-negative samples, the 1343 bp PCR identified 19 of the 23 (83%) bacterial 16S rRNA genes to the species level while the 762/598 bp PCR identified 14 out of 38 (37%) (Tables 2 and 3).

### Comparison of 1343 bp PCR and 762/598 bp PCR sequencing results

Both the 1343 bp PCR and 762/598 bp PCR detected the bacterial 16S rRNA gene in 23 samples. Concordant identification was obtained in 21 samples. The two discordant pleural fluids (samples 6 and 29) were *Prevotella oris* (1343 bp PCR) and *Haemophilus influenzae* (762/598 bp PCR) and *Prevotella oris* (1343 bp PCR) and *Fusobacterium* species (762/598 bp PCR). Culture results were available for sample 6 only and this sample grew a mixture of *Prevotella* species, *Neisseria cinerea* and *H. influenzae*.

### DISCUSSION

The results of this study showed an increase in the number of bacterial infections diagnosed by using the 762/598 bp PCR.
PCR (51%) and 1343 bp PCR (31%) assays compared to culture (23%). However, neither the 1343 bp nor the 762/598 bp PCR protocols detected all significant infections. Failure of the PCR methods to detect the 16S rRNA gene in culture-positive samples may be due to the number of bacteria present in a sample being lower than the detection limit of the method (Harris & Hartley, 2003; Schuurman et al., 2004). Schuurman et al. (2004) also reported PCR-negative and culture-positive samples and in all cases these discrepant samples had low numbers of colonies on primary plates. Alternatively, PCR inhibitors may have been present in the sample but were not detected using our spiked control sample.

It has been reported that Gram-positive bacteria are not detected as efficiently as Gram-negative bacteria, because the Gram-positive cell wall is difficult to break down during extraction (Harris & Hartley, 2003; Schuurman et al., 2000). Rantakokko-Jalava et al. (2000) showed that there was no systematic discrimination of Gram-positive species with their 16S rRNA primers although Zucol et al. (2006) showed that PCR sensitivities varied between 10 and 10\(^6\) c.f.u. for the different bacterial species being investigated. In this study, in samples that were positive by both PCR protocols, sequencing analysis identified different pathogens in the 1343 bp PCR protocol when compared to the 762/598 bp PCR protocol (samples 6 and 29 in Tables 2 and 3). It is possible that more than one type of bacterium was present in these samples and that the different PCRs preferentially amplified different pathogens. This is further evidence that none of the broad-range PCRs are truly universal (Schabereiter-Gurtner et al., 2008; Maskell et al., 2006; Baker et al., 2003).

Discrepant results were also observed when positive PCR results were compared to culture. Pre-culture antibiotics can render samples negative by culture falsely (Rantakokko-Jalava et al., 2000; Schuurman et al., 2004), and some bacteria, such as obligate anaerobes, are difficult to culture or, as in the case of *Streptococcus pneumoniae*, may undergo autolysis (Harris & Hartley, 2003). The 16S rRNA gene sequencing detected a variety of anaerobic bacteria including species of *Prevotella* and *Fusobacterium* (Tables 2 and 3).

Certain bacteria are difficult to identify because they share more than 99% identity in their 16S rRNA gene sequences and this was observed in 15 of the samples analysed in this study. For example, it was not possible to distinguish between the 16S rRNA sequences of *E. coli* and *Shigella* sp. in sample 1 (although, clinically, *E. coli* was thought to be the most likely pathogen in this sample type) or between *Streptococcus intermedius*, *Streptococcus anginosus* or *Streptococcus constellatus* in sample 5. *Streptococcus pneumoniae*, *Streptococcus mitis* and *Streptococcus oralis* are also difficult to distinguish by 16S rRNA gene sequencing and the 16S rRNA identification in samples 9, 25, 27, 35 and 36 was confirmed by a *Streptococcus pneumoniae*-specific PCR targeting the pneumolysin and autolysin genes (Rudolph et al., 1993).

This study shows that performing 16S rRNA PCR assays has the potential to make an important contribution to patient management by detecting the presence of bacterial pathogens in culture-negative clinical samples. The results support the suggestion that the sensitivity of the assay is affected by the size of the 16S rRNA gene fragment amplified. Although not statistically significant, the 762/598 bp PCR assay, amplifying two shorter fragments, detected the bacterial 16S rRNA gene in more samples than the 1343 bp PCR assay; however, this increase was associated with a compromise in the ability of the assay to identify the bacteria to the species level. The 762/598 bp PCR is the method of choice in our laboratory, although the 1343 bp PCR may be employed when a more specific identification would have an impact on patient care.

**REFERENCES**


Schuurman, T., de Boer, R. F., Kooistra-Smid, A. M. D. & van Zwet, A. A. (2004). Prospective study of use of PCR amplification and sequencing of 16S ribosomal DNA from cerebrospinal fluid for...


