

# Is it Essential to Sequence the Entire 16S rRNA Gene for Bacterial Identification?

Sunhee Hong, PhD and  
Christine E. Farrance, PhD

*Charles River*

---

## Introduction

---

Bacterial Identification in the biopharmaceutical industry, especially in manufacturing facilities, is very important because an occurrence of a problematic microorganism in the final product could be harmful for the end user and detrimental to a company's finances and reputation. Environmental Monitoring (EM) programs are the cornerstone of understanding the microbial ecology in a manufacturing facility and have become a regulatory requirement for most manufacturers. The EM program is a biological surveillance system which enables companies to quickly identify organisms which are transient or resident in their facilities before these organisms have an opportunity to contaminate a product. A properly executed EM program provides an early warning of potential contamination problems due to equipment failure, inadequate cleaning, or deficiencies in staff hygiene training, for example, so that problems can be corrected to prevent adulteration of the end product. The Food and Drug Administration (FDA) has published guidelines for the production of sterile drugs by aseptic processing which includes a section on EM programs, and the USP general information chapter "Microbiological Control and Monitoring of Aseptic Processing Environments" also contains detailed information regarding EM programs<sup>1</sup>.

The EM program is only effective if the organisms recovered from the facility are accurately identified, so the information gathered can be used to understand the microbial control through tracking and trending and dictate appropriate remediation activities. There are several different options available for bacterial identification; however, the use of 16S rRNA gene sequences has been considered the most powerful and accurate tool, while conventional phenotypic methods often show major weaknesses<sup>2-5</sup>. The entire 16S rRNA gene sequence is approximately 1500 base pairs (bp) and consists of highly conserved regions which provide a broad taxonomic spectrum, and nine hypervariable regions (V1 – V9) that allow high taxonomic level discrimination<sup>6,7</sup>. For the publication and taxonomic classification

for novel bacterial species, the full length of the 16S rRNA gene sequence followed by phylogenetic relationship analysis with other closely related species is necessary<sup>8</sup>. For the identification of bacteria, it is reported that the full length 16S rRNA gene sequence analysis may provide better resolution for certain species<sup>4,9,10</sup>. On the other hand, Clarridge (2004)<sup>4</sup> claimed that for most clinical bacterial isolates, evaluating the first 500 bp is sufficient for identification. The first 500 bp may even show increased resolution between certain species since this region can show slightly more diversity than the remaining sequence<sup>4</sup>. Traditionally, sequence-based identifications made on samples from EM programs have been made based on the first 500 bp of the 16S rRNA gene. Use of the entire 1500 bp versus the first 500 bp 16S rRNA gene sequence in bacterial identification has attracted interest from the biopharmaceutical industry, but the number of studies and the data supporting this option are limited.

The purpose of this study was to examine the hypothesis that the first 500 bp of the 16S rRNA gene is sufficient for identification of bacteria recovered during the execution of EM programs.

## Study Approach

A total of 208 diverse “in house” sequences covering 131 genera were randomly selected for further analysis. The sequences for both the first 500 bp (5F to 531R) and the nearly full gene sequences (5F to 1492R) were compared against the EzTaxon database. EzTaxon is a public database containing type strains with valid published prokaryotic names and representative sequences from uncultured phylotypes<sup>11</sup>. The identification of phylogenetic neighbors was initially carried out by the BLASTN<sup>12</sup> program against the EzTaxon database. The top 30 sequences with the highest scores were selected for the calculation of pairwise sequence similarity using a global alignment algorithm<sup>13</sup>, which was implemented on the EzTaxon server (<http://www.ezbiocloud.net/eztaxon><sup>11</sup>). After obtaining the alignment results for each sample’s sequence, the information for 30-60 phylogenetic neighbors was downloaded for phylogenetic analysis. A phylogenetic

**Table 1. Presumptive identification and the category as indicated below**

Presumptive ID	Category	Presumptive ID	Category
<i>Achromobacter denitrificans</i>	b	<i>Acidovorax konjaci</i>	a
<i>Acidovorax temperans</i>	a	<i>Actinobacillus hominis</i>	a
<i>Actinomyces howellii</i>	a	<i>Aeromonas enteropelogenes</i>	b
<i>Aeromonas hydrophila hydrophila</i>	a	<i>Aeromonas veronii</i>	b
<i>Afipia felis</i>	a	<i>Alishewanella agri</i>	d
<i>Alteromonas macleodii</i>	d	<i>Aminobacter ciceronei</i>	d
<i>Amycolatopsis mediterranei</i>	a	<i>Anaerococcus hydrogenalis</i>	a
<i>Anaerococcus prevotii</i>	a	<i>Aquaspirillum polymorphum</i>	a
<i>Arcanobacterium haemolyticum</i>	a	<i>Arthrobacter agilis</i>	a
<i>Asticcacaulis excentricus</i>	a	<i>Atopobium minutum</i>	a
<i>Azohydromonas lata</i>	a	<i>Azospirillum brasilense</i>	a
<i>Bacillus cereus</i>	b	<i>Bacillus coagulans</i>	a
<i>Bacillus horikoshii</i>	a	<i>Bacillus pseudofirmus</i>	a
<i>Bacillus xiamenensis</i>	a	<i>Bartonella henselae</i>	d
<i>Bifidobacterium animalis lactis</i>	a	<i>Bifidobacterium boum</i>	a
<i>Bifidobacterium catenulatum</i>	a	<i>Bifidobacterium pseudolongum globosum</i>	a
<i>Brachybacterium paraconglomeratum</i>	a	<i>Brevibacillus agri</i>	a
<i>Brevibacterium epidermidis</i>	a	<i>Brochothrix campestris</i>	a
<i>Brochothrix thermosphacta</i>	a	<i>Burkholderia andropogonis</i>	a
<i>Burkholderia gladioli</i>	a	<i>Burkholderia plantarii</i>	a
<i>Campylobacter concisus</i>	a	<i>Campylobacter mucosalis</i>	a
<i>Campylobacter upsaliensis</i>	a	<i>Capnocytophaga cynodegmi</i>	a
<i>Capnocytophaga haemolytica</i>	a	<i>Capnocytophaga ochracea</i>	a
<i>Chromatocurvus halotolerans</i>	b	<i>Citrobacter koseri</i>	a
<i>Citrobacter sedlakii</i>	d	<i>Citrobacter youngae</i>	c
<i>Clostridium sporogenes</i>	b	<i>Comamonas thiooxydan</i>	a
<i>Corynebacterium accolens</i>	a	<i>Corynebacterium stationis</i>	a
<i>Corynebacterium tuberculostearicum</i>	a	<i>Cronobacter dublinensis dublinensis</i>	a
<i>Cronobacter malonaticus</i>	a	<i>Cronobacter muytensii</i>	a
<i>Cronobacter sakazakii</i>	a	<i>Cronobacter turicensis</i>	a
<i>Dactylosporangium roseum</i>	b	<i>Deinococcus proteolyticus</i>	a
<i>Deinococcus radiopugnans</i>	a	<i>Dermacoccus nishinomiyaensis</i>	a
<i>Dermatophilus congolensis</i>	a	<i>Devosia riboflavina</i>	a
<i>Dichotomicrobium thermohalophilum</i>	a	<i>Effusibacillus pohliae</i>	a
<i>Enterobacter asburiae</i>	a	<i>Enterobacter cancerogenus</i>	d
<i>Enterobacter cloacae dissolvens</i>	a	<i>Enterobacter hormaechei</i>	a
<i>Enterobacter ludwigii</i>	a	<i>Enterobacter xiangfangensis</i>	d
<i>Enterococcus faecalis</i>	a	<i>Enterococcus faecium</i>	d
<i>Enterococcus hirae</i>	a	<i>Enterococcus sulfureus</i>	a
<i>Erwinia rhapontici</i>	a	<i>Erysipelothrix tonsillarum</i>	a
<i>Escherichia coli</i>	b	<i>Escherichia fergusonii</i>	a
<i>Escherichia vulneris</i>	a	<i>Eubacterium multiforme</i>	a
<i>Eubacterium tenue</i>	a	<i>Exiguobacterium acetylicum</i>	a
<i>Flavobacterium pectinovorum</i>	a	<i>Fluoribacter gormanii</i>	a
<i>Fusobacterium naviforme</i>	a	<i>Fusobacterium nucleatum animalis</i>	a
<i>Gemella haemolysans</i>	b	<i>Gemella morbillorum</i>	a

**Table 1. Presumptive identification and the category as indicated below (continued)**

Presumptive ID	Category	Presumptive ID	Category	Presumptive ID	Category
<i>Geobacillus kaustophilus</i>	b	<i>Geobacillus stearothermophilus</i>	a	<i>Geobacillus thermocatenuatus</i>	b
<i>Geobacillus thermodenitrificans</i>	b	<i>Gluconobacter cerinus</i>	a	<i>Gordonia amarae</i>	a
<i>Gordonia</i> sp.	b	<i>Gordonia sputi</i>	a	<i>Haemophilus paracuniculus</i>	a
<i>Halomonas venusta</i>	a	<i>Hamadaea tsunoensis</i>	a	<i>Hyphomicrobium zavarzinii</i>	d
<i>Insolitipirillum peregrinum peregrinum</i>	a	<i>Kocuria flava</i>	a	<i>Kocuria kristinae</i>	a
<i>Kytococcus sedentarius</i>	a	<i>Lactobacillus dextrinicus</i>	a	<i>Lactobacillus pentosus</i>	b
<i>Lactobacillus rhamnosus</i>	a	<i>Leclercia adecarboxylata</i>	a	<i>Legionella erythra</i>	a
<i>Leuconostoc citreum</i>	a	<i>Leuconostoc mesenteroides cremoris</i>	b	<i>Listeria grayi</i>	a
<i>Loktanelia sediminilitoris</i>	a	<i>Maricaulis maris</i>	a	<i>Methylobacterium mesophilicum</i>	b
<i>Micromonospora coerulea</i>	a	<i>Micromonospora nigra</i>	a	<i>Micromonospora sagamiensis</i>	b
<i>Moraxella lacunata</i>	a	<i>Moraxella lincolnii</i>	a	<i>Mycobacterium caprae</i>	b
<i>Mycobacterium chitae</i>	a	<i>Mycoplasma orale</i>	a	<i>Neisseria weaveri</i>	a
<i>Neptuniibacter caesariensis</i>	a	<i>Nocardia otitidiscaviarum</i>	a	<i>Oceanobacillus kimchii</i>	a
<i>Ochrobactrum cytisi</i>	b	<i>Ochrobactrum pseudogrignonense</i>	a	<i>Paenibacillus thiaminolyticus</i>	a
<i>Paenibacillus urinalis</i>	a	<i>Pantoea eucrina</i>	a	<i>Pantoea septica</i>	a
<i>Pantoea</i> sp.	b	<i>Paracoccus sphaerophysae</i>	a	<i>Pasteurella bettyae</i>	a
<i>Pasteurella caballi</i>	a	<i>Pasteurella canis</i>	a	<i>Pasteurella dagmatis</i>	a
<i>Pediococcus acidilactici</i>	a	<i>Peptococcus niger</i>	a	<i>Photorhabdus luminescens luminescens</i>	a
<i>Planococcus rifietoensis</i>	a	<i>Prevotella zoogeleoformans</i>	a	<i>Propionibacterium acnes</i>	a
<i>Providencia heimbachae</i>	a	<i>Pseudoalteromonas carrageenovora</i>	a	<i>Pseudomonas aeruginosa</i>	a
<i>Pseudomonas balearica</i>	a	<i>Pseudomonas beteli</i>	a	<i>Pseudomonas boreopolis</i>	a
<i>Pseudomonas carboxydohydrogena</i>	a	<i>Pseudomonas japonica</i>	a	<i>Pseudomonas libanensis</i>	b
<i>Pseudorhodobacter ferrugineus</i>	a	<i>Pseudoxanthomonas taiwanensis</i>	a	<i>Psychrobacter phenylpyruvicus</i>	a
Rhizobiales	b	<i>Rhodococcus erythropolis</i>	a	<i>Roseomonas aestuarii</i>	a
<i>Rothia mucilaginoso</i>	a	<i>Rouxiella chamberiensis</i>	a	<i>Saccharothrix coeruleofusca</i>	a
<i>Salimicrobium halophilum</i>	a	<i>Salinispora tropica</i>	a	<i>Salmonella enterica</i>	b
<i>Serratia ficaria</i>	d	<i>Serratia marcescens marcescens</i>	a	<i>Shewanella algae</i>	b
<i>Shewanella hanedai</i>	a	<i>Sphingobacterium psychroaquaticum</i>	a	<i>Sphingobium yanoikuyae</i>	a
<i>Sphingopyxis macrogoltabida</i>	c	<i>Sporosarcina pasteurii</i>	a	<i>Sporosarcina ureae</i>	a
<i>Staphylococcus argenteus</i>	b	<i>Staphylococcus aureus aureus</i>	b	<i>Staphylococcus auricularis</i>	a
<i>Staphylococcus caprae</i>	b	<i>Staphylococcus epidermidis</i>	a	<i>Stenotrophomonas maltophilia</i>	a
<i>Streptococcus pyogenes</i>	a	<i>Streptococcus sanguinis</i>	a	<i>Streptomyces castelarensis</i>	b
<i>Streptomyces specialis</i>	a	<i>Streptomyces vinaceusdrappus</i>	b	<i>Streptosporangium album</i>	b
<i>Streptosporangium vulgare</i>	b	<i>Tatlockia micdadei</i>	a	<i>Tatumella ptyseos</i>	a
<i>Tetrasphaera duodecadis</i>	a	<i>Trabulsiella guamensis</i>	a	<i>Trueperella bernardiae</i>	a
<i>Trueperella pyogenes</i>	a	<i>Tsakamurella inchonensis</i>	a	<i>Tsakamurella paurometabola</i>	a
<i>Turicella otitidis</i>	a	<i>Ureibacillus terrenus</i>	a	<i>Ureibacillus thermosphaericus</i>	a
<i>Vasilyevaeva enhydra</i>	a	<i>Vasilyevaeva enhydra</i>	a	<i>Vibrio pelagius</i>	d
<i>Yersinia rohdei</i>	a	<i>Yersinia rohdei</i>	a	<i>Yersinia ruckeri</i>	a
<i>Yokenella regensburgei</i>	a				

a) the species-level identification was the same using either the first 500 or the 1500 bp of 16S rRNA gene sequences, b) the results were the same when using the first 500 bp or the entire 1500 bp of the 16S rRNA gene sequences, but neither query provided an identification at the species level c) species-level identification was designated, but there were different identifications resulting from the first 500 and 1500 bp of the 16S rRNA gene sequences, d) the first 500 bp of 16S rRNA gene sequence could not provide an identification at the species level, but the nearly full length 1500 bp sequence was able to assign a species-level identification.

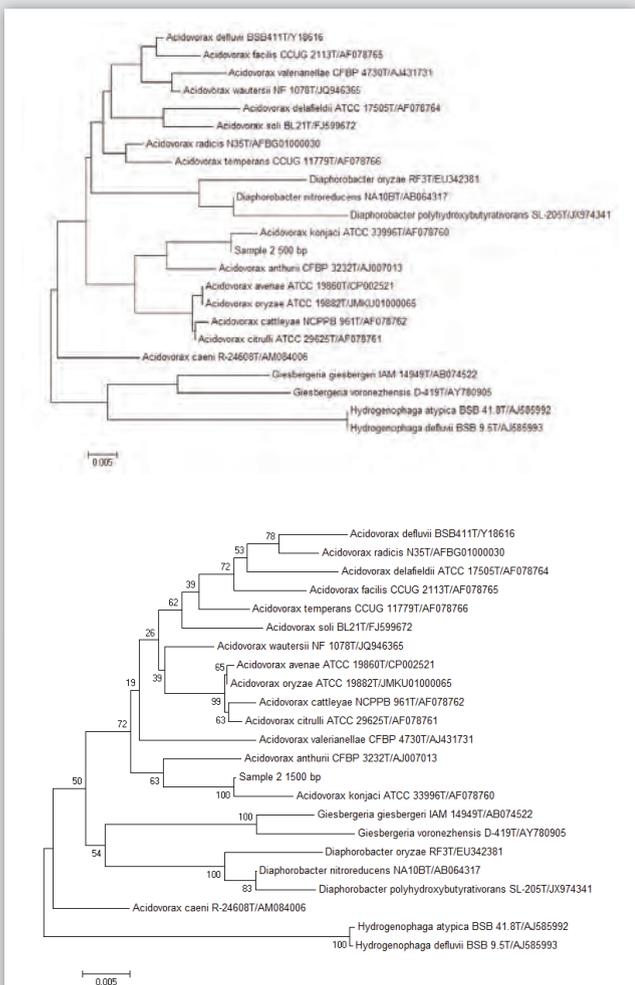
tree was constructed for each sample based on the neighbor-joining method using the MEGA software package version 6<sup>14</sup> and the

evolutionary distance was calculated based on the Jukes & Cantor model<sup>15</sup>.

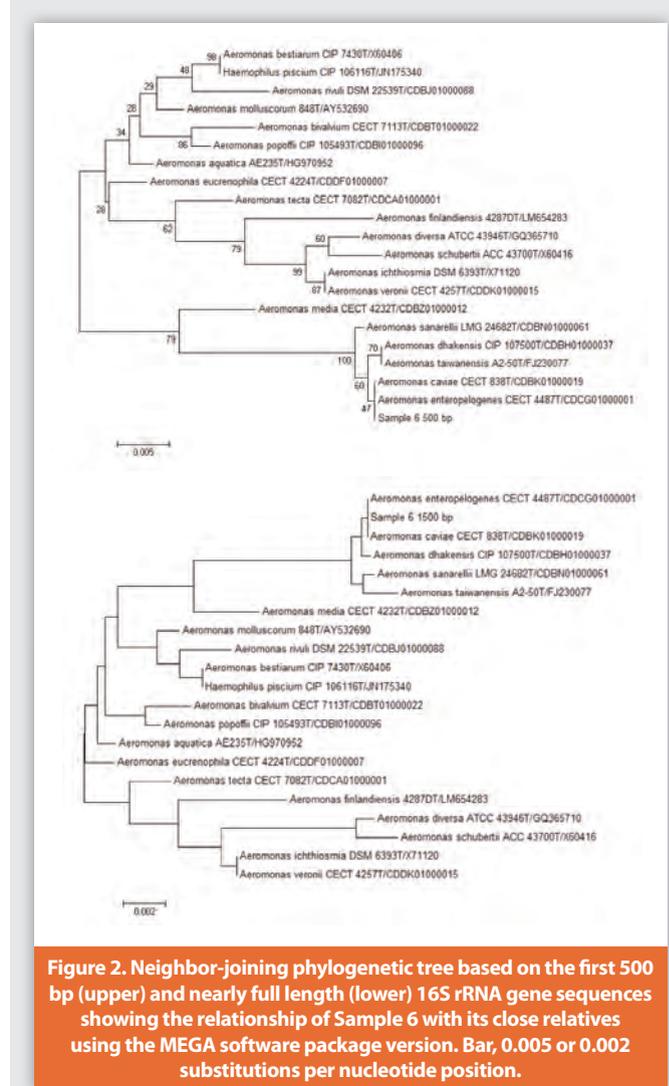
**Table 2. The number of sequences and the percentage of samples in each category resulting from 500 bp and 1500 bp of the 16S rRNA gene sequence analysis.**

Category	Number of Sequences	Ratio of each category (%)	Summary (%)
a	164	78.8	93.7
b	28	13.5	
	3	1.4	
c	2	1.0	3.6
d	11	5.3	
Total	208	100.00	100.00

a) the species-level identification was the same using either the first 500 or the 1500 bp of 16S rRNA gene sequences, b) the results were the same when using the first 500 bp or the entire 1500 bp of the 16S rRNA gene sequences, but neither query provided an identification at the species level c) species-level identification was designated, but there were different identifications resulting from the first 500 and 1500 bp of the 16S rRNA gene sequences, d) the first 500 bp of 16S rRNA gene sequence could not provide an identification at the species level, but the nearly full length 1500 bp sequence was able to assign a species-level identification.



**Figure 1. Neighbor-joining phylogenetic tree based on the first 500 bp (upper) and nearly full length (lower) 16S rRNA gene sequences showing the relationship of Sample 2 with its close relatives using the MEGA software package version. Bar, 0.005 substitutions per nucleotide position.**

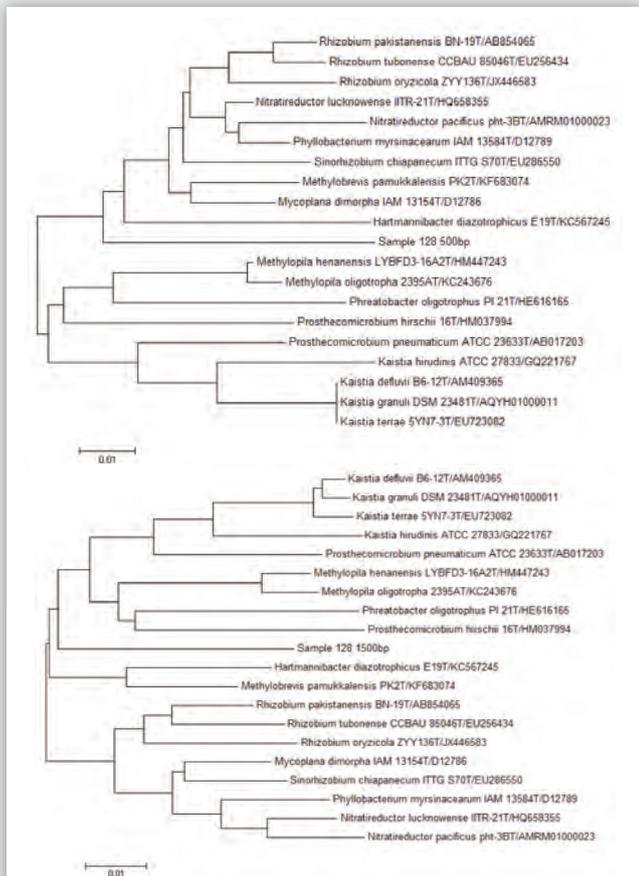


**Figure 2. Neighbor-joining phylogenetic tree based on the first 500 bp (upper) and nearly full length (lower) 16S rRNA gene sequences showing the relationship of Sample 6 with its close relatives using the MEGA software package version. Bar, 0.005 or 0.002 substitutions per nucleotide position.**

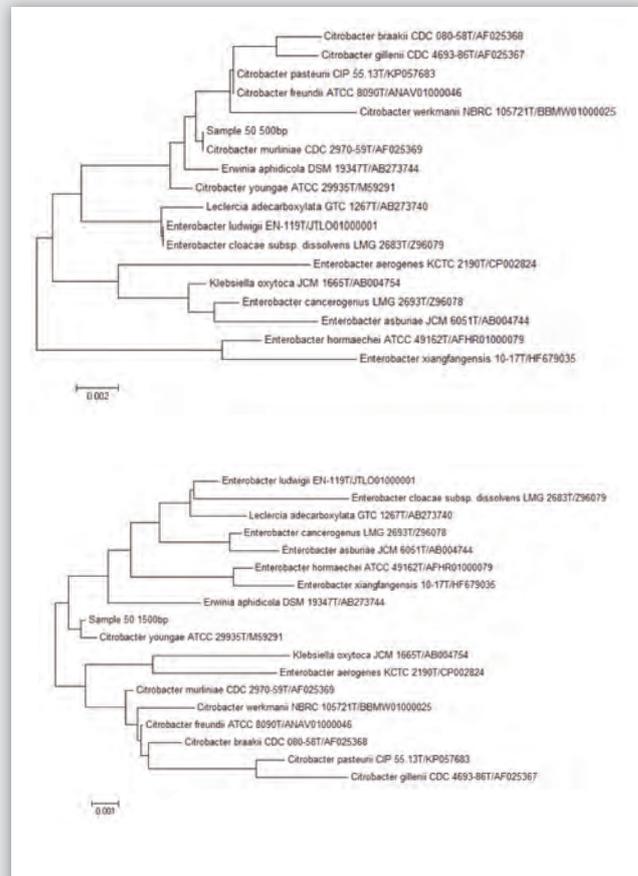
## Results and Discussion

### Comparison of the identification generated by the phylogenetic analyses using the first 500 and 1500 bp 16S rRNA gene sequences

Table 1 shows the presumptive identification for all 208 sequences that were evaluated. The results were categorized into 4 different cases: a) the species-level identification was the same using either the first 500 or the 1500 bp of 16S rRNA gene sequences, b) the results were the same when using the first 500 bp or the entire 1500 bp of the 16S rRNA gene sequences, but neither query provided an identification at the species level due to i) inability to differentiate two or more species that were closely related to the unknown sequence or ii) classification was to the genus or order level, c) species-level identification was designated, but there were different identifications resulting from the first 500 and 1500 bp of the 16S



**Figure 3.** Neighbor-joining phylogenetic tree based on the first 500 bp (upper) and nearly full length (lower) 16S rRNA gene sequences showing the relationship of Sample 128 with its close relatives using the MEGA software package version. Bar, 0.01 substitutions per nucleotide position.



**Figure 4.** Neighbor-joining phylogenetic tree based on the first 500 bp (upper) and nearly full length (lower) 16S rRNA gene sequences showing the relationship of Sample 50 with its close relatives using the MEGA software package version. Bar, 0.002 or 0.001 substitutions per nucleotide position.

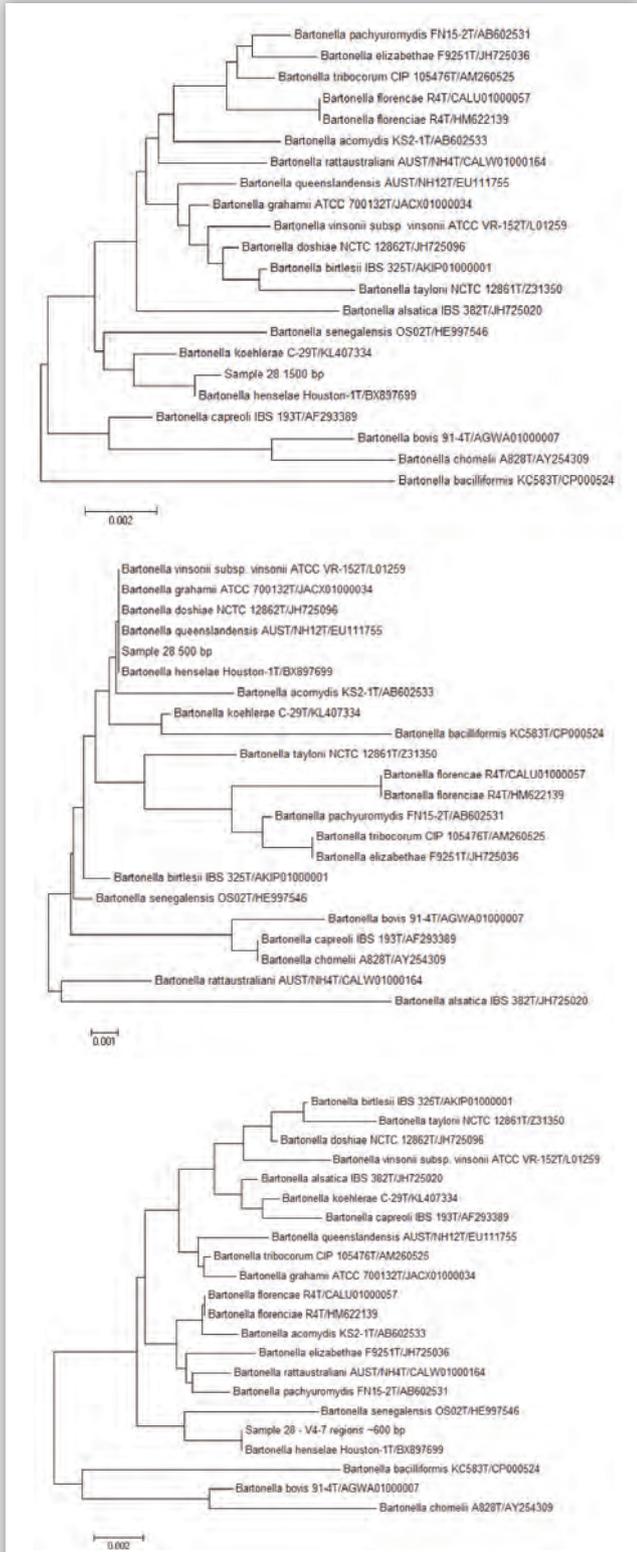
rRNA gene sequences, and *d*) the first 500 bp of 16S rRNA gene sequence could not provide an identification at the species level, but the nearly full length 1500 bp sequence was able to assign a species-level identification.

### Comparison of the results obtained from the first 500 and 1500 bp of 16S rRNA gene

Among the 208 sequences assessed in this study, we found that 93.7% of the samples resulted in the same identification regardless of which sample sequence, 500 bp or 1500 bp, was used for the initial comparison query (combined Categories “a” and “b”, Table 2). Of this, 78.8% resulted in the same exact match to a single species (Category a), 13.5% of the queries had the same outcome but neither the 500 bp nor the 1500 bp 16S rDNA sequence was able to result in a definitive species designation, that is the unknown matched two or more closely related species (Category b-i), and the remaining 1.4% of the unknowns also have the same outcome but were only able to be identified to the genus or order level (Category b-ii).

In the case of Category “a” isolates, both the 500 bp sequence and the nearly full length gene sequence clearly showed the same presumptive identification at the species level. Figure 1 illustrates the phylogenetic relationships for Sample 2, which fell into Category “a”. Both the first 500 and 1500 bp of 16S rRNA gene sequences directly linked to the type strain for *Acidovorax konjaci* with more than a 99% sequence similarity.

In most of the cases for which the 16S rRNA gene sequence data cannot provide definitive resolution (Category b-i and b-ii) it is because the next neighbor(s) are very closely related to the unknown sequence. Neither the first 500 bp nor the 1500 bp 16S rRNA gene sequences could provide resolution. An example is shown with Sample 6 in Figure 2. It is not surprising that we observed 28 cases (13.5%, Category b-i) where we could not separate two or more species even using the entire 16S rRNA gene sequence analysis (Table 2). In fact, this is the major criticism of utilizing the 16S rDNA gene sequence for bacterial identification. It is often cited that 16S rDNA sequencing presents difficulties in identification for certain groups of bacteria because they exhibit indistinguishable 16S rRNA gene sequences<sup>16-18</sup>. In these cases,



**Figure 5. Neighbor-joining phylogenetic tree based on the nearly full length (upper), first 500 bp (middle) 16S rRNA gene sequences, and V4-V7 region (lower) showing the relationship of Sample 28 with its close relatives using the MEGA software package version. Bar, 0.002, 0.001 or 0.002 substitutions per nucleotide position.**

using an alternative marker such as a protein coding gene, preferably a house-keeping gene, could help discriminate closely related species<sup>19-21</sup>.

We also observed 3 cases for which it was not possible to conclusively identify the sample at the species level (Category *b*-ii, Table 2). Two samples showed genus-level identifications while one sample (number 128) showed a low taxonomic, "Order", level identification (Fig 3). The results of the analysis did not change when utilizing the first 500 bp or nearly full length 16S rDNA sequence for the analysis, which indicate the possibility of a new species.

Interestingly, we found two instances which showed different presumptive identification at the species level between the first 500 bp and nearly full length of 16S rRNA gene sequences (Category "c"). Figure 4 shows the unexpected results for Sample 50. Both the 500 bp and the 1500 bp analysis resulted in an identification in the genus *Citrobacter*; however, the species designated were different. In general, GenBank is an adequate reference database; however, the sequences and taxonomic classification in GenBank are not curated and not always valid. Thus, there are many sequences that include uncertain bases (N) or missing bases so the quality of data from GenBank is not acceptable for many sequence entries. Even though we used the EzTaxon database for our study, the origin of the sequences in EzTaxon is GenBank. Indeed, we have observed a significant number of sequences used in our phylogenetic analyses which showed errors such as a missing gap, incorrectly inserted base or uncertain base. We suspect that these two samples in Category "c" may have occurred due to poor quality sequences from EzTaxon being used in the phylogenetic analyses. Further investigation of this case will need to be initiated.

Eleven instances of Category "d" in this study showed that for 5.3% of the samples the 1500 bp of the 16S rRNA gene sequences had better resolution than the first 500 bp alone. The nearly full gene was able to provide a definitive species designation, while the first 500 bp of the 16S rRNA gene sequence matched two or more closely related species (Table 2). Figure 5 illustrates one of the cases, Sample 28. For many *Bartonella* species, evaluation of the entire 16S gene sequence is required. All the *Bartonella* species are well separated from one another in the 1500 bp sequence phylogenetic analysis (Fig 5, upper panel), while many species in this genus share the same sequence in their first 500 bp (Fig 5, middle panel). As stated earlier, some individuals expect the entire 16S rRNA gene sequence would always have better resolution than the first 500 bp sequence. However, this is evidently not correct, although it does occur, as shown here. Additionally, with the desire for high throughput and prompt time-to-result, sequencing the entire 16S rRNA gene is not a practical approach<sup>4,22</sup>. Thus, we strive to determine alternative markers which can overcome the limitations of 16S rDNA sequencing. Previously, we suggested the use of a protein coding gene for discrimination in closely related species which have the same 16S rRNA gene sequences. However, in cases of very closely related organisms such as within the *Bartonella* genus, using different regions of the 16S rRNA gene for identification can provide improved resolution. For example, the first 500 bp of the 16S rRNA gene contains the V1 to V3 variable regions which were not able to separate all *Bartonella* species. However, the use of different regions such as the

region containing V4 to V7 (approximately 600 bp) exhibits the same discrimination power as the entire 16S rRNA gene sequence (Fig 5, lower panel). Thus facilitating implementation of the sequencing of the alternate target within today's requirements for rapid, high volume sample processing.

## Conclusion

While it is correct that in some instances (5.3% of the samples tested in this study) the entire 16S rRNA gene sequence will provide better resolution than the first 500 bp, there are also many cases (93.7% of the samples in this study) in which the entire 16S rRNA gene sequence does not provide any additional information over that obtained with the first 500 bp sequence. The result for both input sequences can be a clear species identification or no definitive answer because the closely related species have a high percentage of sequence similarity. To the best of our knowledge, our work represents the first comparison study between the use of the first 500 bp and the entire 1500 bp sequence of the 16S rRNA gene in bacterial identification for a diverse sample collection. The number of samples evaluated in this study cannot represent all bacterial species diversity. Therefore, the frequency of each categorical result will change as more samples are analyzed and added. However, it will be surprising if the ratio of each case changes significantly. The overall performance of the first 500 bp sequence of the 16S rRNA gene, compared to the entire 1500 bp sequence for identification, is very high; 93.7% of the samples showed no differences between the two approaches. Considering these data and the discussion of the limitation of using the 16S rRNA gene sequence in general, generating the full length sequence data for the 16S rRNA gene is inefficient and impractical. When using the first 500 bp of the 16S rRNA gene sequence for identification and additional resolution is needed, targeting alternate regions with increased variability, whether within the 16S rRNA gene or in a protein coding gene, makes more scientific and operational sense. It must be noted, the quality of sequences used for phylogenetic analysis is critical. Finally, other factors not discussed here, such as the importance of a curated reference database and the overarching issue with current bacterial nomenclature and taxonomy, can have an impact on the identification of microorganisms. Considering all these factors, the sequence from the first 500 bp of the 16S rRNA gene is sufficient for bacterial identification.

## References

1. <http://www.rapidmicrobiology.com>
2. Gerner-Smidt, P., I. Tjernberg, and J. Ursing, Reliability of phenotypic tests for identification of *Acinetobacter* species. *J Clin Microbiol*, 1991. 29(2): p. 277-82.
3. Mohania, D., et al., Molecular approaches for identification and characterization of lactic acid bacteria. *J Dig Dis*, 2008. 9(4): p. 190-8.
4. Clarridge, J.E., 3rd, Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev*, 2004. 17(4): p. 840-62.
5. Boivin-Jahns, V., et al., Comparison of phenotypic and molecular methods for the identification of bacterial strains isolated from a deep subsurface environment. *Appl Environ Microbiol*, 1995. 61(11): p. 4140.
6. Mizrahi-Man, O., E.R. Davenport, and Y. Gilad, Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS One*, 2013. 8(1): p. e53608.
7. Cox, M.J., W.O. Cookson, and M.F. Moffatt, Sequencing the human microbiome in health and disease. *Hum Mol Genet*, 2013. 22(R1): p. R88-94.
8. Stackebrandt, E., et al., Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol*, 2002. 52(Pt 3): p. 1043-7.
9. Chakravorty, S., et al., A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods*, 2007. 69(2): p. 330-9.
10. Janda, J.M. and S.L. Abbott, 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol*, 2007. 45(9): p. 2761-4.
11. Kim, O.S., et al., Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int J Syst Evol Microbiol*, 2012. 62(Pt 3): p. 716-21.
12. Altschul, S.F., et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 1997. 25(17): p. 3389-402.
13. Myers, E.W. and W. Miller, Optimal alignments in linear space. *Comput Appl Biosci*, 1988. 4(1): p. 11-7.
14. Tamura, K., et al., MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol*, 2013. 30(12): p. 2725-9.
15. Jukes, T. H. & Cantor, C. R. Evolution of protein molecules. In *Mammalian protein metabolism*. 1969. vol. 3, pp. 21-132. Edited by H. N. Munro. New York: Academic Press.
16. Ash, C., Farrow, J. A., Dorsch, M., Stackebrandt, E. & Collins, M. D. Comparative analysis of *Bacillus anthracis*, *Bacillus cereus*, and related species on the basis of reverse transcriptase sequencing of 16S rRNA. *Int J Syst Bacteriol* 1991. 41, 343-346.
17. Martínez-Murcia, A. J., Benlloch, S. & Collins, M. D. Phylogenetic interrelationships of members of the genera *Aeromonas* and *Plesiomonas* as determined by 16S ribosomal DNA sequencing: lack of congruence with results of DNA-DNA hybridization. *Int J Syst Bacteriol* 1992. 42, 412-421.
18. Christensen, H., Nordentoft, S. & Olsen, J. E. Phylogenetic relationships of *Salmonella* based on rRNA sequences. *Int J Syst Bacteriol* 1998. 48, 605-610.
19. Yamamoto, S. & Harayama, S. Phylogenetic relationships of *Pseudomonas putida* strains deduced from the nucleotide sequences of *gyrB*, *rpoD* and 16S rRNA genes. *Int J Syst Bacteriol* 1998. 48, 813-819.
20. Wang, L.-T., Lee, F.-L., Tai, C.-J. & Kasai, H. Comparison of *gyrB* gene sequences, 16S rRNA gene sequences and DNA-DNA hybridization in the *Bacillus subtilis* group. *Int J Syst Evol Microbiol* 2007. 57, 1846-1850.
21. Payne, G. W., Vandamme, P., Morgan, S. H., LiPuma, J. J., Coenye, T., Weightman, A. J., Jones, T. H. & Mahenthalingam, E. Development of a *recA* gene-based identification approach for the entire *Burkholderia* genus. *Appl Environ Microbiol* 2005. 71, 3917-3927.
22. Patel, J. B. 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Mol. Diagn.* 2001. 6:313-321.