

# 16S Direct workflow: Microbial identification using 16S gene sequencing on the SeqStudio Genetic Analyzer

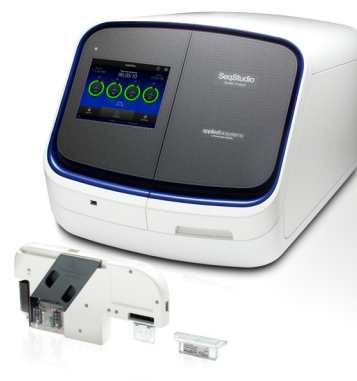
## In this application note we present:

- A fast and economical workflow called “16S Direct” for bacterial identification at the species level by PCR and Sanger sequencing of the 16S rRNA gene
- The use of the Applied Biosystems™ BigDye™ Direct kit for high-resolution Sanger sequencing
- The use of the innovative Applied Biosystems™ SeqStudio™ Genetic Analyzer for capillary electrophoresis
- The use of the freely available Applied Biosystems™ MicrobeBridge™ software with link to the US Centers for Disease Control (CDC) and MicrobeNet for data analysis

## Introduction

Rapid and accurate identification of infectious, fastidious, or noncultivable bacteria is a major challenge for clinical practice and public health surveillance. Based on the groundbreaking phylogenetic research work of Woese et al. [1] and others, sequencing of the 16S ribosomal RNA (rRNA) gene has emerged as the preferred method for taxonomic classification and identification of bacteria. Today, many products and literature reviews demonstrate the use of sequencing of the 16S gene as the gold standard for identification at the species level (for reviews see references 2 and 3). There are numerous protocols and a published guideline [4] for PCR and sequencing of the 16S gene for research and applied use. In addition, several databases from public institutions and commercial companies are available for comparing 16S sequences with typed or annotated entries.

The Applied Biosystems™ MicroSEQ™ 16S gene PCR sequencing kit and MicroSEQ™ database have long been trusted for microbial identification in pharmaceutical production settings. Although the MicroSEQ system works well on high-throughput instruments, some investigators or laboratories may not need high-powered solutions.



The recently introduced Applied Biosystems™ SeqStudio™ Genetic Analyzer provides a low-cost alternative to high-throughput instruments.

In this application note, we present a novel, fast, and economical 16S gene sequencing workflow called “16S Direct” that is optimized for use on the SeqStudio Genetic Analyzer.

## SeqStudio Genetic Analyzer and 16S Direct workflow

The SeqStudio Genetic Analyzer is an affordable capillary electrophoresis (CE) system that can perform both automated Sanger DNA sequencing and high-precision sizing and analysis of multicolor fluorescent DNA fragments. The instrument can accommodate a 96-well plate loaded with sequencing reactions that are sequentially electrophoresed in batches of four samples per run. The sequencing samples can be processed in a 30 minute “short” run cycle, resulting in DNA sequences of about 500 bases each. This is sufficient to reconstruct the almost complete sequence (~1.5 kb) of the 16S gene and thus identify the species of a particular sample in only 4 sequencing reactions and 1 run.

The procedure described here can be adapted to individual needs, such as different primer designs or running the sequencing reactions on a CE instrument such as an Applied Biosystems™ 3500 series genetic analyzer. Other established or commercial 16S sequencing methods and protocols using Applied Biosystems™ BigDye™ sequencing chemistry will also work on the SeqStudio instrument, but may require more steps and time to prepare additional CE runs as more sequencing reactions for assembling the complete gene sequence are needed (Table 1).

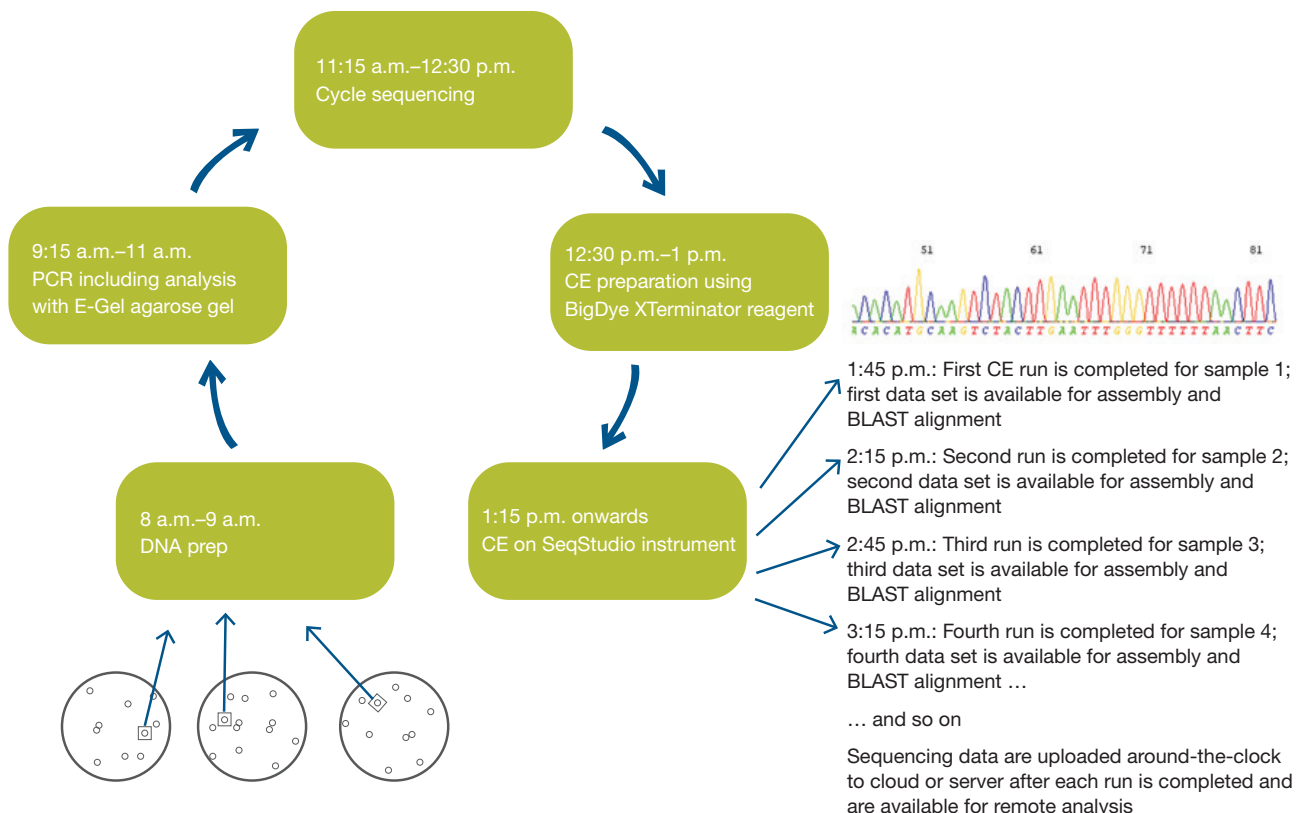
**Table 1. Number of PCR reactions, sequencing reactions, and runs on the SeqStudio Genetic Analyzer (30 min each) required to identify a number of bacteria after 4 hr CE time (arbitrary time point).**

	No. of PCR reactions	No. of sequencing reactions	No. of CE runs needed on SeqStudio instrument	No. of bacteria identified after 4 hr (i.e., after 8 CE runs)
16S Direct workflow	2	4	1	8
MicroSEQ Full Gene 16S rDNA Sequencing Kit	3	6	2	5
Traditional lab-developed protocol	1	8	2	4

The basic idea behind the 16S Direct workflow is to reduce the number of sequencing reactions to 4 in order to fully use the running capacity of 4 capillaries on the SeqStudio instrument. This maximizes the throughput of samples within a given time window.

### The 16S Direct workflow

An overview and an example timeline of the 16S Direct workflow are shown in Figure 1 and in more detail in Figure 2. Starting in the morning, individual bacterial colonies are picked from agar culture plates and PCR-ready DNA is extracted using, for example, the Applied Biosystems™ PrepMan™ Ultra Sample Preparation Reagent. PCR reactions “A” and “B” are then set up followed by a quick check for quantity and expected size of the two amplicons by standard agarose or Invitrogen™ E-Gel™ agarose gels. A total of 4 sequencing reactions are then set up using the BigDye Direct sequencing reagent and cycled. The reactions are then purified with Applied Biosystems™ BigDye™ XTerminator™ reagent, and the samples are electrophoresed on the SeqStudio Genetic Analyzer in the early afternoon. The instrument continuously generates a data set of 4 sequences every 30 min when operating in short-run mode. The sequence files are then immediately available for secondary data analysis using MicrobeBridge software and BLAST™ alignment with a microbial database.



**Figure 1. The 16S Direct workflow—an example time line.**

DNA	<ul style="list-style-type: none"> <li>• Prepare gDNA from bacterial isolate</li> <li>• Use your preferred or verified method to prepare PCR-ready DNA</li> </ul>	
PCR	<ul style="list-style-type: none"> <li>• Set up PCR reactions “A” and “B” with BigDye Direct kit (as described in “Materials and methods” section)</li> </ul>	Time (manual): 2–10 min Time (on PCR cycler): 60 min
E-Gel agarose gel	<ul style="list-style-type: none"> <li>• Check PCR success with an E-Gel or other agarose gel</li> </ul>	Time (manual): 2–5 min Time (on device): 10–15 min
Sequencing	<ul style="list-style-type: none"> <li>• Set up forward and reverse sequencing reactions with amplicons “A” and “B” using the BigDye Direct sequencing kit</li> </ul>	Time (manual): 2–10 min Time (on PCR cycler): 70 min
BigDye XTerminator reagent	<ul style="list-style-type: none"> <li>• Purify sequencing reaction with BigDye XTerminator reagent</li> <li>• Transfer supernatant to CE plate</li> </ul>	Time (manual): 2–5 min Time (on device): 20–30 min
CE	<ul style="list-style-type: none"> <li>• Set up CE run on the SeqStudio Genetic Analyzer</li> <li>• Short run (each run of 4 sequencing samples = 1 specimen) or</li> <li>• Long run (each run of 4 sequencing samples = 1 specimen)</li> </ul>	Time (manual): 1–5 min Time (on instrument): 30 min Time (on instrument): 100 min
MicrobeBridge software	<ul style="list-style-type: none"> <li>• Assemble sequencing files with MicrobeBridge software</li> <li>• Copy into PC memory</li> </ul>	Time (manual): 1–10 min
BLAST alignment	<ul style="list-style-type: none"> <li>• Paste into preferred 16S database (e.g., NCBI microbial genomes or CDC MicrobeNet)</li> <li>• Review or verify top hit</li> </ul>	Time (manual): 1–2 min Time (intellectual): 1–5 min

Figure 2. Steps of the 16S Direct workflow.

## Materials and methods

### DNA

The genomic DNA (gDNA) used for this research was from bacterial strains of known species obtained from ATCC and diluted to 10 ng/μL in Invitrogen™ TE buffer (Tris 10 mM, EDTA 0.1 mM). To prepare PCR-ready DNA from bacteria growing on an agar plate or other biological or environmental matrix, the PrepMan Ultra Sample Preparation Reagent is a suitable product that is fast, economical, and easy to use. Other options are described at [thermofisher.com/gdna](http://thermofisher.com/gdna)

### PCR

For PCR of the 16S rRNA gene, the following primers were custom ordered from [thermofisher.com/oligos](http://thermofisher.com/oligos):

16S\_A\_FWD (F8): 5'-tgtaaacgacgccagcAGAGTTTGATC-MTGGCTCAG-3'

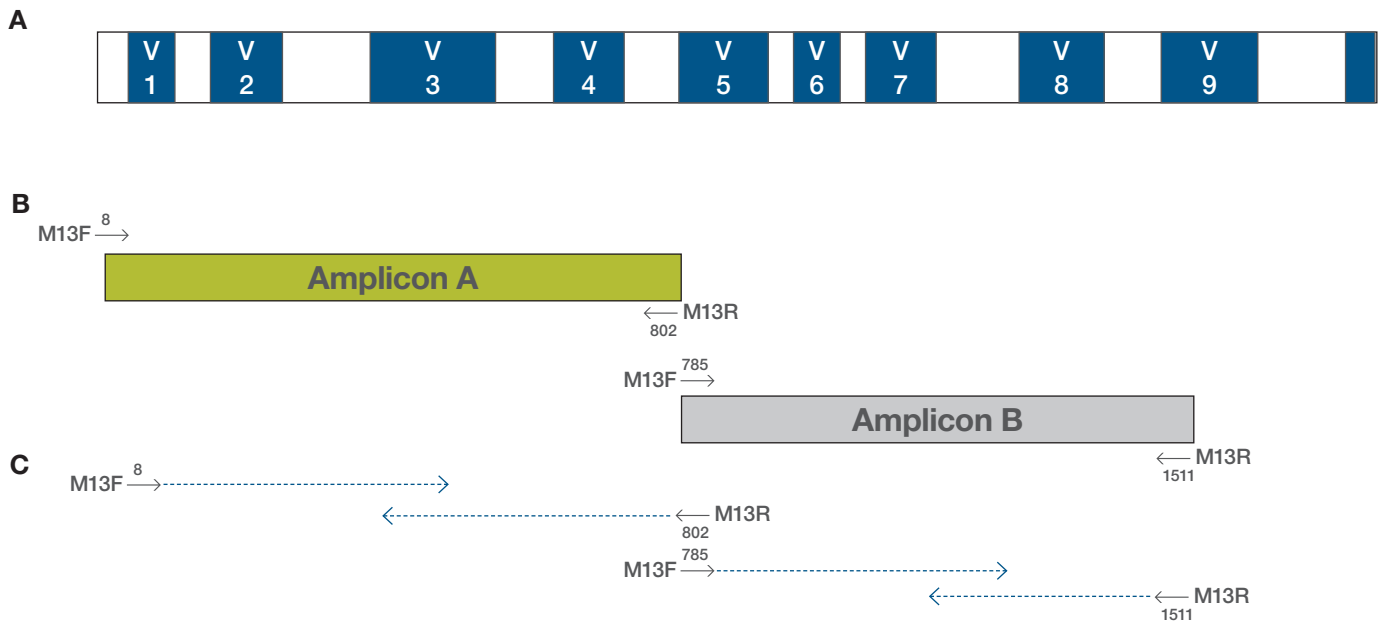
16S\_A\_REV (R802): 5'-caggaaacagctatgaccTACCAGGG-TATCTAATCC-3'

16S\_B\_FWD (F785): 5'-tgtaaacgacgccagcGGATT-AGATACCCTGGTA-3'

16S\_B\_REV (R1511): 5'-caggaaacagctatgaccCGGT-TACCTTGTTACGACTT-3'

PCR set “A” amplifies the 5′ half of the 16S rRNA gene from position 8 to 802, and PCR set “B” amplifies the 3′ half of the 16S rRNA gene from position 785 to 1,511 (Figure 3). Forward PCR primers are modified at the 5′ end with the M13 forward primer sequence and likewise, reverse PCR primers with the M13 reverse primer sequence. M13 primer sequences are shown in lowercase letters. Note that the internal 16S gene-specific primers are complementary (Figure 3) and thus form a short overlapping segment of 18 bp that can be used to link segments “A” and “B” for full gene sequence assembly using the MicrobeBridge software. The advantage of this particular primer design is that only 4 sequencing reactions are required for near full-length coverage of the 16S gene, which matches the 4-capillary capacity of the SeqStudio instrument for a single run.

For PCR, 5 μL of the BigDye Direct PCR reagent was combined with 3.5 μL water and 0.5 μL primer pair “A” or “B” (each primer at 10 μM) and 1 μL DNA (typically 1–10 ng) and amplified in an Applied Biosystems™ ProFlex™ PCR System using these cycling conditions: initial hot start at 95°C for 5 min followed by 8 cycles at 95°C for 15 sec, 50°C for 15 sec, and 68°C for 90 sec, followed by 27 cycles at 95°C for 15 sec, 65°C for 15 sec, and 72°C for 90 sec.



**Figure 3. Sequencing primer-binding sites.** Schematic representation of the 16S rRNA gene showing (A) variable regions (V1–V9) with high sequence diversity between species, and general and conserved regions (white segments) that are targets for the M13-tagged primers. The segment between nucleotides 785–802 where the sequences overlap serves as a primer-binding site for the (B) PCR amplicons “A” and “B”. (C) The M13 tags are used as sequencing primer-binding sites.

After PCR, 15  $\mu\text{L}$  of low-EDTA TE buffer was added to the reaction and 2.5  $\mu\text{L}$  (~10%) was used for electrophoresis on a 2% E-Gel agarose gel to verify the presence of amplicon “A” (~830 bp) or amplicon “B” (~750 bp) as a clearly visible single band in an estimated amount of 20–50 ng.

### High-resolution Sanger sequencing with the BigDye Direct sequencing reagent

The BigDye Direct kit is comprehensive and economical, containing PCR reagent (see above), BigDye Direct

sequencing reagent, and unique M13 sequencing primers that bind to the M13 forward and reverse sequences present in the PCR amplicon. The BigDye Direct reagent provides high resolution of the nucleotide sequences close to the 5′ end and does not require a separate PCR primer purification step, thereby saving time and extra reagents. A sample setup of 4 cycle sequencing reactions for sequencing M13-tagged PCR amplicons “A” and “B” in both directions is shown in Table 2.

**Table 2. Setting up the forward and reverse BigDye Direct sequencing reactions.** Note that the Applied Biosystems™ BigDye™ 5X sequencing buffer is not included in the BigDye Direct kit and must be obtained separately. The 5X buffer is part of the standard BigDye sequencing v1.1 or v3.1 kits or can be purchased as a stand-alone item.

Reagent	Volume for 2 forward sequencing reactions (1 specimen)	Volume for 2 reverse sequencing reactions (1 specimen)
Molecular biology–grade water (not included in BigDye Direct kit)	12 $\mu\text{L}$	12 $\mu\text{L}$
BigDye v1.1 and v3.1 sequencing buffer 5X (not included in the BigDye Direct kit)	3 $\mu\text{L}$	3 $\mu\text{L}$
BigDye Direct M13 forward primer (included in the BigDye Direct kit)	1 $\mu\text{L}$	–
BigDye Direct M13 reverse primer (included in the BigDye Direct kit)	–	1 $\mu\text{L}$
BigDye Direct sequencing reagent (included in the BigDye Direct kit)	2 $\mu\text{L}$	2 $\mu\text{L}$
Total volume	18 $\mu\text{L}$ (=2 x 9 $\mu\text{L}$ )	18 $\mu\text{L}$ (=2 x 9 $\mu\text{L}$ )
	↓	↓
Well or tube 1	9 $\mu\text{L}$ forward sequencing reaction + 1 $\mu\text{L}$ amplicon A	
Well or tube 2	9 $\mu\text{L}$ reverse sequencing reaction + 1 $\mu\text{L}$ amplicon A	
Well or tube 3	9 $\mu\text{L}$ forward sequencing reaction + 1 $\mu\text{L}$ amplicon B	
Well or tube 4	9 $\mu\text{L}$ reverse sequencing reaction + 1 $\mu\text{L}$ amplicon B	

Cycle sequencing on a ProFlex PCR System was performed using this profile: 1 cycle comprised of 37°C for 15 min, 80°C for 2 min, and 96°C for 1 min; 13 cycles at 96°C for 10 sec, 50°C for 5 sec, and 60°C for 1 min 15 sec; 6 cycles at 96°C for 10 sec, 50°C for 5 sec, and 60°C for 1 min 30 sec; 6 cycles at 96°C for 10 sec, 50°C for 5 sec, and 60°C for 2 min 30 sec. After cycle sequencing, 50 µL of BigDye XTerminator suspension was added to each sequencing reaction. It consisted of 1 part BigDye Xterminator beads (10 µL/sample) and 4 parts of SAM Solution (40 µL/sample) and was prepared in bulk for all samples shortly before use. The sequencing plate with the BigDye XTerminator suspension-treated samples was vigorously vortexed for 30 min and then centrifuged for 2 min to pellet the beads; 20 µL of the supernatant was transferred to a fresh plate, centrifuged for 1 min to remove any residual air bubbles, and then placed on the SeqStudio Genetic Analyzer for CE.

### Sequencing by CE on the SeqStudio Genetic Analyzer

The samples can be run either as short (30 min), medium (45 min), or long runs (105 min) using BigDye Direct chemistry and a regular sequencing run module (note: do not use a BigDye XTerminator run module). Long runs yield nearly complete bidirectional coverage, whereas short runs yield approximately complete, 40% bidirectional coverage. Complete and accurate assembly of near full-length 16S gene (~1,450 bases) sequences can be reliably achieved with high-quality (quality value (QV) >35) short-run data.

After completion of a CE run cycle, the 4 sequencing files for PCR segments “A” and “B” are immediately available for data analysis and also remotely retrievable through internet cloud connectivity. This helps enable urgent identification analysis away from the lab or after lab hours.

### Data analysis

The output files from the SeqStudio Genetic Analyzer are sequencing files in .ab1 format, which can be viewed in electropherogram trace mode or in text mode. In order to reconstitute the entire sequence of the 16S rRNA gene from both strands of amplicons “A” and “B”, the .ab1 files need to be assembled, i.e., an overlap needs to be established and a contiguous consensus sequence (“contig”) needs to be generated. This can be conveniently done using the MicrobeBridge v.1.1 software, which is available as free download at [thermofisher.com/order/catalog/product/A29238](http://thermofisher.com/order/catalog/product/A29238). Briefly, the workflow is as follows:

1. Import and organize the 4 associated sequence files (i.e., forward and reverse for amplicon “A” and “B”) for each specimen.
2. Click on the “Review” tab, which initiates base calling and the assembly process that results in an overview of the sequence assembly (Figure 4), details of the traces (Figure 5), and quality metrics.

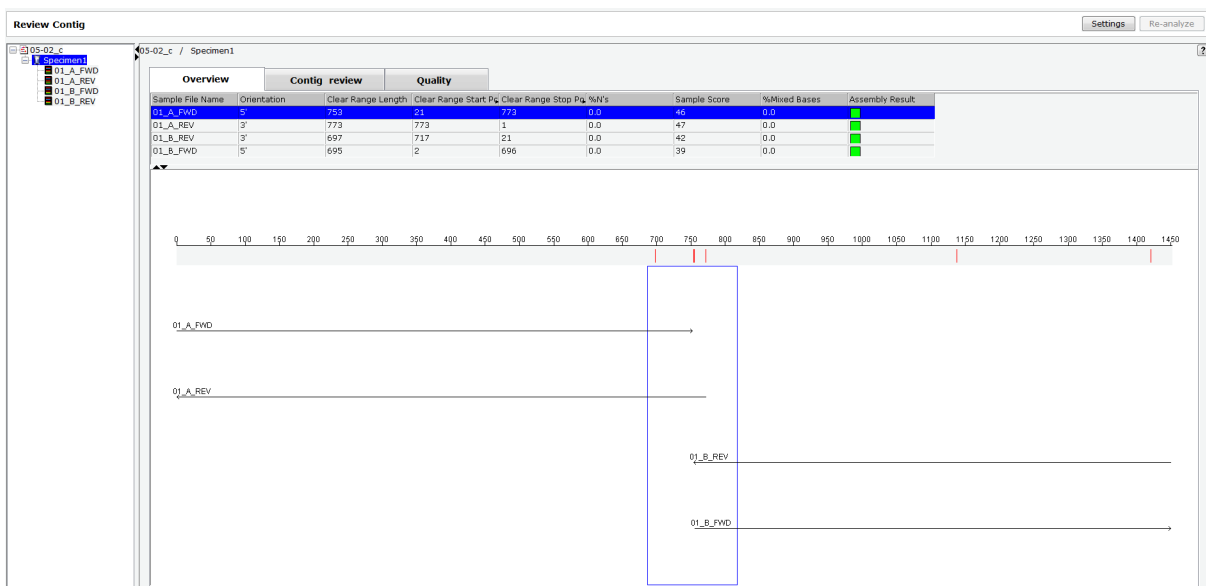


Figure 4. Overview of the assembly of the 4 sequencing traces: forward and reverse for amplicon “A” and forward and reverse for amplicon “B”.

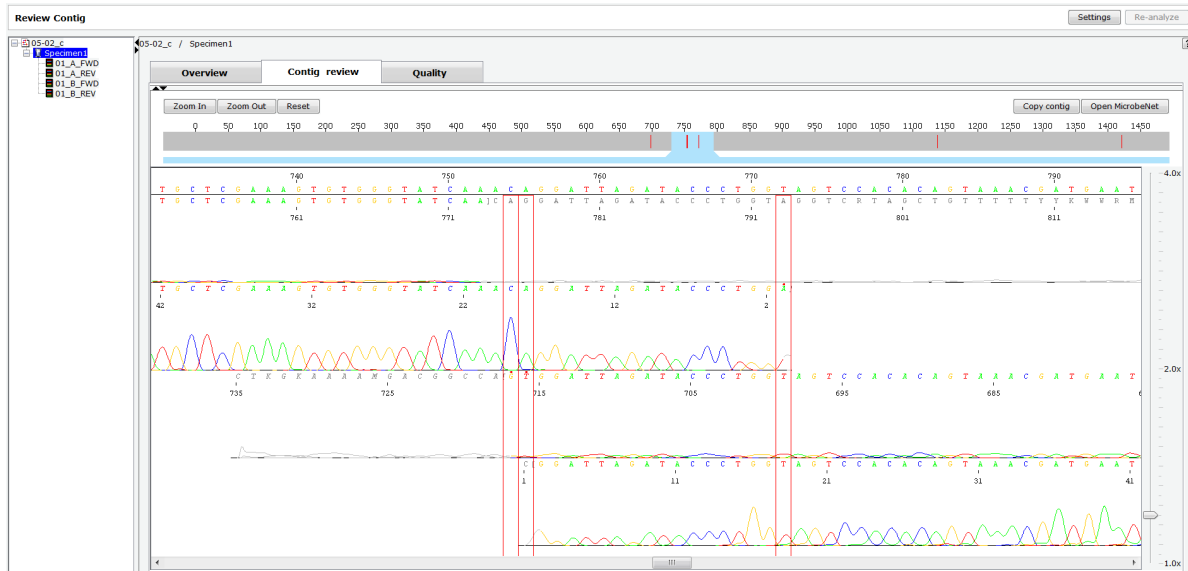


Figure 5. Trace view of the junction between amplicons “A” and “B”.

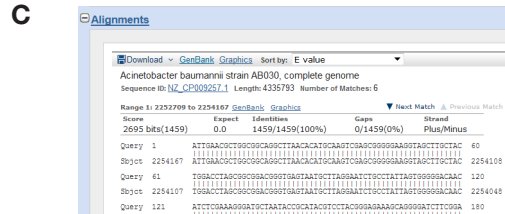
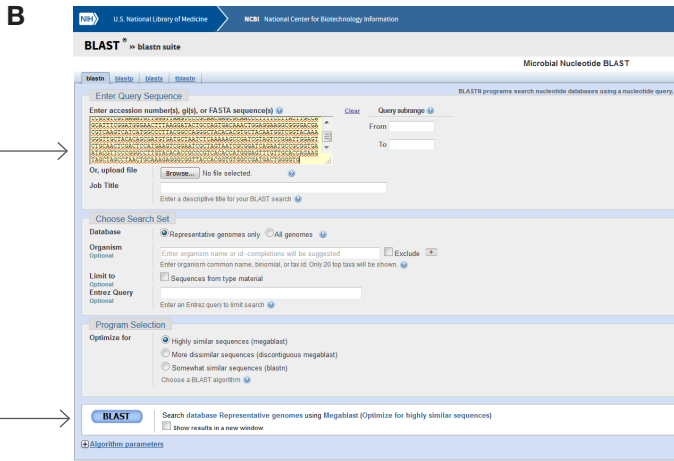
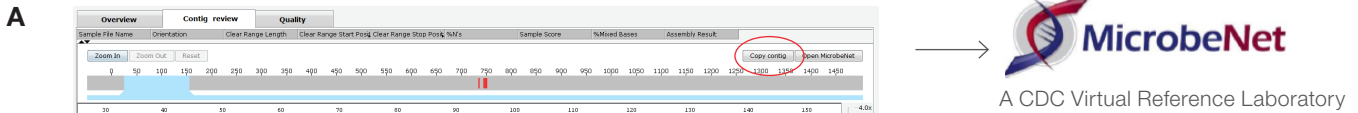
1. Review the assembly output, which enables correcting any miscalls in the consensus sequence. Ambiguities in the assembly are indicated by a red bar (Figure 4), which can be inspected and edited in the rare event that it is necessary. The MicrobeBridge assembler usually determines the consensus base call by the sequence trace with the highest quality and thus the lowest probability of error. Note the correct orientation of the forward and reverse traces for amplicons “A” and “B” (Figure 4) and verify that the overall length of the assembled sequence reaches ~1,450 bases.
2. When the assembly results look satisfactory, open the “Contig review” tab and click “Copy contig” (Figure 6A). This will copy the assembled sequence onto the memory clipboard for subsequent pasting into a BLAST alignment of personal preference. For feasibility of demonstration in this application note, we have used the NCBI Microbial Nucleotide BLAST tool, which compares a query sequence to fully sequenced microbial genomes for identification (Figure 6B and C). For additional information, please refer to [blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE\\_TYPE=BlastSearch&BLAST\\_SPEC=MicrobialGenomes](http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=MicrobialGenomes).

The “Open MicrobeNet” button (Figure 6A) provides a link to the CDC MicrobeNet portal for pasting the query sequence into the CDC localized alignment tool for identification and further detailed microbiological information on the lead hit. This free and useful public service is provided by the CDC at [microbenet.cdc.gov/](http://microbenet.cdc.gov/).

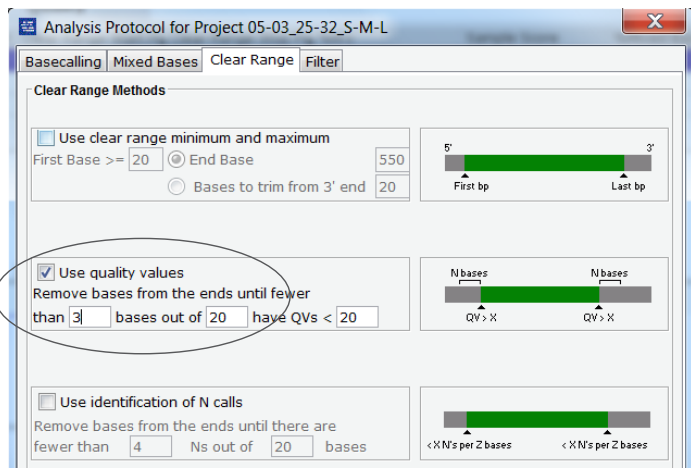
### Adjusting the trimming stringency to facilitate alignment and contig assembly

When the overlap in the center between amplicons “A” and “B” is shorter than 10 nucleotides, the alignment fails and a scrambled assembly with a cluster of red bars will occur. This undesired outcome can often be rectified by adjusting (i.e., relaxing) the “clear range” trim parameters in the project analysis settings (Figure 7). By default, the clear range is set to 3 bases, which is fairly stringent and sufficient to generate a consensus sequence between the amplicon “A” forward and “A” reverse strands or likewise amplicon “B” forward and “B” reverse strands. Due to the relatively short sequence overlap between the 3’ end of segment “A” and the 5’ end of segment “B”, the default trimming can sometimes undercut this limit when the sequence quality of the terminal bases is not ideal. In this case, the alignment can be achieved by setting the “clear range” to a higher value (e.g., 4, 5, 6) until the alignment is successful. In rare cases, an alignment and assembly between amplicons “A” and “B” may not be achieved. In such cases, the species identification by nucleotide alignment has to be done separately with amplicons “A” and “B”, which should produce the same result. The only difference with an alignment with a joint “A + B” segment is that the total number of aligned nucleotides will be 15–20 nt shorter (i.e., missing the overlap segment between “A” and “B” consisting of the primer region F785–R802 for which the sequence is known).





**Figure 6. Data analysis workflow.** (A) The “Copy contig” button (circled in red) copies the assembled consensus sequence onto a memory clipboard ready for pasting into either (B) the NCBI Microbial Nucleotide BLAST portal or any other search application. For identification, the top hit of the results is selected and (C) the species ID is shown with alignment and metrics (e.g., % identity). Alternatively, using the “Open MicrobeNet” button (6A →), the contig can be pasted into the CDC MicrobeNet BLAST tool for identification and further information on the species.



**Figure 7. Accessing the project analysis settings.** The project analysis settings can be accessed by clicking the “Settings” button. The stringency of the base trimming can be relaxed by increasing the default value from “3” to “4”, “5”, “6”, or “7”.

## Results

### Long sequencing runs

To demonstrate the feasibility of the primer design and workflow described above, we have processed DNA derived from 6 representative microbial organisms belonging to 5 different phyla: Bacteroidetes, Fusobacteria, Proteobacteria, Actinobacteria, Firmicutes, and Verrucomicrobia. Table 3 shows a summary of the long-read sequencing results from the SeqStudio Genetic Analyzer. Sequencing metrics for each of the 4 sequencing reactions are listed.

The quartet of sequencing files was analyzed by MicrobeBridge software in 3 ways:

1. Consensus of forward and reverse reads of PCR amplicon “A” (sequencing reactions 1 and 2)
2. Consensus of forward and reverse reads of PCR amplicon “B” (sequencing reactions 3 and 4)
3. Consensus of forward and reverse reads of PCR segments “A and B” (sequencing reactions 1, 2, 3, and 4).

**Table 3. Summary of long CE runsequence data metrics, alignment results, and identification of the microbial DNA test panel representing 5 bacterial phyla.** “Sample score” is a consolidated quality measure: the target is a value of 35 or higher. CRL is contiguous read length in nucleotides; QV20 is the longest number of nucleotides with a minimum quality value (QV) of 20 (i.e., probability of error of 1 in 100); RFU is the average signal strength of all 4 dye traces in relative fluorescent units. The MicrobeBridge software trim setting is the value set as the clear range for trimming end bases in the analysis protocol (Figure 7). Top ID is the first and highest-ranking listing in the result table of the BLAST search.

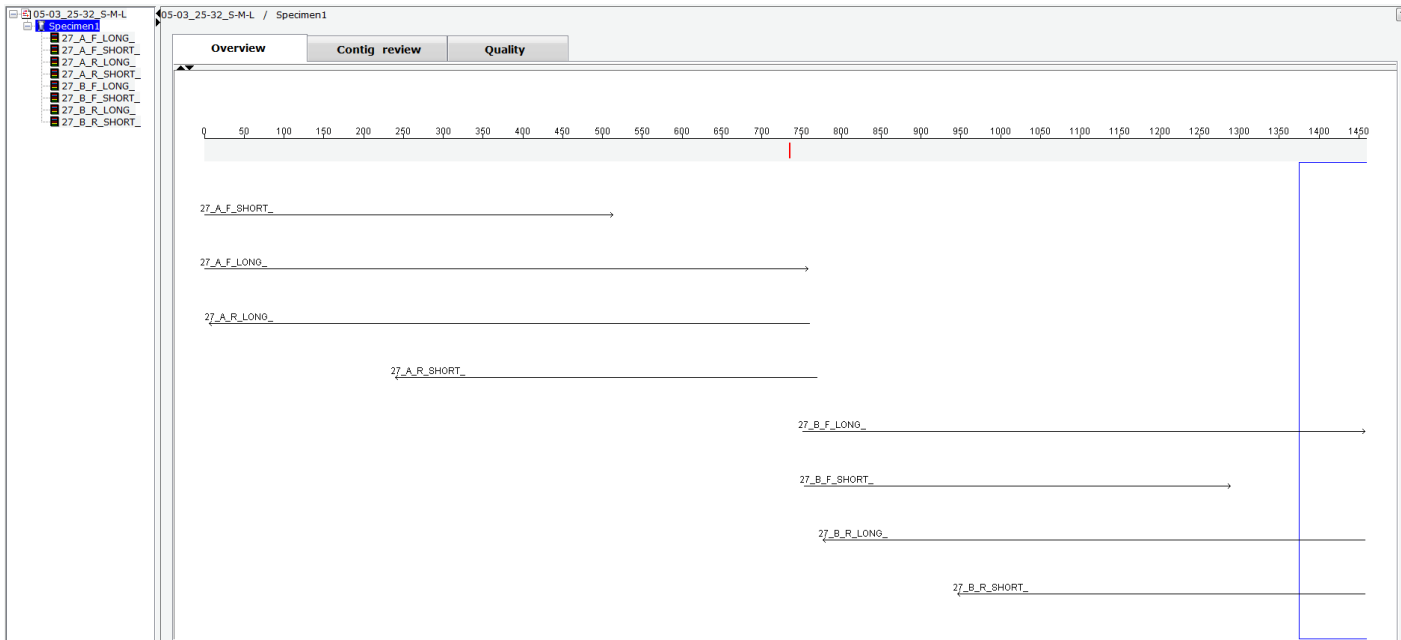
Sample number, organism of DNA input	PCR primer set	Sequencing primer	Sequencing trace No.	SeqStudio run mode	Sequencing trace metrics				MicrobeBridge software	BLAST results				
					Sample score	CRL	QV20+	Signal RFU		Trim setting for contig assembly	Top ID	Identity (%)	No. of nt aligned	No. of gaps
1 <i>Bacteroides fragilis</i> ATCC 25285D-5	A	Forward	1	Long sequence	49	772	756	1,388	}	3 (default)	<i>Bacteroides fragilis</i>	99.87%	756	0
		Reverse	2	Long sequence	47	783	780	1,099						
	B	Forward	3	Long sequence	39	717	703	710	}	3 (default)	<i>Bacteroides fragilis</i>	99.56%	677	0
		Reverse	4	Long sequence	43	711	705	688						
	Sequencing traces No. 1 + 2 + 3 + 4 for contig assembly BLAST into NCBI microbial portal									5	<i>Bacteroides fragilis</i>	99.72%	1,451	0
	2 <i>Fusobacterium nucleatum</i> ATCC 25286D-5	A	Forward	1	Long sequence	47	760	761	1,012	}	3 (default)	<i>Fusobacterium nucleatum</i>	99.73%	734
Reverse			2	Long sequence	47	761	749	763						
B		Forward	3	Long sequence	43	720	710	782	}	3 (default)	<i>Fusobacterium nucleatum</i>	100%	690	0
		Reverse	4	Long sequence	48	729	724	551						
Sequencing traces No. 1 + 2 + 3 + 4 for contig assembly BLAST into NCBI microbial portal									3 (default)	<i>Fusobacterium nucleatum</i>	99.86%	1,442	0	
3 <i>Acinetobacter baumannii</i> ATCC 17978D-5		A	Forward	1	Long sequence	34	53	329	1,088	}	3 (default)	<i>Acinetobacter baumannii</i>	99.73%	749
	Reverse		2	Long sequence	47	787	777	975						
	B	Forward	3	Long sequence	43	724	717	877	}	3 (default)	<i>Acinetobacter baumannii</i>	100%	689	0
		Reverse	4	Long sequence	50	697	708	533						
	Sequencing traces No. 1 + 2 + 3 + 4 for contig assembly BLAST into NCBI microbial portal									3 (default)	<i>Acinetobacter baumannii</i>	99.86%	1,456	0
	4 <i>Actinomyces naeslundii</i> ATCC 12104D-5	A	Forward	1	Long sequence	34	54	133	762	}	3 (default)	<i>Actinomyces oris</i>	97.80%	773
Reverse			2	Long sequence	47	806	794	656						
B		Forward	3	Long sequence	43	722	710	942	}	3 (default)	<i>Actinomyces oris</i>	99.28%	692	2
		Reverse	4	Long sequence	49	720	718	617						
Sequencing traces No. 1 + 2 + 3 + 4 for contig assembly BLAST into NCBI microbial portal									3 (default)	<i>Actinomyces oris</i>	98.52%	1,483	3	
4 <i>Actinomyces naeslundii</i> ATCC 12104D-5		A	Forward	1	Long sequence	34	54	133	762	}	3 (default)	<i>Actinomyces naeslundii</i>	100%	773
	Reverse		2	Long sequence	47	806	794	656						
	B	Forward	3	Long sequence	43	722	710	942	}	3 (default)	<i>Actinomyces naeslundii</i>	100%	692	0
		Reverse	4	Long sequence	49	720	718	617						
	Sequencing traces No. 1 + 2 + 3 + 4 for contig assembly BLAST into NCBI standard nucleotide database									3 (default)	<i>Actinomyces naeslundii</i>	100%	1,483	0
	5 <i>Akkermansia muciniphila</i> ATCC BAA-835D-5	A	Forward	1	Long sequence	48	761	760	611	}	3 (default)	<i>Akkermansia muciniphila</i>	100%	725
Reverse			2	Long sequence	50	754	759	2,067						
B		Forward	3	Long sequence	32	722	683	245	}	3 (default)	<i>Akkermansia muciniphila</i>	100%	689	0
		Reverse	4	Long sequence	43	701	703	285						
Sequencing traces No. 1 + 2 + 3 + 4 for contig assembly blast into NCBI microbial portal									4	<i>Akkermansia muciniphila</i>	99.79%	1,434	0	
6 <i>Streptococcus agalactiae</i> ATCC BAA-1138D-5		A	Forward	1	Long sequence	47	770	760	558	}	3 (default)	<i>Streptococcus agalactiae</i>	100%	761
	Reverse		2	Long sequence	46	793	776	2,719						
	B	Forward	3	Long sequence	46	719	714	2,863	}	3 (default)	<i>Streptococcus agalactiae</i>	100%	689	0
		Reverse	4	Long sequence	48	725	729	5,812						
	Sequencing traces No. 1 + 2 + 3 + 4 for contig assembly BLAST into NCBI microbial portal									4	<i>Streptococcus agalactiae</i>	99.79%	1,471	0



Each consensus sequence was aligned using the NCBI Microbial Genomes BLAST tool that contains the genome sequences of over 7,000 bacterial species. All BLAST queries yielded the expected species identification at very high (>99.5%) identity rate with the exception of sample no. 4. Here the *Actinomyces naeslundii* sample was identified as *Actinomyces oris* but with lower identity values and with gaps in the alignment. These data result from an anomaly in the sequence deposition in databases. At the time these experiments were performed, the genome for *A. naeslundii* was not entered in the NCBI Microbial Genomes database yet, but the 16S rRNA gene sequence was available in the NCBI standard nucleotide database. When these sequences were aligned using the NCBI standard nucleotide database, the correct identification at a 100% identity match was obtained. Therefore, great care should be taken to choose the appropriate and valid database for the 16S rRNA gene sequence queries as certain microbial species may differ by only a few nucleotides in their 16S genes, may be incorrectly listed, may not be included at all, or may be listed under a different species name.

### Short sequencing runs

The same set of DNA samples used for long sequencing runs was again used for another PCR or sequencing run, and electrophoresed using the short sequencing run mode on the SeqStudio instrument. The short sequencing run only takes about 30 minutes for the 4 sequencing reactions and yields a total of over 2,000 nucleotides with high-quality values. About 40% of the sequences covered both forward- and reverse-direction strands and about 60% of the sequence covered only one direction (Figure 8). The sample score values for the single-coverage portion were usually greater than 40 and thus exhibited very high confidence in base call accuracy. Table 4 shows that all samples were identified correctly above 99.5% identity rate using the consensus sequences for PCR segments “A” or “B” alone or in combination: “A joined with B”. This demonstrates that a full 16S gene sequence can be obtained for a given specimen with a short CE run of 30 minutes (compared to about 105 minutes with a long CE run).



**Figure 8. Overview of the contig assembly for short run and long run sequencing files for sample No. 27 (*Acinetobacter baumannii*).** The long run (105 min) yielded almost full bidirectional coverage whereas the short run (30 min) yielded full 1x coverage and about 40% bidirectional coverage.

**Table 4. Summary of short CE run sequence data metrics, alignment results, and identification of the microbial DNA test panel representing 8 species from 5 bacterial phyla.**

DNA input of known organism (sample)	PCR primer set	Sequencing primer	Sequencing trace No.	SeqStudio run mode	Sequencing trace metrics				MicrobeBridge software	BLAST results				
					Sample score	CRL	QV20+	Signal RFU		Trim setting for contig assembly	Top ID	Identity (%)	No. of nt aligned	No. of gaps
1  <i>Bacteroides fragilis</i> ATCC 25285D-5	A	Forward	1	Short sequence	51	535	538	3,890	}	3 (default)	<i>Bacteroides fragilis</i>	99.87%	756	0
		Reverse	2	Short sequence	50	521	531	1,696						
	B	Forward	3	Short sequence	47	539	540	2,935	}	3 (default)	<i>Bacteroides fragilis</i>	99.56%	677	0
		Reverse	4	Short sequence	50	532	534	1,394						
	Sequencing traces No. 1 + 2 + 3 + 4 for contig assembly BLAST into NCBI microbial portal									4	<i>Bacteroides fragilis</i>	99.72%	1,451	0
2  <i>Fusobacterium nucleatum</i> ATCC 25586D-5	A	Forward	1	Short sequence	50	538	544	3,375	}	3 (default)	<i>Fusobacterium nucleatum</i>	99.73%	734	0
		Reverse	2	Short sequence	48	539	539	2,558						
	B	Forward	3	Short sequence	46	529	536	2,843	}	3 (default)	<i>Fusobacterium nucleatum</i>	100%	690	0
		Reverse	4	Short sequence	50	528	540	2,783						
	Sequencing traces No. 1 + 2 + 3 + 4 for contig assembly BLAST into NCBI microbial portal									3 (default)	<i>Fusobacterium nucleatum</i>	99.86%	1,442	0
3  <i>Acinetobacter baumannii</i> ATCC 17978D-5	A	Forward	1	Short sequence	50	530	541	2,535	}	3 (default)	<i>Acinetobacter baumannii</i>	100%	752	0
		Reverse	2	Short sequence	48	529	532	2,106						
	B	Forward	3	Short sequence	45	532	533	2,139	}	3 (default)	<i>Acinetobacter baumannii</i>	100%	689	0
		Reverse	4	Short sequence	49	525	538	2,476						
	Sequencing traces No. 1 + 2 + 3 + 4 for contig assembly BLAST into NCBI microbial portal									3 (default)	<i>Acinetobacter baumannii</i>	100%	1,459	0
4  <i>Actinomyces naeslundii</i> ATCC 12104D-5	A	Forward	1	Short sequence	50	526	530	2,729	}	3 (default)	<i>Actinomyces naeslundii</i>	100%	773	0
		Reverse	2	Short sequence	49	542	543	1,632						
	B	Forward	3	Short sequence	44	536	541	2,035	}	3 (default)	<i>Actinomyces naeslundii</i>	100%	692	0
		Reverse	4	Short sequence	50	517	536	1,948						
	Sequencing traces No. 1 + 2 + 3 + 4 for contig assembly BLAST into NCBI standard									3 (default)	<i>Actinomyces naeslundii</i>	100%	1,483	0
5  <i>Akkermansia muciniphila</i> ATCC BAA-835D-5	A	Forward	1	Short sequence	50	535	543	2,253	}	3 (default)	<i>Akkermansia muciniphila</i>	100%	725	0
		Reverse	2	Short sequence	47	522	531	1,561						
	B	Forward	3	Short sequence	42	526	527	1,390	}	3 (default)	<i>Akkermansia muciniphila</i>	99.86%	691	0
		Reverse	4	Short sequence	48	525	544	1,322						
	Sequencing traces No. 1 + 2 + 3 + 4 for contig assembly BLAST into NCBI microbial portal									3 (default)	<i>Akkermansia muciniphila</i>	99.79%	1,434	0
6  <i>Streptococcus agalactiae</i> ATCC BAA-1138D-5	A	Forward	1	Short sequence	50	539	536	4,107	}	3 (default)	<i>Streptococcus agalactiae</i>	99.74%	761	0
		Reverse	2	Short sequence	49	529	531	2,108						
	B	Forward	3	Short sequence	45	515	526	2,825	}	3 (default)	<i>Streptococcus agalactiae</i>	100%	689	0
		Reverse	4	Short sequence	49	528	540	3,346						
	Sequencing traces No. 1 + 2 + 3 + 4 for contig assembly BLAST into NCBI microbial portal									3 (default)	<i>Streptococcus agalactiae</i>	99.86%	1,471	0

## Conclusions

In this application note, we demonstrated the feasibility of sequencing the 16S rRNA gene on the SeqStudio Genetic Analyzer for the purpose of bacterial identification. The 16S Direct workflow that was presented here for its use on the SeqStudio instrument could also be applicable on other Applied Biosystems™ CE instruments such as the 3130 or 3500 series instruments with similar run modules. The 16S Direct workflow includes a novel and streamlined primer design that reduces the number of PCR amplicons to 2 and the number of sequencing reactions to 4 to obtain near full-length DNA sequence information for querying genomic databases such as the NCBI Microbial Genomes or CDC MicrobeNet. To that end, the BigDye Direct PCR and sequencing reagents generate high-resolution and high-quality sequence reads right from the beginning of the primer. A short CE sequencing run of 30 min loaded with 4 sequencing reactions is sufficient to yield approximately 2,000 high-quality bases (i.e., 4 x ~500 bases) for subsequent 16S gene contig assembly using the free MicrobeBridge software. This fast and economical workflow will benefit microbiology laboratories that use DNA sequencing as their ultimate bacterial identification and characterization tool. The affordable, compact, and versatile SeqStudio Genetic Analyzer makes it easy to introduce molecular genetics techniques in the general microbiology laboratory.

## Ordering information

Product	Cat. No.
SeqStudio Genetic Analyzer	A35644
SeqStudio Cartridge v1	A33671
SeqStudio Cartridge v2	A41331
BigDye Direct Cycle Sequencing Kit	4458687
BigDye Terminator v1.1 & v.3.1 5X Sequencing Buffer	4336697
BigDye XTerminator Purification Kit	4376486
ProFlex 3 x 32-well PCR System	4484073
E-Gel General Purpose Agarose Gels, 2%	G501802
TE Buffer	12090015
Nuclease-Free Water	AM9937

## References

1. Woese CR et al. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 87:4576-4579.
2. Clarridge JE 3<sup>rd</sup> (2004) Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev* 17:840-862.
3. Woo PC et al. (2008) Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clinical Microbiol Infect* 14: 908-934.
4. Petti CA et al. (2008) Interpretive criteria for identification of bacteria and fungi by DNA target sequencing; Approved Guideline. Clinical and Laboratory Standards Institute. CLSI Document MM18-A 28(12).
5. Weisburg WG et al. (1991) 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* 173: 697-703.

## Additional information

### MicroSEQ Microbial Identification system

Please note that the use of MicroSEQ ID products in combination with the SeqStudio Genetic Analyzer is for Research Use Only; not verified or supported for pharmaceutical QC purposes. Find out more at [thermofisher.com/microseq](http://thermofisher.com/microseq).

Find out more about the SeqStudio Genetic Analyzer at [fishersci.com/seqstudio](http://fishersci.com/seqstudio) or [fishersci.ca/seqstudio](http://fishersci.ca/seqstudio)

Contact us today:

#### In the United States

Order online: [fishersci.com](http://fishersci.com)

Fax an order: 1-800-926-1166

Call customer service: 1-800-766-7000

#### In Canada

Order online: [fishersci.ca](http://fishersci.ca)

Fax an order: 1-800-463-2996

Call customer service: 1-800-234-7437

