ion torrent
Sequencing for all.™

LONGER READ LENGTHS IMPROVE BACTERIAL IDENTIFICATION USING 16S rRNA GENE SEQUENCING ON THE ION PGM™ SYSTEM

16S rRNA sequencing is a fast, inexpensive profiling technique based on variation in the bacterial 16S ribosomal RNA (rRNA) gene. This method has a wide range of uses, including the characterization of bacteria populations, taxonomical analysis, and species identification. To support diverse projects such as the study of microbes present in foot ulcers and the bioremediation of arsenic-contaminated water, Dr. George Watts (Genomics Shared Service at the University of Arizona Cancer Center, Tucson, AZ) collaborated with Ion Torrent researchers to optimize the amplicon region targeted in the 16S gene (Figure 1) so he could obtain improved species-level discrimination of a set of 38 bacterial DNAs. The resulting comparison demonstrated that targeting two hypervariable (V1 and V2) regions in a single sequencing read with Ion Torrent's new 400-base chemistry improved the species-level discriminatory power of 16S rRNA profiling.

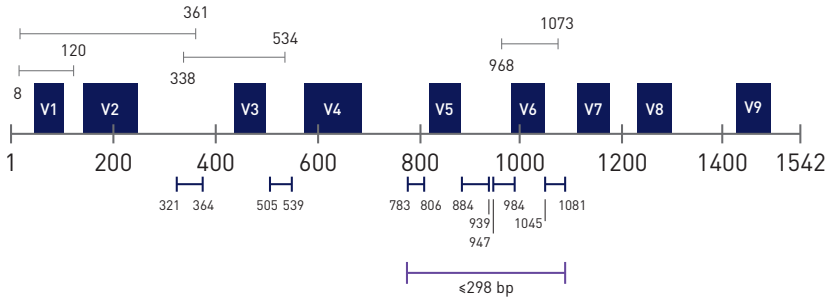


Figure 1. *E. coli* 16S rRNA gene with variable regions illustrated (blue squares). Top panel, amplicons (thin grey bars) targeted in this study. Bottom panel, the thick blue bars indicate conserved regions for primer design due to high coverage for the intervening amplicons for almost all taxa present in the Ribosomal Database Project (RDP) database[24]. The purple bar illustrates an additional hypothetical region that could be used to generate amplicons for 400-base pair–sequencing that would target two variable regions.

Introduction

Understanding the diversity of microbial communities, in almost any conceivable environment, has been greatly facilitated by rapid advances in next-generation sequencing (NGS) technologies and associated bioinformatics approaches. For example, the Ion PGM™ System has been used to characterize the bacterial communities important for human health in diverse sites including the mouth[1], airway[2, 3], diabetic mycetoma (or "madura foot")[4], and the human and murine intestinal tract[5, 6, 7]. Semiconductor sequencing has also revealed the composition of bacterial populations in a wide variety of environmental sources, such as waterways in response to oil sands mining and sewage[8, 9, 10], northern Gulf of Mexico[10] and English Channel[11, 12], soil and rhizosphere of a cactus and willow[13, 14], uranium mine tailings[15, 16], contaminated arctic soils[17], aquaculture systems[18], biowaste reservoirs[19], bioproduction settings[20, 21, 22], and during the life cycle of the lone star tick[23].

Significant to assessing microbial diversity is the bacterial 16S rRNA gene. Sequencing the 16S rRNA gene is used to identify bacterial species and perform grouping by shared sequence characteristics (taxonomical assignment). The 16S rRNA gene is highly conserved across domains of Bacteria and Archaea, and taxonomical assignment is possible due to the presence of nine hypervariable regions (V1-V9) that contain sufficient sequence diversity to classify microbes (Figure 1). Moreover, since these variable regions are flanked by conserved regions, PCR amplification using universal primers is possible.

Universal primers targeting the 16S rRNA gene allow a single, or a few, PCR reactions to be used when amplifying bacterial communities as a whole. Due to the clonal nature of massively parallel NGS workflows, each PCR reaction and subsequent sequencing read can be considered to be representative of a single bacterium within a mixed population. This approach has allowed characterization of bacterial communities without isolation and culturing. Through a deep sequencing approach, 16S rRNA profiling has fundamentally changed our understanding of countless microbial communities and has proven to be an important discovery tool, revealing what is called the "microbial dark matter" of our planet due to the difficulty of culturing most bacterial species.

life technologies™

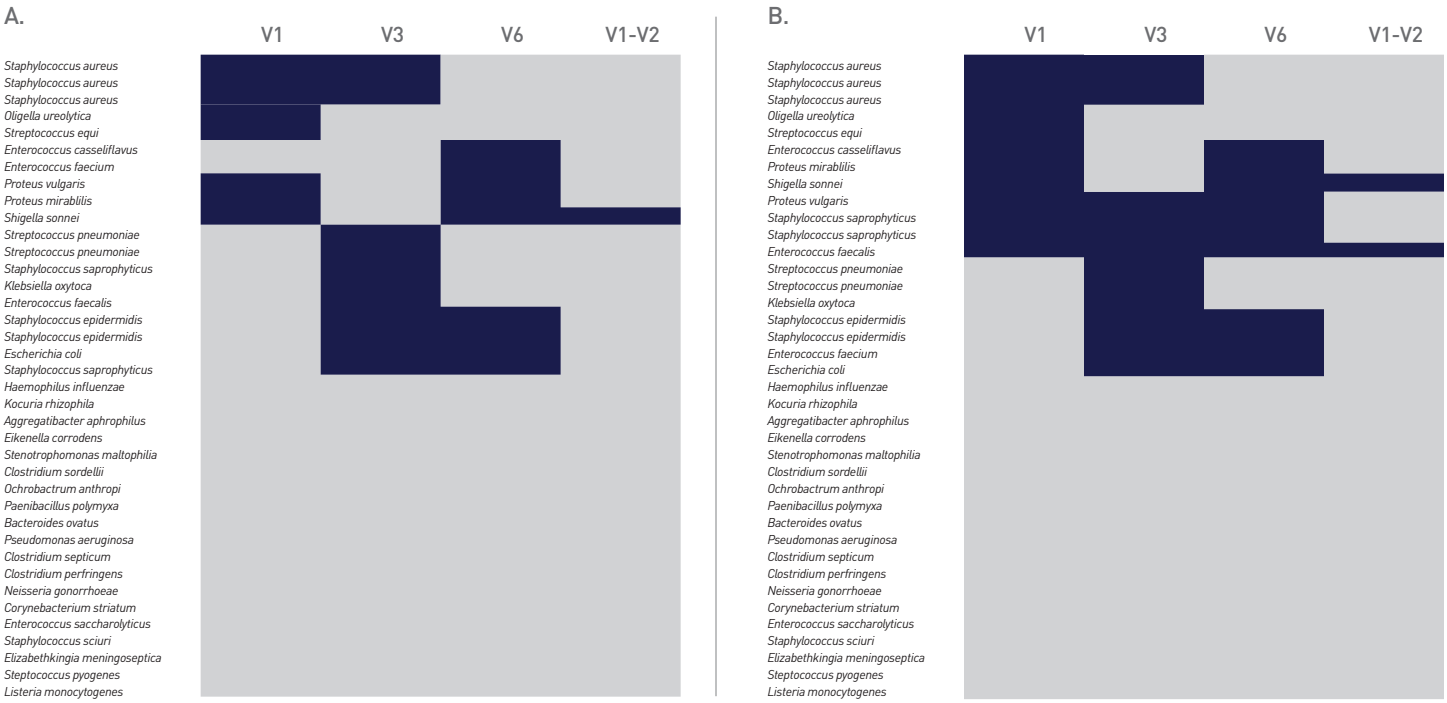


Figure 2. Samples that were correctly identified by amplicons targeting variable regions V1, V3, V6, and V1-V2. (A) Positive genus-level unique identification is indicated in grey for a particular sample with blue indicating ambiguous identification. **(B)** Positive species-level unique identification is indicated in grey for a particular sample with blue indicating ambiguous identification.

However, no single hypervariable region is sufficiently diverse to differentiate among all bacterial species. This fact, coupled with read length restrictions on most NGS platforms, require concessions to be made in terms of the discriminatory power of 16S rRNA profiling and the hypervariable regions investigated when planning an NGS project. In this study we demonstrate that, with the new 400-base read lengths possible on the Ion PGM™ System, targeting of two hypervariable regions (V1 and V2) in a single read increases the discriminatory power of 16S rRNA profiling.

Sequencing analysis of the 16S rRNA gene regions

A biologically relevant set of 38 bacterial DNAs, comprising 34 unique species, were used to determine if the discriminatory power of 16S rRNA profiling could be improved using 400-base sequencing of amplicons targeting multiple variable regions. To provide a baseline using the 200-base chemistry, the Genomics Shared Service at the Arizona Cancer Center amplified single variable regions V1 (175 bp insert), V3 (259 bp insert), and V6 (168 bp insert) from gDNA obtained from the 38 isolates (for primer and strain information see Supplementary information, Table 1 and Table 2). The PCR amplification products were used to create a library using the Ion Plus Fragment Library Kit (Cat. No. 4471252) with sample indexing using the Ion Xpress™ Barcode Adapters 1-96 Kit (Cat. No. 4474517). Template preparation was performed using the Ion OneTouch™ DL System and the Ion PGM™ Template OT2 200 Kit (Cat. No. 4480974). Sequencing was performed using the Ion PGM™ Sequencing 200 Kit (Cat. No. 4474004) on the Ion PGM™ System using three Ion 314™ Chips (Cat. No. 4482261) with ≤14 barcoded samples per chip. Primary data analysis was performed with Torrent Suite™ Software v3.4.

Variable regions V1-V2 (416 bp insert) of the 16S rRNA gene were amplified from gDNA obtained from the 38 isolates (Supplementary information, Table 1 and Table 2). PCR amplification products were used to create a library using the Ion Plus Fragment Library Kit (Cat. No. 4471252) with sample indexing using the Ion Xpress™ Barcode Adapter 1-96 Kit (Cat. No. 4474517). Template preparation was performed using the Ion OneTouch™ 2 System and the Ion PGM™ Template OT2 400 Kit (Cat. No. 4479878). Sequencing was performed using the Ion PGM™ Sequencing 400 Kit (Cat. No. 4482002) on the Ion PGM™ System using the Ion 318™ Chip v2 (Cat. No. 4484355) with all 38 barcoded samples on a single chip. Primary data analysis was performed with Torrent Suite™ Software v3.4.

Data analysis of the 16S rRNA gene regions

Reference sequences for the 38 bacterial species were parsed using custom scripts to isolate 16S rRNA gene regions targeted by the primers for V1, V3, V6, and V1-V2 regions (Supplementary information, Table 1). Using the TMAP alignment tool, the file output from the sequencing runs were aligned to the parsed reference sequences. For each barcoded sample, the following was determined: a) total throughput; b) proportion of reads aligning to specific reference accession. The most significant hit from the aligned reads for each barcode was evaluated to determine if the majority of the sequence reads correctly matched the known bacterial species for a particular barcode.

Results

Out of 38 samples, 19 (50%) were correctly identified at the species level by amplicons targeting any of the variable regions investigated (V1, V3, V6, and V1-V2) (Figure 2B). Sequencing the amplicon targeting the V3 region, 24 (63%) samples were correctly identified at the species level. The amplicon targeting the V6 region demonstrated similar discrimination with 24 distinct species and 27 samples (71%) correctly identified at the species level.

Likewise, the amplicon targeting V1 correctly identified 26 (68%) samples (24 separate species) at the species level. Amplicons V1 and V1-V2 use the same forward primer but differ in length due to the reverse primers targeting a single variable region or two variable regions, respectively (Figure 1 and Supplementary Information, Table 1). The additional variable region

targeting V1-V2 was sequenced using 400-base chemistry and improved the discrimination of 11 additional samples (8 distinct species) at the species level when compared to the single V1 region sequenced with 200-base chemistry (Figure 2B). The amplicon targeting the V1-V2 region demonstrated the highest discrimination with 36 (95%) of samples correctly identified at the species level.

A single species *(Enterococcus faecalis)* was not correctly identified at the species level by the majority of sequencing reads for any amplicon or variable region (Figure 2B). However, *E. faecalis* was uniquely identified at the genus level by the majority of sequencing reads using the longer amplicon targeting variable regions V1 and V2 (Figure 2A).

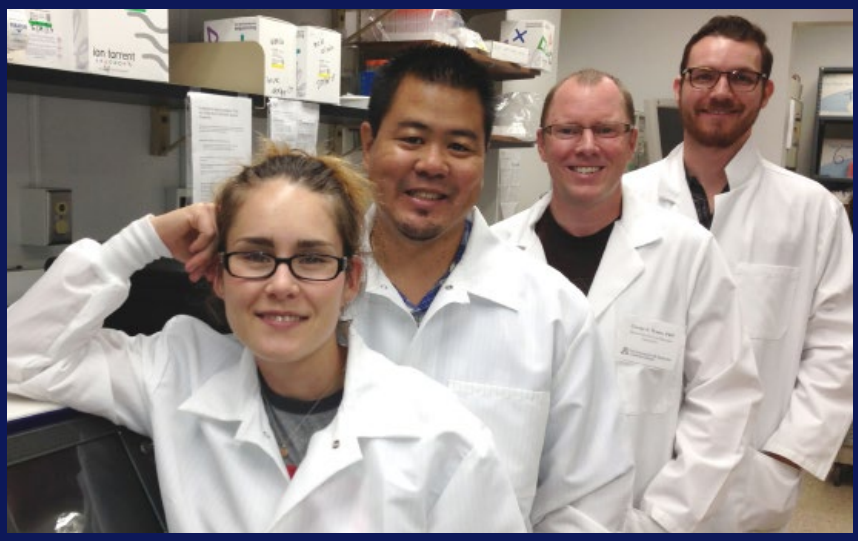
The remaining sample not correctly identified at the species nor genus level using the amplicon targeting V1-V2 was *Shigella sonnei* (Figure 2). For this sample, the correct identification at both the genus and species level occurred using the amplicon targeting the V3 region. When compared to V1-V2, the improved discrimination occurred using the amplicon targeting the V3 region was unique to this sample.

Conclusions

Ion Torrent technology delivers two optimized sequencing solutions for Ion PGM™ System: the Ion PGM™ Sequencing 200 Kit and the Ion PGM™ Sequencing 400 Kit. The Ion PGM™ Sequencing 400 Kit is ideal for *de novo* sequencing assembly and some targeted sequencing research applications. These applications include HLA gene sequencing and 16S rRNA profiling that benefit from longer reads since the key target regions are long contiguous stretches that span large exons or contain multiple adjacent hypervariable regions, respectively. The improved read lengths possible with 400-base chemistry will help the Genomics Shared Service at the University of Arizona Cancer Center investigate microbial populations important for human health and the environment.

Acknowledgements

Ion Torrent would like to acknowledge Dr. Donna Wolk, Director of Microbiology, Geisinger Medical Laboratories.



Genomics Shared Service at the Arizona Cancer Center

Genomics Shared Service at the Arizona Cancer Center is directed by George Watts, PhD and Bernard Futscher, PhD. In this picture are members of Genomics Shared Service core facility. Left to right: Candice J. Clark; Marc M. Oshiro, PhD; George S. Watts, PhD; Greg Metzger.

Genomics Shared Service core facility provides genomics services to funding members composed of the Arizona Cancer Center, the Southwest Environmental Health Sciences Center, and the BIO5 Institute, as well as to other federally funded investigators at the University of Arizona and other research investigators.

The facility provides complete support for sequencing and gene expression profiling as well as support for other genomics research applications, including: comparative genomic hybridization (CGH), chromatin immunoprecipitation on microarrays (ChIP-on-chip), single nucleotide polymorphism (SNP) analysis, and resequencing. Additional services include sample quality control assessment, real-time PCR, consultation in experimental design, and data archiving.

Key to the sequencing services offered by the Genomics Shared Service is the Ion PGM™ System. The core facility offers the Ion AmpliSeq™ Cancer Hotspot Panel v2 for targeted sequencing of 50 oncogenes and tumor suppressor genes. In addition, customers of the core facility can order custom panels for variant detection. Ion AmpliSeq™ technology, a massively multiplexed PCR amplicon methodology, can be customized using the free assay design tool, Ion AmpliSeq™ Designer, for the creation of primer pools to target any region of interest within the human genome.

The Genomics Shared Service is providing support to several researchers in their studies to identify bacteria using 16S rRNA sequencing. Initially, when read lengths were <400 bp, these studies targeted the V6 region of the 16S rRNA gene. With longer read lengths allowing the targeting of V1 and V2 regions, the Genomics Shared Service is expanding project support to include the analysis of the microbes present in retrospective samples from infected and healing foot ulcers. Another project involves defining the composition of bacterial communities that can bioremediate arsenic in drinking water. In the future, Genomics Shared Service hopes to use 16S rRNA sequencing to improve bacterial identification in cancer patients, especially those bacterial species for which culture-based methods perform poorly.

References

1. Junemann S, Prior K, Szczepanowski R et al. (2012) Bacterial community shift in treated periodontitis patients revealed by Ion Torrent 16S rRNA gene amplicon sequencing. *PLoS ONE* 7:e41606.

2. Blainey PC, Milla CE, Cornfield DN, Quake SR (2012) Quantitative analysis of the human airway microbial ecology reveals a pervasive signature for cystic fibrosis. *Sci Transl Med* 4:153ra130.

3. Salipante SJ, Sengupta DJ, Rosenthal C et al. (2013) Rapid 16S rRNA next-generation sequencing of polymicrobial clinical samples for diagnosis of complex bacterial infections. *PLoS ONE* 8:e65226.

4. Salipante SJ, Sengupta DJ, Hoogestraat DR, et al. (2013) Molecular diagnosis of *Actinomadura madurae* infection by 16S rRNA deep sequencing. *J Clin Microbiol* In Press, Accepted Manuscript.

5. Petrof EO, Gloor GB, Vanner SJ et al. (2013) Stool substitute transplant therapy for the eradication of *Clostridium difficile* infection: 'RePOOPulating' the gut. *Microbiome* 1:3.

6. Milani C, Hevia A, Foroni E et al. (2013) Assessing the fecal microbiota: an optimized Ion Torrent 16S rRNA gene-based analysis protocol. *PLoS ONE* 8:e68739.

7. Mir SA, Nagy-Szakal D, Dowd SE et al. (2013) Prenatal methyl-donor supplementation augments colitis in young adult mice. *PLoS ONE* 8:e73162.

8. Yergeau E, Lawrence JR, Sanschagrin S et al. (2012) Next-generation sequencing of microbial communities in the Athabasca River and its tributaries in relation to oil sands mining activities. *Appl Environ Microbiol* 78:7626–7637.

9. Yergeau E, Lawrence JR, Sanschagrin S et al. (2013). Aerobic biofilms grown from Athabasca watershed sediments are inhibited by increasing bituminous compounds concentrations. *Appl Environ Microbiol* In Press, Accepted Manuscript.

10. Kapoor V, Smith C, Santo Domingo JW et al. (2013) Correlative assessment of fecal indicators using human mitochondrial DNA as a direct marker. *Environ Sci Technol* 47:10485–10493.

11. Ortmann A, Ortell N (2013) Changes in free-living bacterial community diversity reflect the magnitude of environmental variability. *FEMS Microbiol Ecol* In Press, Accepted Manuscript.

12. Taylor JD, Cottingham SD, Billinge J, Cunliffe M (2013) Seasonal microbial community dynamics correlate with phytoplankton-derived polysaccharides in surface coastal waters. *ISME J* In Press, Accepted Manuscript.

13. Nessner Kavamura V, Taketani RG, Lançoni MD et al. (2013) Water regime influences bulk soil and rhizosphere of *Cereus jamacaru* bacterial communities in the Brazilian Caatinga biome. *PLoS ONE* 8: e73606.

14. Yergeau E, Sanschagrin S, Maynard C et al. (2013). Microbial expression profiles in the rhizosphere of willows depend on soil contamination. *ISME J* In Press, Accepted Manuscript.

15. Bondici VF, Lawrence JR, Khan NH et al. (2013) Microbial communities in low-permeability, high pH uranium mine tailings: characterization and potential effects. *J Appl Microbiol* 114:1671–1686.

16. Khan N, Bondici V, Medihala P et al. (2013) Bacterial diversity and composition of an alkaline uranium mine tailings-water interface. *J Microbiol* In Press, Accepted Manuscript.

17. Bell TH, Yergeau E, Maynard C et al. (2013) Predictable bacterial composition and hydrocarbon degradation in Arctic soils following diesel and nutrient disturbance. *ISME J* 7:1200–1210.

18. Auffret M, Yergeau E, Pilote A et al. (2012) Impact of water quality on the bacterial populations and off-flavours in recirculating aquaculture systems. *FEMS Microbiol Ecol* 84:235–247.

19. Whiteley AS, Jenkins S, Waite I et al. (2012) Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. *J Microbiol Methods* 91:80–88.

20. Visi D, D'Souza N, Ayre B et al. (2013) Investigation of the bacterial retting community of kenaf (*Hibiscus cannabinus*) under different conditions using next-generation semiconductor sequencing. *J Ind Microbiol Biotechnol* 40:465–475.

21. Wang W, Xie L, Luo G et al. (2013) Performance and microbial community analysis of the anaerobic reactor with coke oven gas biomethanation and *in situ* biogas upgrading. *Bioresour Technol* 146:234–239.

22. Luo G, Wang W, Angelidaki I (2013) Anaerobic digestion for simultaneous sewage sludge treatment and CO biomethanation: process performance and microbial ecology. *Environ Sci Technol* 47:10685–10693.

23. Menchaca AC, Visi DK, Strey OF et al. (2013) Preliminary assessment of microbiome changes following blood-feeding and survivorship in the *Amblyomma americanum* nymph-to-adult transition using semiconductor sequencing. *PLoS ONE* 8:e67129.

24. Wang Y, Qian PY (2009) Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS ONE* 4:e7401.

ion torrent
Sequencing for all.™

Table 1. Primer sequences, *E. coli* 16S rRNA gene numbering, and references.

Region (*E. coli* 16S numbering)	Forward primer (*E. coli* 16S numbering)	Forward primer sequence	Reverse primer (*E. coli* 16S numbering)	Reverse primer sequence	Reference	Notes
V1 (8–120)	8F (8–27)	AGAGTTTGATCMTGGCTCAG	120R (101–120)	TTACTCACCCGTICGCCRCT	[1]	F8 is used for V1, V1-V2
V3 (338–534)	F338 (338–357)	ACTCCTACGGGAGGCAGCAG	R534 (518–534)	ATTACCGCGGCTGCTGG	[1]	R534 is used for V3
V6 (968–1073)	F968 (968–985)	ACGCGARGAACCTTACC	R1073 (1054–1073)	ACGAGCTGACGACARCCATG	[2]	
V1-V2 (8–361)	8F (8–27)	AGAGTTTGATCMTGGCTCAG	R361 (342–361)	CYIACTGCTGCCTCCCGTAG	[1]	R361 is reverse complement of F338

Table 2. Bacterial species and strains assessed.

Genus	Species	Strain	Accession
Aggregatibacter	*aphrophilus*	ATCC 33389	gi|174771|gb|M75041.1|HEARRNAD
Bacteroides	*ovatus*	JCM 5824	gi|284925050|dbj|AB510705.1|
Clostridium	*perfringens*	ATCC 13124	gi|110798562:10167-11688
Clostridium	*septicum*	Pasteur III	gi|219846429|ref|NR_026020.1|
Clostridium	*sordellii*	ATCC 9714	gi|17529654|dbj|AB075771.1|
Corynebacterium	*striatum*	Minnett	gi|310975177|ref|NR_037041.1|
Eikenella	*corrodens*	JCM 12952	gi|261228524|dbj|AB525415.1|
Elizabethkingia	*meningoseptica*	ATCC 13253	gi|343201541|ref|NR_042267.1|
Enterococcus	*casseliflavus*	9199	gi|2828136|gb|AF039903.1|
Enterococcus	*faecalis*	NRIC 0112	gi|157907333|dbj|AB362601.1|
Enterococcus	*faecium*	DO	gi|389867183:463164-464711
Enterococcus	*saccharolyticus*	ATCC 43076	gi|89357462|gb|DQ411816.1|
Escherichia	*coli*	ATCC 25922	gi|86278349|gb|DQ360844.1|
Haemophilus	*influenzae*	PittEE	gi|148825133:376402-377942
Klebsiella	*oxytoca*	KCTC 1686	gi|375256816:3821-5374
Kocuria	*rhizophila*	DC2201	gi|184199646:1752303-1753835
Listeria	*monocytogenes*	Scott A	gi|265932|gb|S55472.1|
Neisseria	*gonorrhoeae*	TCDC-NG08107	gi|385334895:1177616-1179168
Ochrobactrum	*anthropi*	SMrs31	gi|407080005|gb|JX485750.1|
Oligella	*ureolytica*	ATCC 43534	gi|7407118|gb|AF227164.1|
Paenibacillus	*polymyxa*	M1	gi|343094728:11940-13481
Proteus	*mirabilis*	HI4320	gi|172046403:452966-454507
Proteus	*vulgaris*	ATCC 6380	gi|175698|gb|J01874.1|PRMRRD

life technologies™

Table 2. Bacterial species and strains assessed. (Cont.)

Genus	Species	Strain	Accession
Pseudomonas	*aeruginosa*	ATCC 27853	gi|10567490|gb|AF094719.1|
Shigella	*sonnei*	ATCC 25931	gi|1255976|emb|X96964.1|
Staphylococcus	*aureus*	ATCC 43300	gi|211920447|emb|AM980864.1|
Staphylococcus	*aureus*	MRSA252	gi|49240382:514251-515805
Staphylococcus	*aureus*	MSSA476	gi|49243355:490196-491750
Staphylococcus	*epidermidis*	ATCC 12228	gb|AE015929.1|:1598006-1599559
Staphylococcus	*epidermidis*	RP62A	gb|CP000029.1|:105734-107287
Staphylococcus	*saprophyticus*	ATCC 15305	gi|576602|gb|L37596.1|STARGDA
Staphylococcus	*saprophyticus*	MAFF 911473	gi|197304660|dbj|D83371.2|STA16SRR19
Staphylococcus	*sciuri*	B723-3	gi|83764381|dbj|AB212276.1|
Stenotrophomonas	*maltophilia*	D457	gi|386716467:395406-396943
Streptococcus	*equi*	zooepidemicus MGCS10565	gi|195973861|gb|CP001129.1|:17608-20007
Streptococcus	*pneumoniae*	ATCC 49619	gi|32396623|gb|AY281082.1|
Streptococcus	*pneumoniae*	clone 4V4	gi|82617083|emb|AM157442.1|
Streptococcus	*pyogenes*	A20	gi|409913960:17062-18562

References

1. Sundquist A, Bigdeli S, Jalili R et al. (2007) Bacterial flora-typing with targeted, chip-based Pyrosequencing. *BMC Microbiol* 7:108.
2. Wang Y, Qian PY (2009) Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS ONE* 4:e7401.