

Supplementary Materials: TARGETING THE 16S rRNA GENE FOR BACTERIAL IDENTIFICATION IN COMPLEX MIXED SAMPLES: COMPARATIVE EVALUATION OF SECOND (ILLUMINA) AND THIRD (OXFORD NANOPORE TECHNOLOGIES) GENERATION SEQUENCING TECHNOLOGIES

Raf Winand¹, Bert Bogaerts¹, Stefan Hoffman¹, Loïc Lefevre¹, Maud Delvoye¹, Julien Van Braekel¹, Qiang Fu¹, Nancy HC Roosens¹, Sigrid CJ De Keersmaecker^{1,†} and Kevin Vanneste^{1,†}

1. Description of Illumina MiSeq Raw and Processed Read Statistics for All Sequenced 16S rRNA Gene Regions

The total number of raw reads and raw contigs that were generated can be found in Table S1, as well as the number of preprocessed and unique contigs when using either the SILVA and NCBI 16S databases. Nearly no reads were generated for samples 16S9, 16S10, and 16S11, whereas sample 16S6 generated 573,391 raw reads and 550,391 contigs, with other samples varying between these two extremes. Of particular interest are the numbers of unique contigs obtained when using either database, which also varied widely between samples. For instance, sample 16S7 generated 272,475 preprocessed and 31,662 unique contigs using the SILVA database, and 245,850 preprocessed and 29,179 unique contigs using the NCBI 16S database. Closer inspection revealed that the number of ambiguity characters was very high, influencing the alignment and subsequent filtering steps. In contrast, sample 16S6 generated 509,679 preprocessed and 3215 unique contigs using the SILVA database, and 525,024 preprocessed and 3165 unique contigs using the NCBI 16S database, suggesting that not all diversity of the sample was captured by the reads, because the ratio between preprocessed and unique contigs is very large compared to the other samples. The number of reads classified per sample at each taxonomic level is presented in Table S2. The number of reads classified at the family level corresponds roughly with the number of pre-processed contigs for each sample indicating that the majority of pre-processed reads could be classified at the family level. The only exception is the aforementioned sample 16S6, where only 350,136 out of 509,679 reads, and 303,720 out of 525,024 reads, could be identified at the family level using the SILVA and NCBI 16S databases, respectively, which can be explained by the relatively low number of unique reads in this sample. Sample 16S6 was, however, a clear outlier with 30.89% and 41.83% of unclassified reads using the SILVA and NCBI 16S databases, respectively, which can most likely be explained by the relatively low number of unique reads compared to total reads resulting in large fluctuations when even a few unique reads are misclassified.

2. Description of Bacterial Species Detected in the Mock Community for All Sequenced 16S rRNA Gene Regions for the Illumina MiSeq Data

Below, a concise summary is provided for every bacterial species present in the mock community. Note that species identification was always only possible using the NCBI 16S database. For *B. subtilis*, the family *Bacillaceae* was identified in all samples using both the SILVA and NCBI 16S databases. The genus *Bacillus* was identified in all samples using the NCBI and SILVA databases, except for sample 16S1 using the SILVA database. There were no samples in which *B. subtilis* was identified at the species level. For *E. coli*, the family *Enterobacteriaceae* was identified in all samples using both the SILVA and NCBI 16S databases. The genus *Escherichia-Shigella* was identified in all samples except 16S9 using the NCBI 16S database, and in all samples except 16S9 and 16S6 using the SILVA database. Interestingly enough, the species *E. coli* was only identified in sample 16S6 despite its lower sequence quality. For *Enterococcus faecalis*, the family *Enterococcaceae* was identified in all samples using both the NCBI and SILVA databases, except for 16S6 using the SILVA database. The genus *Enterococcus* was identified in all

samples using the NCBI 16S database but not in samples 16S1, 16S4, and 16S6 using the SILVA database. The species *E. faecalis* was identified in samples 16S2, 16S3, 16S7, 16S8, and 16S9. For *Salmonella enterica*, the family *Enterobacteriaceae* was identified in all samples using both the SILVA and NCBI 16S databases, but classification at lower taxonomic levels was more problematic. The genus *Salmonella* was identified in only six samples (16S2, 16S3, 16S4, 16S5, 16S7, and 16S8) using the NCBI 16S database, and two samples (16S5 and 16S7) using the SILVA database. The species *S. enterica* was identified in the same six aforementioned samples. For *Listeria monocytogenes*, both the family *Listeriaceae* and the genus *Listeria* were identified in all samples using the SILVA and NCBI 16S databases. The species *L. monocytogenes* was, however, only identified in sample 16S9. For *Pseudomonas aeruginosa*, the family *Pseudomonadaceae* was identified in all samples using the SILVA and NCBI 16S databases. The genus *Pseudomonas* was identified in all samples using the NCBI 16S database, but only in four samples (16S2, 16S3, 16S5, and 16S7) using the SILVA database. The species *P. aeruginosa* was identified in all samples except 16S6 and 16S9. For *Lactobacillus fermentum*, both the family *Lactobacillaceae* and the genus *Lactobacillus* were identified in all samples using the SILVA and NCBI 16S databases. The species *L. fermentum* was identified in all samples except 16S9. For *S. aureus*, both the family *Staphylococcaceae* and the genus *Staphylococcus* were identified in all samples except 16S9 using the SILVA and NCBI 16S databases. The species *S. aureus* was similarly detected in all samples except 16S9.

3. Description of Bacterial Species Detected in the Non-Template Controls for All Sequenced 16S rRNA Gene Regions for the Illumina MiSeq Data

All gene regions were accompanied by a NTC that was subjected to the same laboratory and bioinformatics workflow as the actual samples. The numbers of reads surviving preprocessing and resulting in classification were generally low in the NTC samples. We found that none of the false positives identified in the actual samples were present in the NTCs. Only bacteria known to be contaminants of the Illumina MiSeq sequencing workflow were detected in the NTCs, and were typically only supported by limited read numbers. At the family level, few reads were classified with sample 16S9, an outlier that had up to 2917 reads classified. At the genus level, for the organisms present in the mock community, only *Pseudomonas* and *Staphylococcus* were identified in between one (16S7) and three (16S1, 16S3, and 16S9) NTC samples, respectively, using both the SILVA and NCBI 16S databases. *Bacillus* was identified in two samples (16S4 and 16S9) only when using the NCBI 16S database. For organisms not present in the mock community, the large majority were only sporadically identified in one or two samples, with the exception of *Delftia* being identified in three samples (16S1, 16S3, and 16S4) using both the SILVA and NCBI 16S databases, and *Bradyrhizobium* and *Sphingomonas* being identified in five samples (16S1, 16S3, 16S4, 16S7, and 16S9) using the NCBI 16S database. Some of the other genera identified only sporadically in the NTC samples included *Actinomyces*, *Corynebacterium*, *Devosia*, *Enhydrobacter*, *Mesorhizobium*, *Methylobacterium*, *Micrococcus*, *Stenotrophomonas*, and *Streptococcus*. All these genera are, however, known contaminants of the Illumina sequencing process, as described previously [1–3]. A complete overview of all reads that were classified in the NTC samples can be found in Supplementary Tables 4–8. These contaminants were also found back in the actual samples, but were filtered out by the bioinformatics workflow during the preprocessing steps due to their extremely low abundance, and did not interfere with identification. Such contaminants could nevertheless present issues because some of them belong to genera such as *Pseudomonas* and *Staphylococcus*, which also contain pathogenic bacteria of interest present in the mock community (i.e., *P. aeruginosa* and *S. aureus*). Because the number of detected reads for the former bacteria in the actual samples was always very high, they could be excluded as contaminants. However, if either *P. aeruginosa* or *S. aureus* would be present in actual samples at low concentrations, this could be problematic through difficulties in differentiating between the actual presence of a highly pathogenic bacteria and a known contaminant.

4. Supplementary tables

Table S1. Overview of read and contigs statistics for MiSeq data. The first, second, third, and fourth columns list the sample name, number of raw reads for the sample, the exact community sequenced, and the number of contigs, respectively. The fifth and sixth columns list the number of preprocessed and unique contigs when the SILVA database was used. The seventh and eighth columns list the numbers of preprocessed and unique contigs when the NCBI database was used.

Sample	Raw reads	Community	Contigs	SILVA		NCBI	
				Preprocessed contigs	Unique contigs	Preprocessed contigs	Unique contigs
16S1	550,363	ZymoBIOMICS	442,223	388,451	7,499	385,571	7,287
16S1	177,859	NTC	882	351	96	352	95
16S2	374,064	ZymoBIOMICS	370,429	309,038	13,446	309,281	13,227
16S2	186	NTC	20	4	4	8	5
16S3	485,522	ZymoBIOMICS	480,933	354,082	19,314	354,730	18,887
16S3	364	NTC	152	120	60	120	61
16S4	344,096	ZymoBIOMICS	336,363	194,088	7,980	207,346	8,171
16S4	779	NTC	210	174	55	174	55
16S5	390,381	ZymoBIOMICS	377,697	115,380	8,184	115,441	8,058
16S5	8,172	NTC	63	8	3	7	3
16S6	573,391	ZymoBIOMICS	550,931	509,679	3,215	525,024	3,165
16S6	842	NTC	43	10	5	10	5
16S7	461,396	ZymoBIOMICS	455,347	272,475	31,662	245,850	29,179
16S7	13,351	NTC	131	63	22	58	20
16S9	562,425	ZymoBIOMICS	557,407	443,250	26,506	443,732	25,814
16S9	193	NTC	30	5	4	7	6
16S10	314	ZymoBIOMICS	259	106	13	52	8
16S10	4,285	NTC	3,399	3,035	60	3,008	60
16S11	235	ZymoBIOMICS	158	0	0	0	0
16S11	20,841	NTC	866	3	3	8	3
16S12	278	ZymoBIOMICS	210	30	17	19	13
16S12	2,441	NTC	2,234	4	4	9	9

NTC: Non-template control

Table S2. Overview of number of reads classified in each sample at different taxonomic levels for the MiSeq data. The first column lists the sample name. The second, third, and fourth columns list the number of reads identified down to the family, genus, and species level, respectively, when using the NCBI 16S database. The fifth and sixth columns list the numbers of reads identified down to the family, and genus level, respectively, when using the SILVA database.

Sample	NCBI			SILVA	
	Family	Genus	Species	Family	Genus
16S1	383,986	311,248	190,338	386,270	186,732
16S2	308,254	308,038	249,264	307,920	271,592
16S3	352,002	350,652	290,912	350,938	306,269
16S4	206,029	186,816	135,323	192,676	137,890
16S5	114,640	106,755	71,825	114,479	114,448
16S6	303,720	299,408	173,996	350,136	339,190
16S7	242,467	242,467	198,856	268,527	268,527
16S9	440,754	440,577	358,968	439,940	353,022
16S10	50	32	20	103	30
16S11	0	0	0	0	0
16S12	0	0	0	0	0

Table S3. Overview of number of reads classified in sample 16S9 of the MiSeq data at all taxonomic levels. The first column lists the taxonomic level, the second column the taxon name, and the third and fourth column, the number of identified reads at that level using the SILVA and NCBI 16S databases, respectively.

Taxonomic level	Taxon	SILVA	NCBI 16S
Kingdom	<i>Bacteria</i>	103	50
Phylum	<i>Firmicutes</i>	30	29
Class	<i>Bacilli</i>	30	29
Order	<i>Bacillales</i>	17	17
Family	<i>Bacillaceae</i>	12	12
Genus	<i>Bacillus</i>	12	12
Species	<i>Bacillus halotolerans</i>	-	12
Family	<i>Listeriaceae</i>	5	5
Genus	<i>Listeria</i>	5	5
Species	<i>Listeria monocytogenes</i>	-	5
Order	<i>Lactobacillales</i>	13	12
Family	<i>Enterococcaceae</i>	3	3
Genus	<i>Enterococcus</i>	3	3
Species	<i>Enterococcus faecalis</i>	-	3
Family	<i>Lactobacillaceae</i>	10	9
Genus	<i>Lactobacillus</i>	10	9
Genus	<i>Lactobacillus_unclassified</i>	-	9
Phylum	<i>Proteobacteria</i>	73	21
Class	<i>Betaproteobacteria</i>	52	-
Order	<i>Neisseriales</i>	52	-
Family	<i>Neisseriaceae</i>	52	-
Genus	<i>Neisseriaceae_unclassified</i>	52	-
Class	<i>Gammaproteobacteria</i>	21	21
Order	<i>Enterobacteriales</i>	18	18
Family	<i>Enterobacteriaceae</i>	18	18
Family	<i>Enterobacteriaceae_unclassified</i>	18	18
Order	<i>Pseudomonadales</i>	3	3
Family	<i>Pseudomonadaceae</i>	3	3
Family	<i>Pseudomonadaceae_unclassified</i>	3	-
Genus	<i>Pseudomonas</i>	-	3
Genus	<i>Pseudomonas_unclassified</i>	-	3

Table S4. False positive genera and the percentage of reads assigned to them in each gene region for the MiSeq data when using the NCBI 16S and SILVA database.

NCBI 16S

[illegible]

SILVA

[illegible]

Table S5. Overview of number of reads that were assigned to each species using the NCBI 16S database when using EPI2ME, Mothur, and GraphMap for classification for the MinION data. For each species present in the mock community (indicated in the first column), all other species belonging to the same genus that were assigned at least one read are presented (indicated in the second column). The theoretical abundance of 16S rRNA belonging to species of the mock community is indicated in the third column. The next columns list the number of reads identified at the taxonomic level of the family, genus, and species for the three different bioinformatics workflows. The entries "OTHER HITS n=1" and "OTHER HITS n>1" refer to all reads identified as species for which a genus was not present in the mock community, including and excluding single hits, respectively.

Species in sample	%sample	Identified species	EPI2ME			Mothur			GraphMap		
			Family	Genus	Species	Family	Genus	Species	Family	Genus	Species
<i>Pseudomonas aeruginosa</i>	4.20%	TOTAL	192	192		85	85		190	190	
		<i>Pseudomonas aeruginosa</i>			179			85			172
		<i>Pseudomonas alcaligenes</i>			1			0			2
		<i>Pseudomonas delhiensis</i>			0			0			1
		<i>Pseudomonas graminis</i>			1			0			0
		<i>Pseudomonas knackmussii</i>			1			0			0
		<i>Pseudomonas mendocina</i>			0			0			1
		<i>Pseudomonas nitroreducens</i>			1			0			0
		<i>Pseudomonas oleovorans</i>			0			0			2
		<i>Pseudomonas otitidis</i>			5			0			11
		<i>Pseudomonas pelagia</i>			0			0			1
		<i>Pseudomonas resinovorans</i>			4			0			0
<i>Escherichia coli</i>	10.10%	TOTAL	1,607	36		1,117	432		2,057	728	
		<i>Escherichia coli</i>			0			0			12
		<i>Escherichia albertii</i>			0			0			1
		<i>Escherichia fergusonii</i>			11			33			324
		<i>Escherichia marmotae</i>			5			88			320
		<i>Shigella boydii</i>			0			0			9
		<i>Shigella dysenteriae</i>			2			103			18
		<i>Shigella flexneri</i>			13			128			16
		<i>Shigella sonnei</i>			5			80			28
<i>Salmonella enterica</i>	10.40%	TOTAL	1,607	289		1,117	684		2,057	1,209	
		<i>Salmonella enterica</i>			288			684			1,168
		<i>Salmonella bongori</i>			1			0			41
<i>Lactobacillus fermentum</i>	18.40%	TOTAL	934	929		193	193		969	968	
		<i>Lactobacillus fermentum</i>			773			193			560
		<i>Lactobacillus alvi</i>			1			0			0
		<i>Lactobacillus collinoides</i>			1			0			0
		<i>Lactobacillus equigenerosi</i>			6			0			4
		<i>Lactobacillus gastricus</i>			85			0			1

Species in sample	%sample	Identified species	EPI2ME			Mothur			GraphMap		
			Family	Genus	Species	Family	Genus	Species	Family	Genus	Species
Enterococcus faecalis	9.90%	<i>Lactobacillus gorillae</i>			6			0			397
		<i>Lactobacillus ingluviei</i>			7			0			0
		<i>Lactobacillus mixtipabuli</i>			1			0			0
		<i>Lactobacillus mucosae</i>			15			0			3
		<i>Lactobacillus nagelii</i>			0			0			2
		<i>Lactobacillus reuteri</i>			31			0			0
		<i>Lactobacillus siliginis</i>			1			0			0
		<i>Lactobacillus vaccinostercus</i>			1			0			0
		<i>Lactobacillus wasatchensis</i>			1			0			1
		TOTAL	886	870		565	565		1,141	1,141	
		<i>Enterococcus faecalis</i>			781			564			1,120
		<i>Enterococcus asini</i>			1			0			0
		<i>Enterococcus avium</i>			8			0			0
		<i>Enterococcus bulliens</i>			1			0			0
		<i>Enterococcus canintestini</i>			1			0			0
		<i>Enterococcus canis</i>			5			0			0
		<i>Enterococcus dispar</i>			8			0			0
		<i>Enterococcus durans</i>			1			0			0
		<i>Enterococcus eurekensis</i>			0			0			1
		<i>Enterococcus faecium</i>			4			0			5
		<i>Enterococcus gallinarum</i>			4			0			0
		<i>Enterococcus haemoperoxidus</i>			2			0			2
		<i>Enterococcus hirae</i>			17			0			0
		<i>Enterococcus lemanii</i>			0			0			1
		<i>Enterococcus malodoratus</i>			1			0			0
		<i>Enterococcus moraviensis</i>			4			0			4
		<i>Enterococcus mundtii</i>			1			0			0
		<i>Enterococcus pseudoavium</i>			4			0			1
		<i>Enterococcus rivorum</i>			3			0			0
		<i>Enterococcus rotai</i>			0			1			0
		<i>Enterococcus saccharolyticus</i>			11			0			1
		<i>Enterococcus silesiacus</i>			1			0			0
		<i>Enterococcus sulfureus</i>			10			0			6
		<i>Enterococcus termitis</i>			2			0			0
<i>Staphylococcus aureus</i>	15.50%	TOTAL	1,547	1,536		863	863		1,555	1,554	
		<i>Staphylococcus aureus</i>			1,416			860			1,412
		<i>Staphylococcus capitis</i>			3			0			0
		<i>Staphylococcus caprae</i>			20			0			5

Species in sample	%sample	Identified species	EPI2ME			Mothur			GraphMap		
			Family	Genus	Species	Family	Genus	Species	Family	Genus	Species
<i>Listeria monocytogenes</i>	14.10%	<i>Staphylococcus carnosus</i>			0			0			2
		<i>Staphylococcus chromogenes</i>			1			0			0
		<i>Staphylococcus cohnii</i>			2			0			1
		<i>Staphylococcus delphini</i>			0			0			1
		<i>Staphylococcus devriesei</i>			3			0			2
		<i>Staphylococcus epidermidis</i>			1			0			0
		<i>Staphylococcus equorum</i>			0			0			2
		<i>Staphylococcus haemolyticus</i>			4			0			15
		<i>Staphylococcus hominis</i>			27			0			5
		<i>Staphylococcus kloosii</i>			1			0			0
		<i>Staphylococcus pasteurii</i>			1			0			1
		<i>Staphylococcus petrasii</i>			10			0			36
		<i>Staphylococcus pettenkoferi</i>			1			0			0
		<i>Staphylococcus piscifermentans</i>			2			0			0
		<i>Staphylococcus pseudintermedius</i>			0			0			3
		<i>Staphylococcus saccharolyticus</i>			34			0			36
		<i>Staphylococcus saprophyticus</i>			5			0			0
		<i>Staphylococcus sciuri</i>			2			0			0
		<i>Staphylococcus simiae</i>			2			3			31
		<i>Staphylococcus simulans</i>			0			0			1
		<i>Staphylococcus stepanovicii</i>			0			0			1
		<i>Staphylococcus xylosus</i>			1			0			0
		TOTAL	1,320	1,280		588	588		1,233	1,233	
		<i>Listeria monocytogenes</i>			26			0			13
		<i>Listeria aquatica</i>			0			0			1
		<i>Listeria grayi</i>			1			0			1
		<i>Listeria innocua</i>			205			1			771
		<i>Listeria ivanovii</i>			19			0			2
		<i>Listeria marthii</i>			3			0			3
		<i>Listeria riparia</i>			1			0			1
		<i>Listeria rocourtiae</i>			0			0			1
		<i>Listeria seeligeri</i>			97			100			149
		<i>Listeria weihenstephanensis</i>			2			0			7
		<i>Listeria welshimeri</i>			926			487			284
<i>Bacillus subtilis</i>	17.40%	TOTAL	2,187	2,133		1,018	1,018		632	626	
		<i>Bacillus subtilis</i>			973			207			282
		<i>Bacillus acidicola</i>			6			0			0
		<i>Bacillus altitudinis</i>			1			0			0

Species in sample	%sample	Identified species	EPI2ME			Mothur			GraphMap		
			Family	Genus	Species	Family	Genus	Species	Family	Genus	Species
		<i>Bacillus amyloliquefaciens</i>			88			0			2
		<i>Bacillus aquimaris</i>			4			0			0
		<i>Bacillus asahii</i>			1			0			0
		<i>Bacillus atrophaeus</i>			89			0			1
		<i>Bacillus australimaris</i>			2			0			0
		<i>Bacillus berkeleyi</i>			1			0			0
		<i>Bacillus carboniphilus</i>			1			0			0
		<i>Bacillus cereus</i>			1			0			0
		<i>Bacillus cytotoxicus</i>			1			0			0
		<i>Bacillus depressus</i>			1			0			0
		<i>Bacillus fengquiensis</i>			1			0			1
		<i>Bacillus firmus</i>			8			0			0
		<i>Bacillus flexus</i>			2			0			0
		<i>Bacillus galliciensis</i>			2			0			0
		<i>Bacillus ginsengihumi</i>			1			0			0
		<i>Bacillus gobiensis</i>			1			0			0
		<i>Bacillus gottheilii</i>			5			0			0
		<i>Bacillus halotolerans</i>			357			675			166
		<i>Bacillus hemicellulosilyticus</i>			1			0			0
		<i>Bacillus isabeliae</i>			1			0			0
		<i>Bacillus kribbensis</i>			1			0			0
		<i>Bacillus lentus</i>			1			0			0
		<i>Bacillus licheniformis</i>			14			0			1
		<i>Bacillus massiliosenegalensis</i>			1			0			0
		<i>Bacillus massiloanorexius</i>			1			0			0
		<i>Bacillus mediterraneensis</i>			0			0			1
		<i>Bacillus mesonae</i>			0			0			1
		<i>Bacillus mojavensis</i>			397			131			24
		<i>Bacillus nakamurai</i>			0			1			91
		<i>Bacillus nematocida</i>			18			1			8
		<i>Bacillus oleronius</i>			7			0			0
		<i>Bacillus pseudofirmus</i>			2			0			0
		<i>Bacillus seohaeanensis</i>			1			0			0
		<i>Bacillus shackletonii</i>			1			0			0
		<i>Bacillus siamensis</i>			11			0			3
		<i>Bacillus songklensis</i>			2			0			0
		<i>Bacillus sonorensis</i>			1			0			0
		<i>Bacillus sporothermodurans</i>			2			0			0

Species in sample	%sample	Identified species	EPI2ME			Mothur			GraphMap		
			Family	Genus	Species	Family	Genus	Species	Family	Genus	Species
		<i>Bacillus stratosphericus</i>			2			0			0
		<i>Bacillus subterraneus</i>			1			0			0
		<i>Bacillus tequilensis</i>			8			0			12
		<i>Bacillus testis</i>			1			0			0
		<i>Bacillus toyonensis</i>			1			0			0
		<i>Bacillus vallismortis</i>			108			3			19
		<i>Bacillus velezensis</i>			1			0			14
		<i>Bacillus vietnamensis</i>			1			0			0
		<i>Bacillus wakoensis</i>			1			0			0
		<i>Bacillus xiamenensis</i>			1			0			0
OTHER HITS n = 1	N/A	N/A	30	68	68	0	1	1	35	163	163
OTHER HITS n >1	N/A	N/A	28	62	50	0	0	0	20	133	121

Table S6. Overview of number of reads that were classified at the genus level using the NCBI and SILVA databases when using EPI2ME, Mothur, and GraphMap for the MinION data. Only genera that were not expected to be present are listed. The first column lists the genus names. The second, third, and fourth columns list the number of reads identified for each genus when using the NCBI 16S database and EPI2ME, Mothur, and GraphMap, respectively. The fifth and sixth columns list the number of reads identified for each genus when using the SILVA database and Mothur, and GraphMap, respectively.

Genus	NCBI			SILVA	
	EPI2ME	Mothur	GraphMap	Mothur	GraphMap
0319-6M6_ge	-	-	-	-	1
<i>Alkalibacterium</i>	1	-	-	-	-
<i>Anoxybacillus</i>	-	-	1	-	13
<i>Arsukibacterium</i>	-	-	1	-	-
<i>Atlantibacter</i>	-	-	1	-	-
<i>Brevibacillus</i>	-	-	-	-	3
<i>Burkholderia-Paraburkholderia</i>	-	-	-	-	4
<i>Buttiauxella</i>	-	-	3	-	-
<i>Caldicellulosiruptor</i>	-	-	-	-	3
<i>Cardiobacterium</i>	-	-	-	-	1
<i>Cedecea</i>	2	-	-	-	-
<i>Chitinophaga</i>	-	-	1	-	-
<i>Citrobacter</i>	18	-	21	-	81
<i>Colwellia</i>	-	-	-	-	1
<i>Cronobacter</i>	1	-	3	-	-
<i>Enterobacter</i>	18	1	49	-	455
<i>Erwinia</i>	-	-	2	-	-
<i>Frischella</i>	-	-	1	-	-
<i>Idiomarina</i>	-	-	1	-	-
<i>Klebsiella</i>	2	-	1	-	510
<i>Kluyvera</i>	-	-	-	-	1
<i>Kosakonia</i>	4	-	10	-	-
<i>Leclercia</i>	1	-	4	-	-
<i>Lysinibacillus</i>	-	-	1	-	1
<i>Macrococcus</i>	4	-	1	-	-
<i>Mangrovibacter</i>	-	-	1	-	-
<i>Massilia</i>	-	-	1	-	-
<i>Moraxella</i>	-	-	1	-	-
<i>Mycoplasma</i>	-	-	1	-	-
<i>Oceanimonas</i>	-	-	-	-	1
<i>Oceanobacillus</i>	3	-	1	-	-
<i>Orenia</i>	-	-	1	-	-
<i>Ornithinibacillus</i>	-	-	1	-	-
<i>Pandoraea</i>	-	-	1	-	-
<i>Pantoea</i>	2	-	4	-	6
<i>Pectobacterium</i>	-	-	1	-	11
<i>Pediococcus</i>	-	-	1	-	-
<i>Phaeochromatium</i>	-	-	1	-	-
<i>Photobacterium</i>	-	-	-	-	1

Genus	NCBI			SILVA	
	EPI2ME	Mothur	GraphMap	Mothur	GraphMap
<i>Photorhabdus</i>	-	-	1	-	-
<i>Planococcus</i>	-	-	-	-	1
<i>Planomicrobium</i>	-	-	-	-	2
<i>Pluralibacter</i>	-	-	19	-	-
<i>Pontibacillus</i>	-	-	-	-	54
<i>Proteus</i>	-	-	-	-	1
<i>Pseudescherichia</i>	-	-	5	-	-
<i>Pseudoalteromonas</i>	-	-	2	-	38
<i>Raoultella</i>	-	-	1	-	21
<i>Rhodovulum</i>	-	-	1	-	-
<i>Rickettsiales mitochondria_ge</i>	-	-	-	-	1
<i>Ruminiclostridium</i>	-	-	-	-	7
<i>Salinococcus</i>	1	-	-	-	-
<i>Serratia</i>	4	-	-	-	3
<i>Shewanella</i>	-	-	3	-	-
<i>Sinobaca</i>	-	-	1	-	-
<i>Sporolactobacillus</i>	-	-	-	-	1
<i>Sporotomaculum</i>	-	-	2	-	-
<i>Streptococcus</i>	-	-	-	-	33
<i>Sva0996_marine_group_ge</i>	-	-	-	-	1
<i>Tatumella</i>	1	-	-	-	-
<i>Terrilactibacillus</i>	-	-	1	-	-
<i>Thauera</i>	-	-	1	-	-
<i>Thiopfundum</i>	-	-	1	-	-
<i>Thiothrix</i>	-	-	-	-	7
<i>Thorsellia</i>	-	-	1	-	1
<i>Trabulsiella</i>	-	-	-	-	3
<i>Ursidibacter</i>	-	-	1	-	-
<i>Vagococcus</i>	-	-	-	-	2
<i>Vibrio</i>	5	-	2	-	37
<i>Virgibacillus</i>	-	-	1	-	1
<i>Xenorhabdus</i>	-	-	2	-	-
<i>Yersinia</i>	1	-	-	-	3
<i>Yokenella</i>	-	-	2	-	-
uncultured	-	-	-	-	1
TOTAL	68	1	163	0	1,311

Table S7. Overview of number of reads that were classified at the family level using the NCBI and SILVA databases when using EPI2ME, Mothur, and GraphMap for the MinION data. Only families that were not expected to be present are listed. The first column lists the family name. The second, third, and fourth columns list the number of reads identified for the family when using the NCBI 16S database and EPI2ME, Mothur, and GraphMap, respectively. The fifth and sixth columns list the number of reads identified for the genus when using the SILVA database and Mothur, and GraphMap, respectively.

Family	NCBI			SILVA	
	EPI2ME	Mothur	GraphMap	Mothur	GraphMap
0319-6M6	-	-	-	-	1
<i>Aeromonadaceae</i>	-	-	-	-	1
<i>Budviciaceae</i>	1	-	-	-	-
<i>Burkholderiaceae</i>	-	-	1	-	4
<i>Cardiobacteriaceae</i>	-	-	-	-	1
<i>Carnobacteriaceae</i>	8	-	-	-	-
<i>Chitinophagaceae</i>	-	-	1	-	-
<i>Chromatiaceae</i>	-	-	2	-	-
<i>Colwelliaceae</i>	-	-	-	-	1
<i>Erwiniaceae</i>	6	-	6	-	-
Family_III	-	-	-	-	3
<i>Halobacteroidaceae</i>	-	-	1	-	-
<i>Idiomarinaceae</i>	-	-	1	-	-
<i>Mitochondria</i>	-	-	-	-	1
<i>Moraxellaceae</i>	-	-	1	-	-
<i>Morganellaceae</i>	-	-	3	-	-
<i>Mycoplasmataceae</i>	-	-	1	-	-
<i>Orbaceae</i>	-	-	1	-	-
<i>Oxalobacteraceae</i>	-	-	1	-	-
<i>Paenibacillaceae</i>	-	-	-	-	3
<i>Pasteurellaceae</i>	-	-	1	-	-
<i>Pectobacteriaceae</i>	-	-	1	-	-
<i>Peptococcaceae</i>	-	-	2	-	-
<i>Planococcaceae</i>	1	-	-	-	4
<i>Pseudoalteromonadaceae</i>	-	-	2	-	38
<i>Rhodobacteraceae</i>	-	-	1	-	-
<i>Ruminococcaceae</i>	-	-	-	-	8
<i>Shewanellaceae</i>	-	-	3	-	-
<i>Sporolactobacillaceae</i>	-	-	1	-	1
<i>Streptococcaceae</i>	-	-	-	-	33
Sva0996_marine_group	-	-	-	-	1
<i>Thiopfundaceae</i>	-	-	1	-	-
<i>Thiotrichaceae</i>	-	-	-	-	7
<i>Thorsellaceae</i>	-	-	1	-	-
<i>Vibrionaceae</i>	8	-	2	-	38
<i>Yersiniaceae</i>	6	-	-	-	-
<i>Zoogloeaceae</i>	-	-	1	-	-
TOTAL	30	0	35	0	145

Table S8. Overview of number of reads classified in the negative-template control samples at the family level using the SILVA database for the MiSeq data. The first column lists the family name, and the next columns all list the negative-template control samples. Bold entries represent families that were part of the mock community.

Family	Sample										
	16S1	16S2	16S3	16S4	16S5	16S6	16S7	16S8	16S9	16S10	16S11
<i>Pseudomonadaceae</i>	6	-	12	-	-	-	-	-	15	-	-
<i>Staphylococcaceae</i>	-	-	-	-	-	-	2	-	-	-	-
<i>Bacillaceae</i>	-	-	-	7	-	-	-	-	1,614	-	-
<i>Actinomycetaceae</i>	-	-	-	-	-	-	-	-	13	-	-
<i>Bradyrhizobiaceae</i>	30	-	9	20	-	-	8	-	115	-	-
<i>Brucellaceae</i>	-	-	14	-	-	-	-	-	84	-	-
<i>Caulobacteraceae</i>	-	-	-	10	-	-	-	-	164	-	-
<i>Chitinophagaceae</i>	20	-	-	-	-	-	-	-	-	-	-
<i>Comamonadaceae</i>	70	-	17	36	-	-	7	-	-	-	-
<i>Corynebacteriaceae</i>	-	-	-	-	-	-	-	-	50	-	-
<i>Deinococcaceae</i>	-	-	-	-	-	-	-	-	34	-	-
Elev-16S-1332	-	-	-	-	-	-	-	-	30	-	-
<i>Halomonadaceae</i>	-	-	-	-	-	-	-	-	92	-	-
<i>Hyphomicrobiaceae</i>	-	-	-	-	-	-	-	-	13	-	-
<i>Intrasporangiaceae</i>	-	-	-	-	-	-	5	-	-	-	-
<i>Methylobacteriaceae</i>	-	-	-	-	-	-	-	-	5	-	-
<i>Micrococcaceae</i>	21	-	-	-	-	5	-	-	79	-	-
<i>Mitochondria</i>	-	-	11	-	-	-	-	-	-	-	-
<i>Moraxellaceae</i>	-	-	-	-	-	-	-	-	37	-	-
<i>Neisseriaceae</i>	2	2	-	-	-	-	-	2	-	-	-
<i>Oxalobacteraceae</i>	36	-	-	22	-	-	4	-	-	-	-
<i>Pasteurellaceae</i>	-	-	-	-	-	-	-	-	75	-	-
<i>Phyllobacteriaceae</i>	37	-	5	2	-	-	-	-	-	-	-
<i>Planococcaceae</i>	-	-	-	-	-	-	-	-	141	-	-
<i>Propionibacteriaceae</i>	3	-	-	-	-	-	19	-	-	-	-
<i>Rhizobiaceae</i>	27	-	-	-	-	-	-	-	-	-	-
<i>Rhodocyclaceae</i>	-	-	-	16	-	-	-	-	-	-	-
<i>Sphingomonadaceae</i>	25	-	-	-	-	-	4	-	-	-	-
<i>Streptococcaceae</i>	-	-	-	-	-	-	-	-	231	-	-
<i>Xanthomonadaceae</i>	-	-	7	-	-	-	2	-	-	-	-
env.OPS_17	69	-	8	25	-	-	3	-	125	-	-
TOTAL	346	2	83	138	0	5	54	2	2,917	0	0

Table S9. Overview of number of reads classified in the negative-template control samples at the family level using the NCBI 16S database the MiSeq data. The first column lists the family name, and the next columns all list the negative-template control samples. Bold entries represent families that were part of the mock community.

Family	Sample										
	16S1	16S2	16S3	16S4	16S5	16S6	16S7	16S8	16S9	16S10	16S11
<i>Pseudomonadaceae</i>	6	-	12	-	-	-	-	-	15	-	-
<i>Staphylococcaceae</i>	-	-	-	-	-	-	2	-	-	-	-
<i>Bacillaceae</i>	-	-	-	7	-	-	-	-	1,755	-	-
<i>Actinomycetaceae</i>	-	-	-	-	-	-	-	-	13	-	-
<i>Bradyrhizobiaceae</i>	30	-	9	20	-	-	8	-	78	-	-
<i>Brucellaceae</i>	-	-	14	-	-	-	-	-	-	-	-
<i>Caulobacteraceae</i>	2	-	-	10	-	-	-	-	164	-	-
<i>Chitinophagaceae</i>	20	-	-	-	-	-	-	-	-	-	-
<i>Comamonadaceae</i>	70	-	14	36	-	-	7	-	-	-	-
<i>Corynebacteriaceae</i>	-	-	-	-	-	-	-	-	22	-	-
<i>Deinococcaceae</i>	-	-	-	-	-	-	-	-	34	-	-
<i>Gordoniaceae</i>	-	-	-	-	-	-	-	-	37	-	-
<i>Halomonadaceae</i>	-	-	-	-	-	-	-	-	92	-	-
<i>Hyphomicrobiaceae</i>	-	-	-	-	-	-	-	-	13	-	-
<i>Methylobacteriaceae</i>	-	-	-	-	-	-	-	-	5	-	-
<i>Micrococcaceae</i>	22	-	-	-	-	5	-	-	79	-	-
<i>Moraxellaceae</i>	-	-	-	-	-	-	-	-	37	-	-
<i>Neisseriaceae</i>	2	-	-	-	-	-	-	-	-	-	-
<i>Oxalobacteraceae</i>	36	-	-	22	-	-	4	-	-	-	-
<i>Pasteurellaceae</i>	-	-	-	-	-	-	-	-	75	-	-
<i>Phyllobacteriaceae</i>	-	-	5	2	-	-	-	-	-	-	-
<i>Propionibacteriaceae</i>	3	-	-	-	-	-	19	-	-	-	-
<i>Rhizobiaceae</i>	27	-	-	-	-	-	-	-	34	-	-
<i>Sphingomonadaceae</i>	25	-	14	31	-	-	4	-	23	-	-
<i>Streptococcaceae</i>	-	-	-	-	-	-	-	-	232	-	-
<i>Xanthomonadaceae</i>	-	-	7	-	-	-	2	-	-	-	-
<i>Zoogloeaceae</i>	-	-	-	16	-	-	-	-	-	-	-
TOTAL	243	0	75	144	0	5	46	0	2,708	0	0

Table S10. Overview of number of reads classified in the negative-template control samples at the genus level using the SILVA database for the MiSeq data. The first column lists the genus name, and the next columns all list the negative-template control samples. Bold entries represent families that were part of the mock community.

Genus	Sample										
	16S1	16S2	16S3	16S4	16S5	16S6	16S7	16S8	16S9	16S10	16S11
<i>Pseudomonas</i>	6	-	12	-	-	-	-	-	-	-	-
<i>Staphylococcus</i>	-	-	-	-	-	-	2	-	-	-	-
<i>Actinobacillus</i>	-	-	-	-	-	-	-	-	75	-	-
<i>Anaerobacillus</i>	-	-	-	7	-	-	-	-	1,607	-	-
<i>Bosea</i>	15	-	-	-	-	-	-	-	37	-	-
<i>Bradyrhizobium</i>	-	-	9	-	-	-	-	-	78	-	-
<i>Brevundimonas</i>	-	-	-	-	-	-	-	-	24	-	-
<i>Brucella</i>	-	-	14	-	-	-	-	-	-	-	-
<i>Deinococcus</i>	-	-	-	-	-	-	-	-	34	-	-
<i>Delftia</i>	53	-	12	36	-	-	-	-	-	-	-
<i>Devosia</i>	-	-	-	-	-	-	-	-	13	-	-
<i>Enhydrobacter</i>	-	-	-	-	-	-	-	-	37	-	-
<i>Herbaspirillum</i>	-	-	-	22	-	-	-	-	-	-	-
<i>Lacibacter</i>	20	-	-	-	-	-	-	-	-	-	-
<i>Lysinibacillus</i>	-	-	-	-	-	-	-	-	141	-	-
<i>Mesorhizobium</i>	-	-	5	2	-	-	-	-	-	-	-
<i>Methylobacterium</i>	-	-	-	-	-	-	-	-	5	-	-
<i>Micrococcus</i>	21	-	-	-	-	-	-	-	-	-	-
<i>Neisseria</i>	2	-	-	-	-	-	-	-	-	-	-
<i>Ornithinimicrobium</i>	-	-	-	-	-	-	5	-	-	-	-
<i>Pelomonas</i>	-	-	-	-	-	-	7	-	-	-	-
<i>Propionibacterium</i>	3	-	-	-	-	-	19	-	-	-	-
<i>Rhizobium</i>	27	-	-	-	-	-	-	-	-	-	-
<i>Sphingomonas</i>	-	-	-	-	-	-	4	-	-	-	-
<i>Stenotrophomonas</i>	-	-	7	-	-	-	2	-	-	-	-
<i>Streptococcus</i>	-	-	-	-	-	-	-	-	231	-	-
<i>Thauera</i>	-	-	-	16	-	-	-	-	-	-	-
<i>Undibacterium</i>	-	-	-	-	-	-	4	-	-	-	-
uncultured	-	-	-	10	-	-	-	-	-	-	-
TOTAL	147	0	59	93	0	0	43	0	2,282	0	0

Table S11. Overview of number of reads classified in the negative-template control samples at the genus level using the NCBI 16S database for the MiSeq data. The first column lists the genus name, and the next columns all list the negative-template control samples. Bold entries represent families that were part of the mock community.

Genus	Sample										
	16S1	16S2	16S3	16S4	16S5	16S6	16S7	16S8	16S9	16S10	16S11
<i>Pseudomonas</i>	6	-	12	-	-	-	-	-	15	-	-
<i>Staphylococcus</i>	-	-	-	-	-	-	2	-	-	-	-
<i>Bacillus</i>	-	-	-	7	-	-	-	-	4	-	-
<i>Actinomyces</i>	-	-	-	-	-	-	-	-	13	-	-
<i>Agrobacterium</i>	-	-	-	-	-	-	-	-	34	-	-
<i>Bosea</i>	15	-	-	-	-	-	-	-	-	-	-
<i>Bradyrhizobium</i>	15	-	9	20	-	-	8	-	78	-	-
<i>Brevundimonas</i>	-	-	-	-	-	-	-	-	24	-	-
<i>Caulobacter</i>	-	-	-	-	-	-	-	-	140	-	-
<i>Corynebacterium</i>	-	-	-	-	-	-	-	-	22	-	-
<i>Cutibacterium</i>	3	-	-	-	-	-	19	-	-	-	-
<i>Deinococcus</i>	-	-	-	-	-	-	-	-	34	-	-
<i>Delftia</i>	53	-	12	36	-	-	-	-	-	-	-
<i>Devosia</i>	-	-	-	-	-	-	-	-	13	-	-
<i>Gordonia</i>	-	-	-	-	-	-	-	-	37	-	-
<i>Halomonas</i>	-	-	-	-	-	-	-	-	92	-	-
<i>Herbaspirillum</i>	36	-	-	22	-	-	-	-	-	-	-
<i>Lysinibacillus</i>	-	-	-	-	-	-	-	-	141	-	-
<i>Mesorhizobium</i>	-	-	5	2	-	-	-	-	-	-	-
<i>Methylobacterium</i>	-	-	-	-	-	-	-	-	5	-	-
<i>Micrococcus</i>	22	-	-	-	-	-	-	-	-	-	-
<i>Moraxella</i>	-	-	-	-	-	-	-	-	37	-	-
<i>Neisseria</i>	2	-	-	-	-	-	-	-	-	-	-
<i>Nesterenkonia</i>	-	-	-	-	-	5	-	-	-	-	-
<i>Ochrobactrum</i>	-	-	14	-	-	-	-	-	-	-	-
<i>Pelomonas</i>	-	-	-	-	-	-	7	-	-	-	-
<i>Phenylobacterium</i>	2	-	-	10	-	-	-	-	-	-	-
<i>Rhizobium</i>	27	-	-	-	-	-	-	-	-	-	-
<i>Sphingomonas</i>	25	-	14	31	-	-	4	-	23	-	-
<i>Stenotrophomonas</i>	-	-	7	-	-	-	2	-	-	-	-
<i>Streptococcus</i>	-	-	-	-	-	-	-	-	232	-	-
<i>Thauera</i>	-	-	-	16	-	-	-	-	-	-	-
<i>Undibacterium</i>	-	-	-	-	-	-	4	-	-	-	-
<i>Variovorax</i>	-	-	2	-	-	-	-	-	-	-	-
TOTAL	206	0	75	144	0	5	46	0	944	0	0

Table S12. Overview of number of reads classified in the negative-template control samples at the species level using the NCBI 16S database for the MiSeq data. The first column lists the species name, and the next columns all list the negative-template control samples. Bold entries represent families that were part of the mock community.

Species	Sample										
	16S1	16S2	16S3	16S4	16S5	16S6	16S7	16S8	16S9	16S10	16S11
<i>Bradyrhizobium japonicum</i>	-	-	-	-	-	-	-	-	78	-	-
<i>Brevundimonas diminuta</i>	-	-	-	-	-	-	-	-	24	-	-
<i>Caulobacter henricii</i>	-	-	-	-	-	-	-	-	140	-	-
<i>Cutibacterium acnes</i>	3	-	-	-	-	-	19	-	-	-	-
<i>Delftia acidovorans</i>	53	-	12	36	-	-	-	-	-	-	-
<i>Gordonia rubripertincta</i>	-	-	-	-	-	-	-	-	37	-	-
<i>Herbaspirillum huttiense</i>	36	-	-	22	-	-	-	-	-	-	-
<i>Lysinibacillus sphaericus</i>	-	-	-	-	-	-	-	-	141	-	-
<i>Micrococcus luteus</i>	22	-	-	-	-	-	-	-	-	-	-
<i>Nesterenkonia lacusekhoensis</i>	-	-	-	-	-	5	-	-	-	-	-
<i>Ochrobactrum grignonense</i>	-	-	14	-	-	-	-	-	-	-	-
<i>Pelomonas saccharophila</i>	-	-	-	-	-	-	7	-	-	-	-
<i>Pseudomonas migulae</i>	-	-	4	-	-	-	-	-	-	-	-
<i>Rhizobium tropici</i>	27	-	-	-	-	-	-	-	-	-	-
<i>Sphingomonas aquatilis</i>	-	-	-	14	-	-	4	-	-	-	-
<i>Sphingomonas echinoides</i>	-	-	-	17	-	-	-	-	23	-	-
<i>Staphylococcus epidermidis</i>	-	-	-	-	-	-	2	-	-	-	-
<i>Stenotrophomonas maltophilia</i>	-	-	7	-	-	-	2	-	-	-	-
<i>Thauera selenatis</i>	-	-	-	16	-	-	-	-	-	-	-
<i>Undibacterium oligocarboniphilum</i>	-	-	-	-	-	-	4	-	-	-	-
TOTAL	141	0	37	105	0	5	38	0	443	0	0

5. Supplementary figures

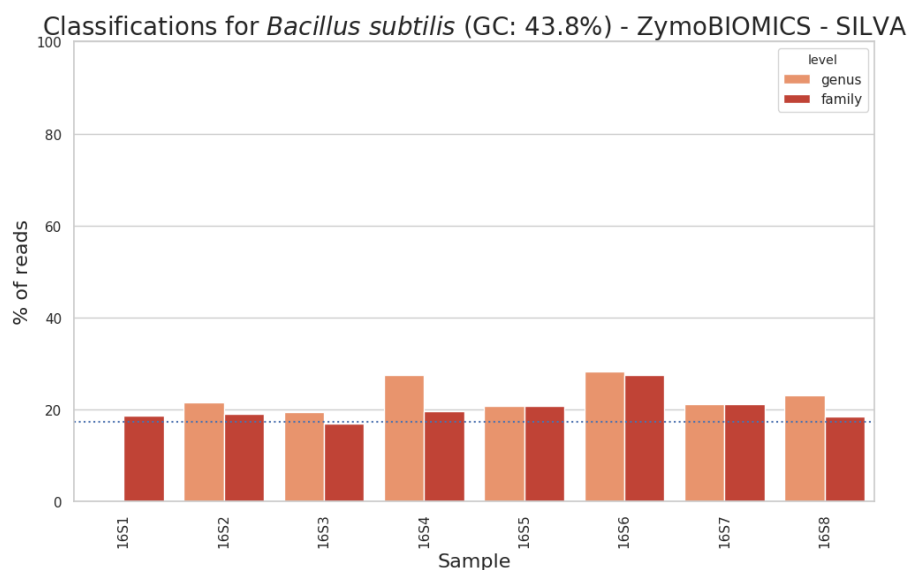


Figure S1. Overview of the correct identification of *Bacillus subtilis* in the different samples when using the SILVA database for the MiSeq data. Bars represent the percentages of reads that are correctly identified at the indicated taxonomic level. The dotted line represents the fraction of 16S rRNA present for this organism in the mock community.

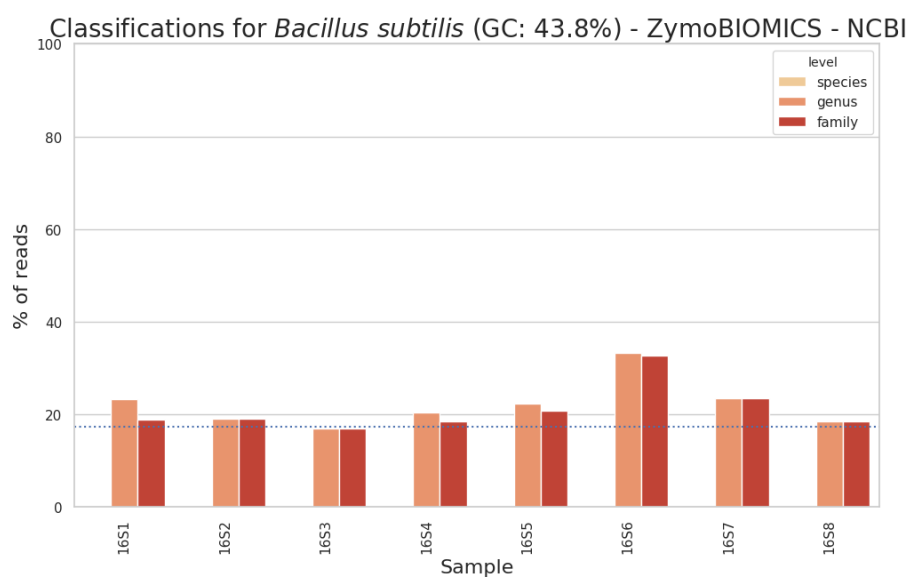


Figure S2. Overview of the correct identification of *Bacillus subtilis* in the different samples when using the NCBI 16S database for the MiSeq data. Bars represent the percentages of reads that are correctly identified at the indicated taxonomic level. The dotted line represents the fraction of 16S rRNA present for this organism in the mock community.

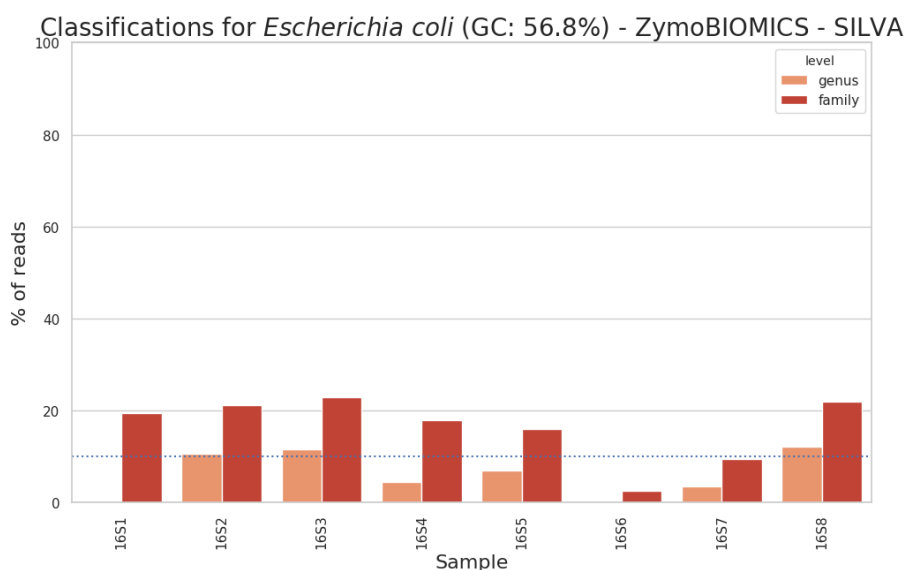


Figure S3. Overview of the correct identification of *Escherichia coli* in the different samples when using the SILVA database for the MiSeq data. Bars represent the percentages of reads that are correctly identified at the indicated taxonomic level. The dotted line represents the fraction of 16S rRNA present for this organism in the mock community.

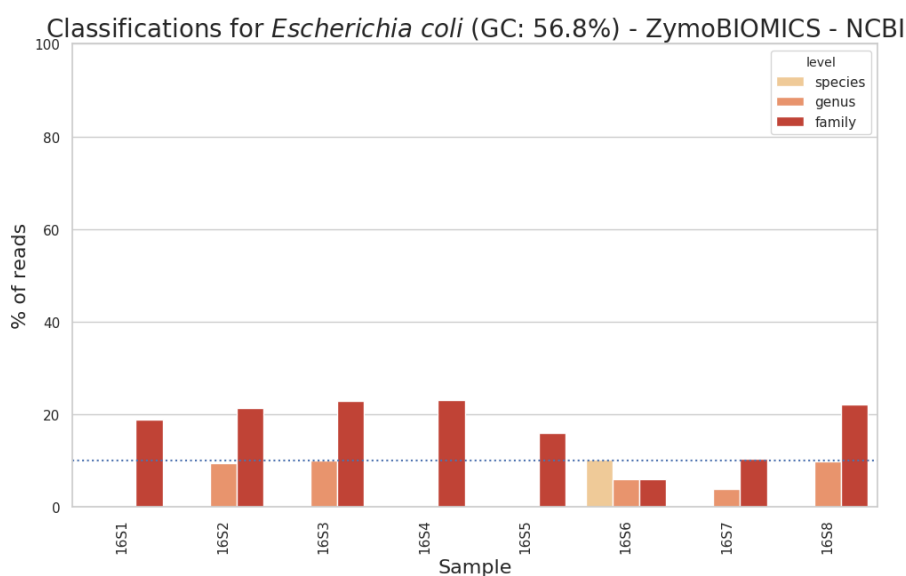


Figure S4. Overview of the correct identification of *Escherichia coli* in the different samples when using the NCBI 16S database for the MiSeq data. Bars represent the percentages of reads that are correctly identified at the indicated taxonomic level. The dotted line represents the fraction of 16S rRNA present for this organism in the mock community.

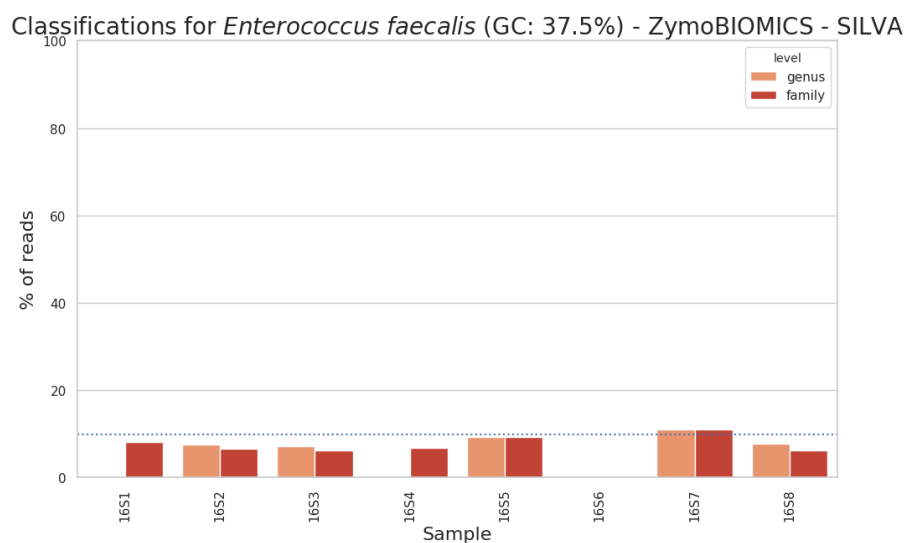


Figure S5. Overview of the correct identification of *Enterococcus faecalis* in the different samples when using the SILVA database for the MiSeq data. Bars represent the percentages of reads that are correctly identified at the indicated taxonomic level. The dotted line represents the fraction of 16S rRNA present for this organism in the mock community.

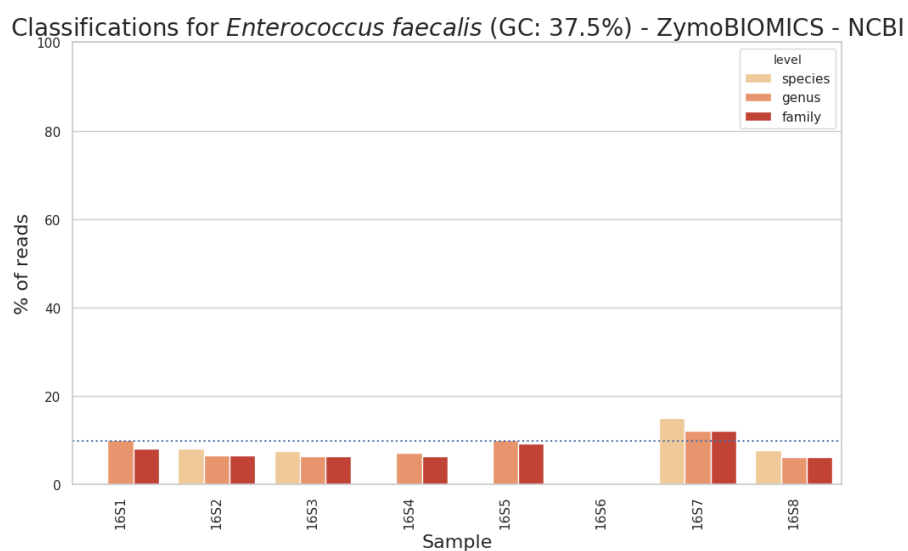


Figure S6. Overview of the correct identification of *Enterococcus faecalis* in the different samples when using the NCBI 16S database for the MiSeq data. Bars represent the percentages of reads that are correctly identified at the indicated taxonomic level. The dotted line represents the fraction of 16S rRNA present for this organism in the mock community.

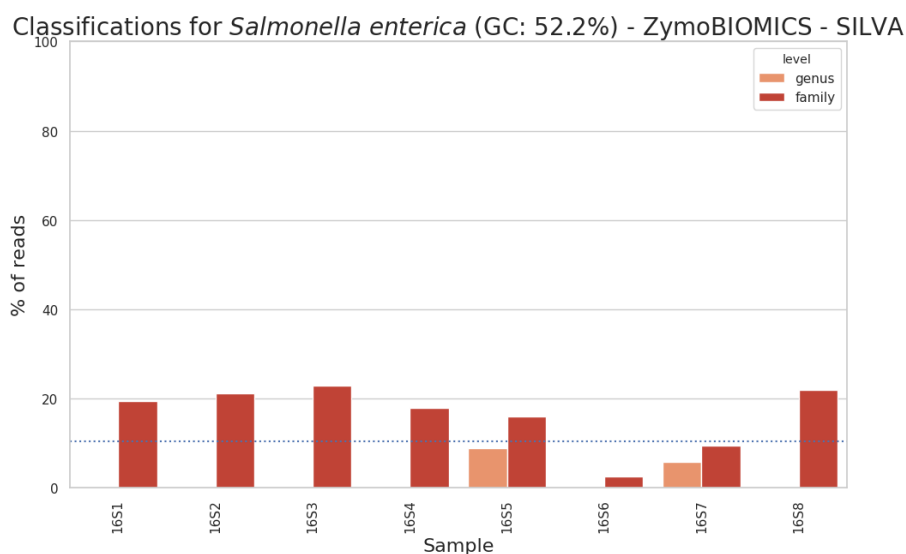


Figure S7. Overview of the correct identification of *Salmonella enterica* in the different samples when using the SILVA database for the MiSeq data. Bars represent the percentages of reads that are correctly identified at the indicated taxonomic level. The dotted line represents the fraction of 16S rRNA present for this organism in the mock community.

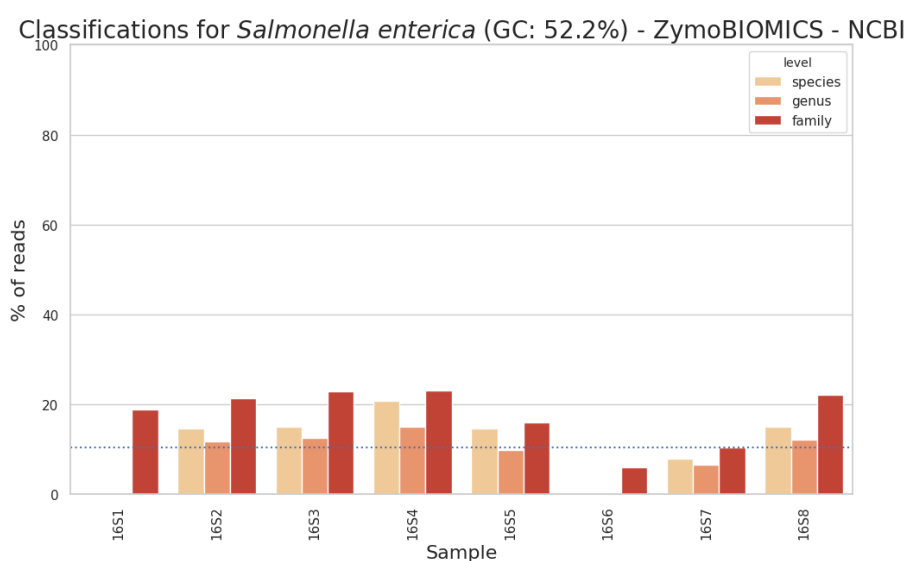


Figure S8. Overview of the correct identification of *Salmonella enterica* in the different samples when using the NCBI 16S database for the MiSeq data. Bars represent the percentages of reads that are correctly identified at the indicated taxonomic level. The dotted line represents the fraction of 16S rRNA present for this organism in the mock community.

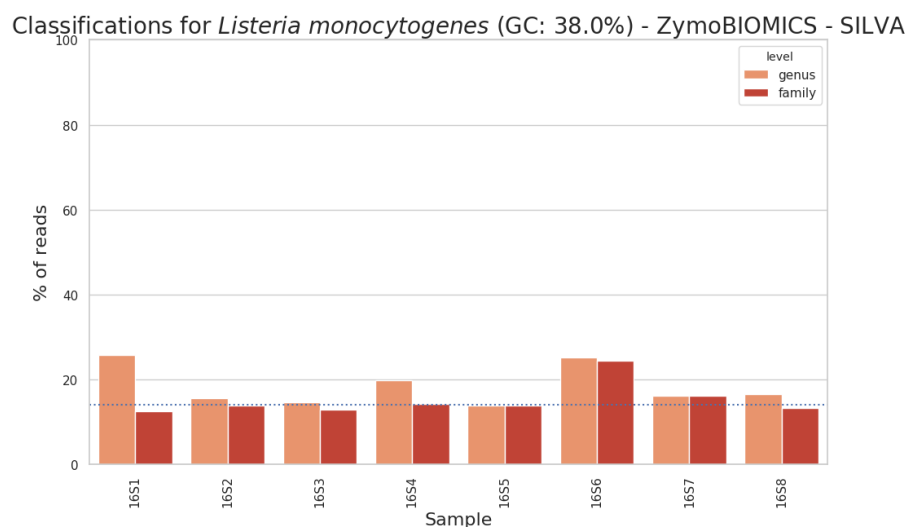


Figure S9. Overview of the correct identification of *Listeria monocytogenes* in the different samples when using the SILVA database for the MiSeq data. Bars represent the percentages of reads that are correctly identified at the indicated taxonomic level. The dotted line represents the fraction of 16S rRNA present for this organism in the mock community.

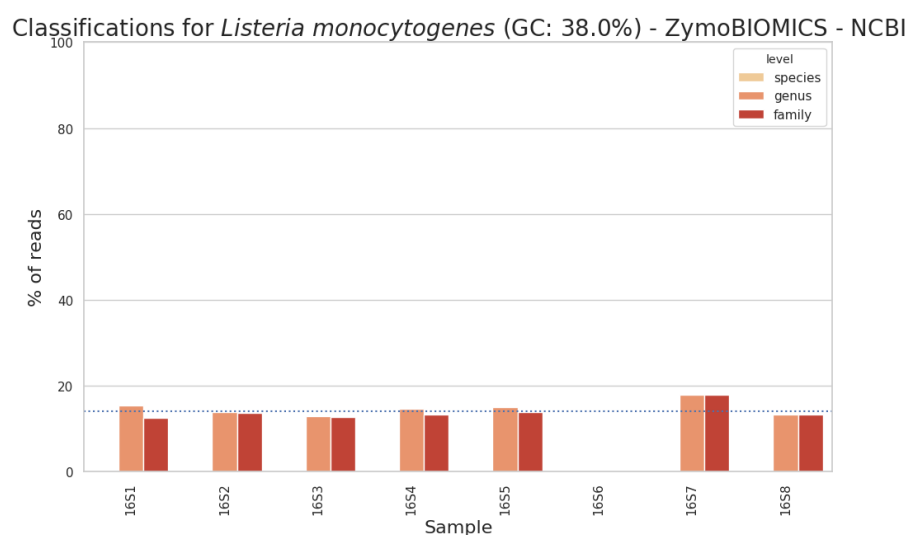


Figure S10. Overview of the correct identification of *Listeria monocytogenes* in the different samples when using the NCBI 16S database for the MiSeq data. Bars represent the percentages of reads that are correctly identified at the indicated taxonomic level. The dotted line represents the fraction of 16S rRNA present for this organism in the mock community.

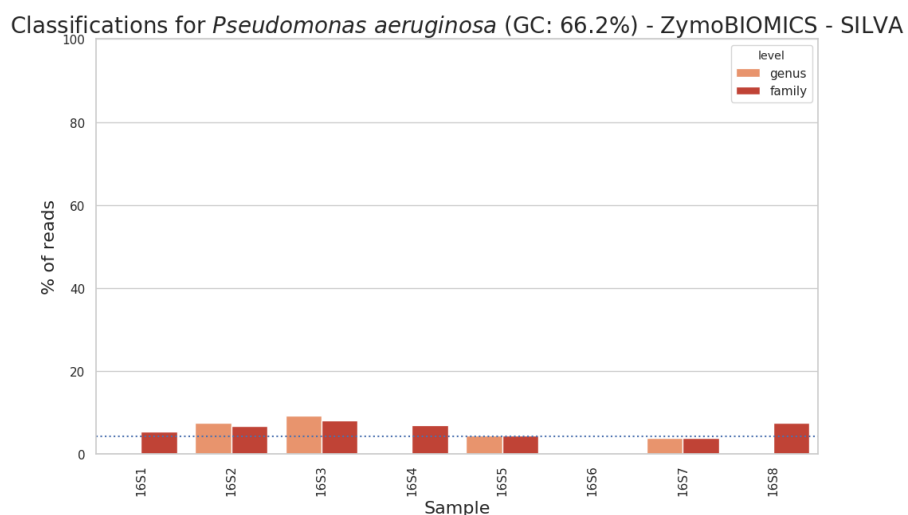


Figure S11. Overview of the correct identification of *Pseudomonas aeruginosa* in the different samples when using the SILVA database for the MiSeq data. Bars represent the percentages of reads that are correctly identified at the indicated taxonomic level. The dotted line represents the fraction of 16S rRNA present for this organism in the mock community.

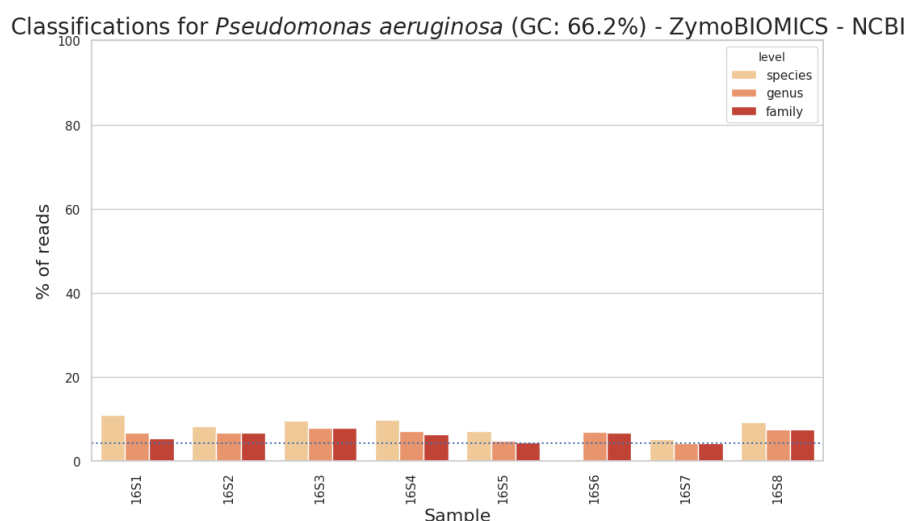


Figure S12. Overview of the correct identification of *Pseudomonas aeruginosa* in the different samples when using the NCBI 16S database for the MiSeq data. Bars represent the percentages of reads that are correctly identified at the indicated taxonomic level. The dotted line represents the fraction of 16S rRNA present for this organism in the mock community.

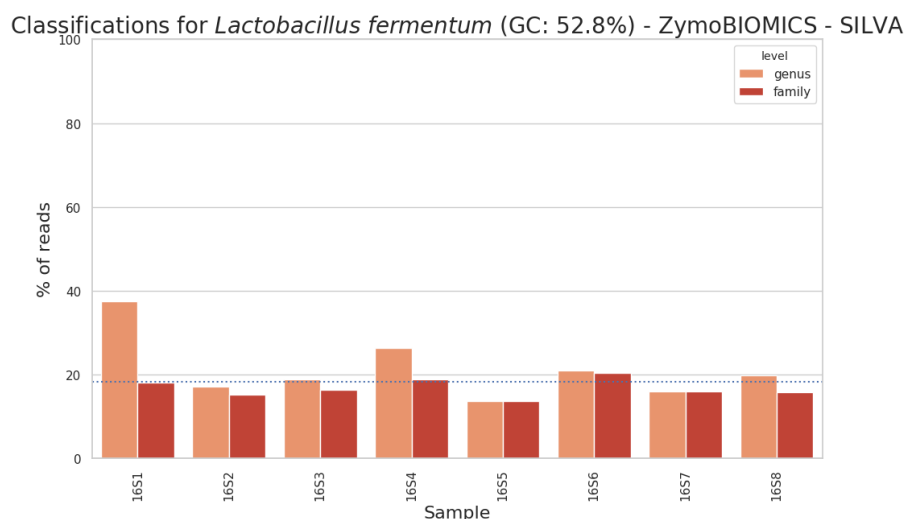


Figure S13. Overview of the correct identification of *Lactobacillus fermentum* in the different samples when using the SILVA database for the MiSeq data. Bars represent the percentages of reads that are correctly identified at the indicated taxonomic level. The dotted line represents the fraction of 16S rRNA present for this organism in the mock community.

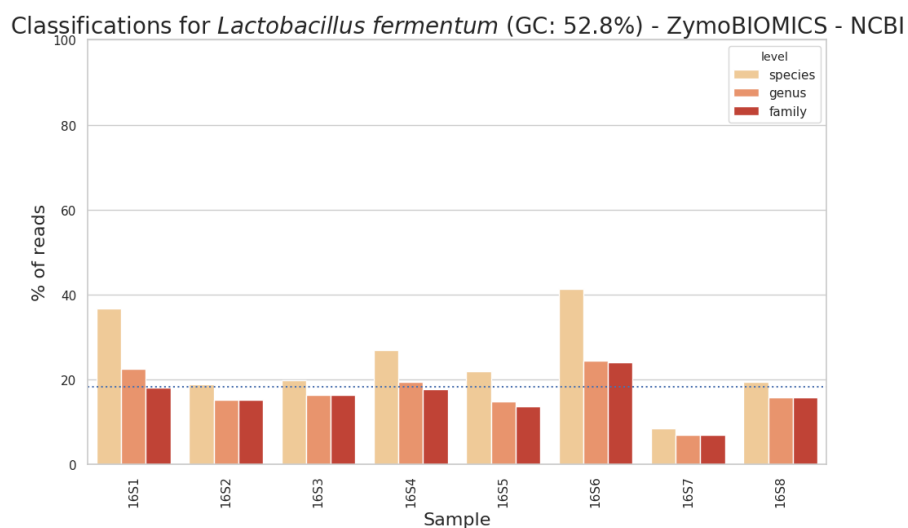


Figure S14. Overview of the correct identification of *Lactobacillus fermentum* in the different samples when using the NCBI 16S database for the MiSeq data. Bars represent the percentages of reads that are correctly identified at the indicated taxonomic level. The dotted line represents the fraction of 16S rRNA present for this organism in the mock community.

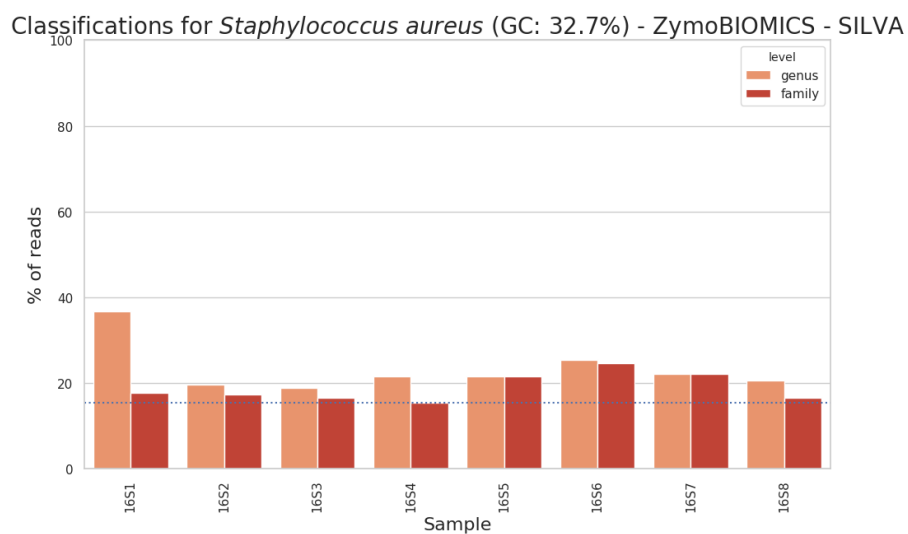


Figure S15. Overview of the correct identification of *Staphylococcus aureus* in the different samples when using the SILVA database for the MiSeq data. Bars represent the percentages of reads that are correctly identified at the indicated taxonomic level. The dotted line represents the fraction of 16S rRNA present for this organism in the mock community.

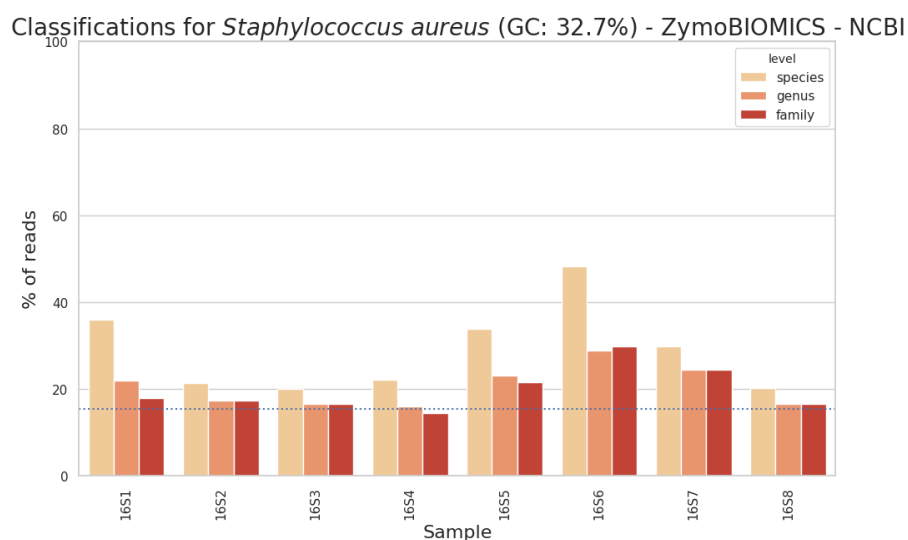


Figure S16. Overview of the correct identification of *Staphylococcus aureus* in the different samples when using the NCBI 16S database for the MiSeq data. Bars represent the percentages of reads that are correctly identified at the indicated taxonomic level. The dotted line represents the fraction of 16S rRNA present for this organism in the mock community.

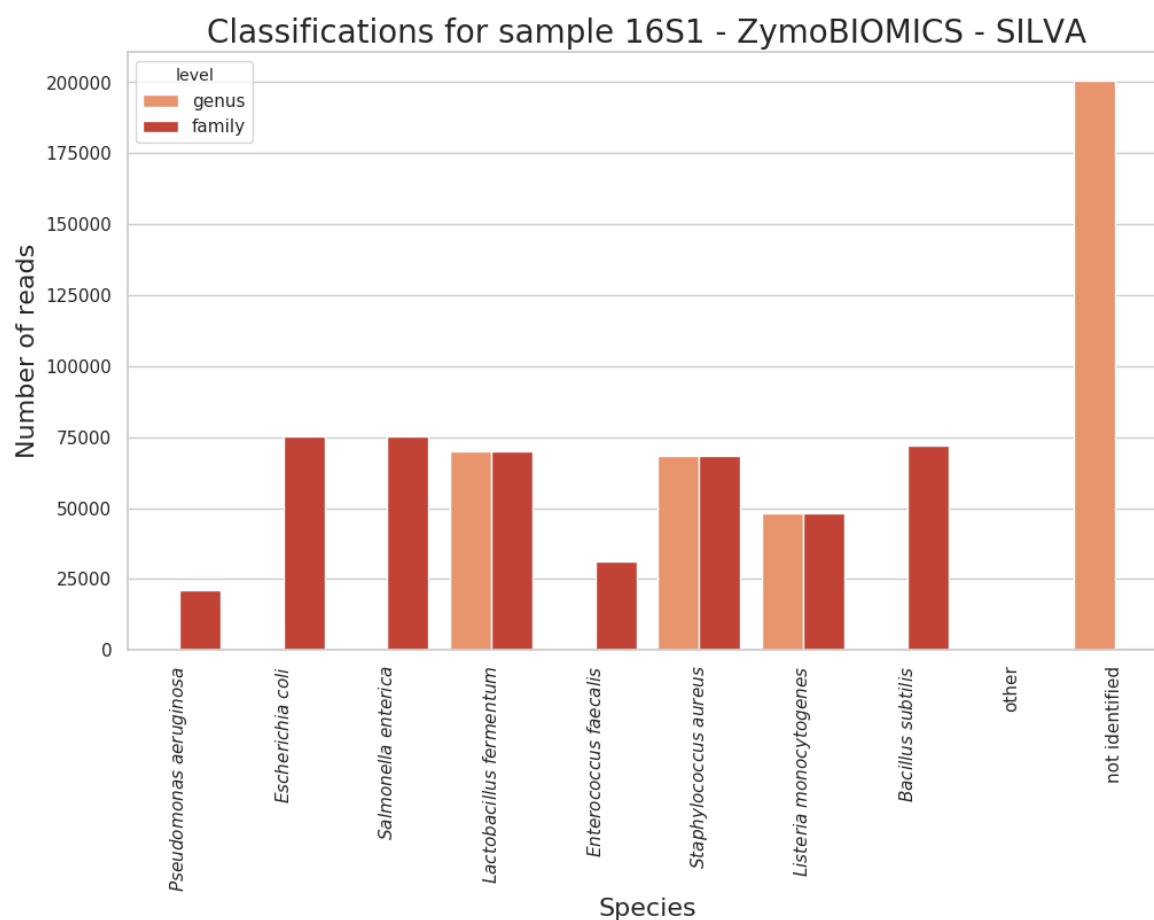


Figure S17. Overview of the number of reads that were correctly identified at the different taxonomic levels for sample 16S1 when using the SILVA database for the MiSeq data. The entry "other" considers all reads that were classified but did not match any of the bacteria present in the sample. The entry "not identified" considers all reads that were not identified at the given taxonomic level but were identified at a higher taxonomic level.

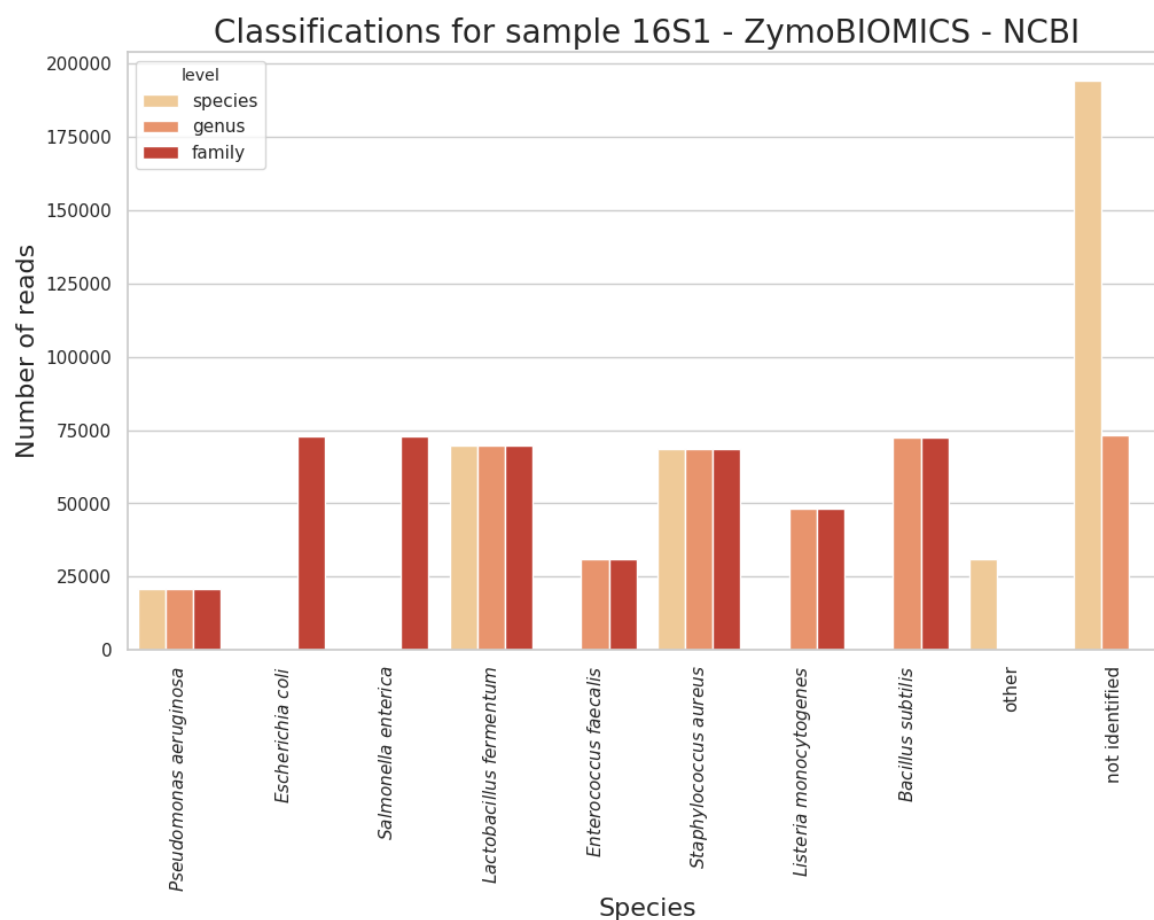


Figure S18. Overview of the number of reads that were correctly identified at the different taxonomic levels for sample 16S1 when using the NCBI 16S database for the MiSeq data. The entry "other" considers all reads that were classified but did not match any of the bacteria present in the sample. The entry "not identified" considers all reads that were not identified at the given taxonomic level but were identified at a higher taxonomic level.

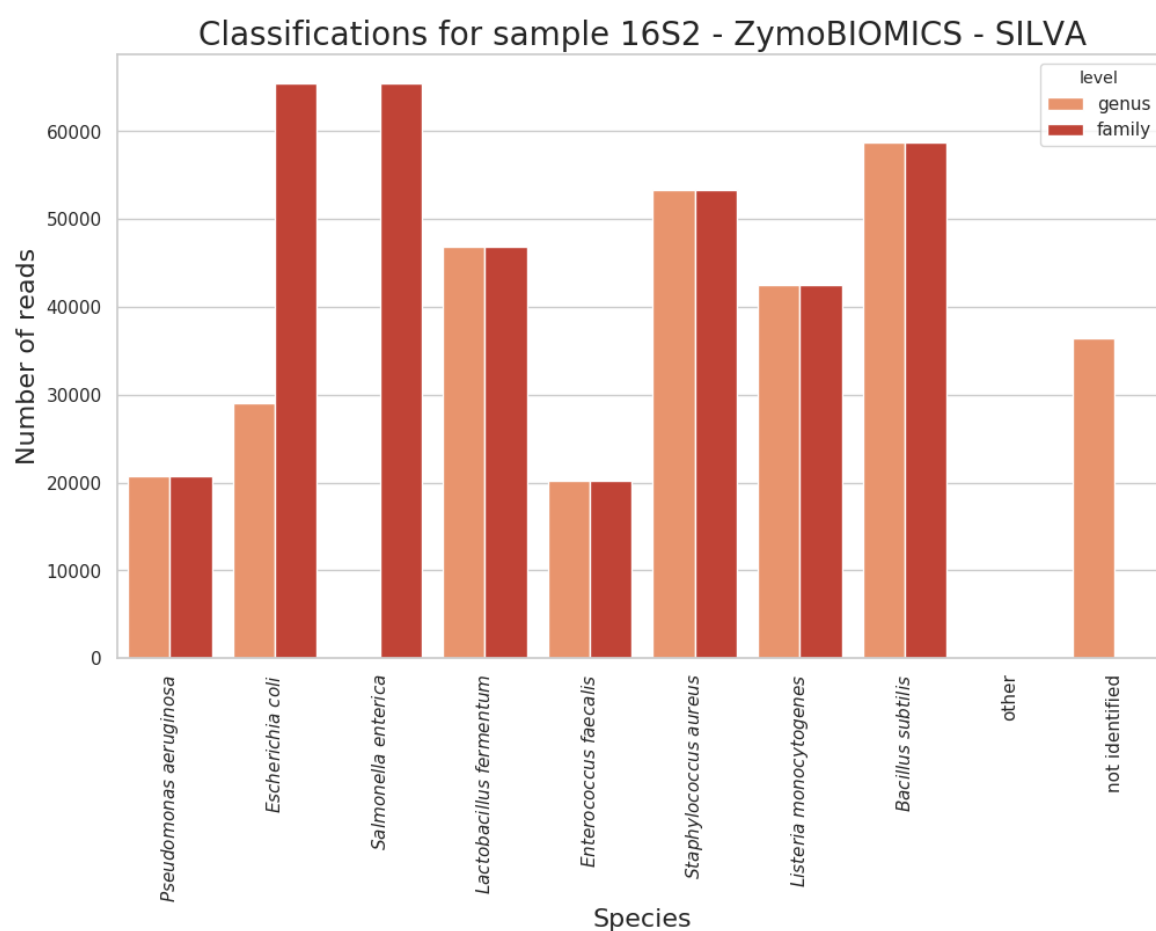


Figure S19. Overview of the number of reads that were correctly identified at the different taxonomic levels for sample 16S2 when using the SILVA database for the MiSeq data. The entry "other" considers all reads that were classified but did not match any of the bacteria present in the sample. The entry "not identified" considers all reads that were not identified at the given taxonomic level but were identified at a higher taxonomic level.

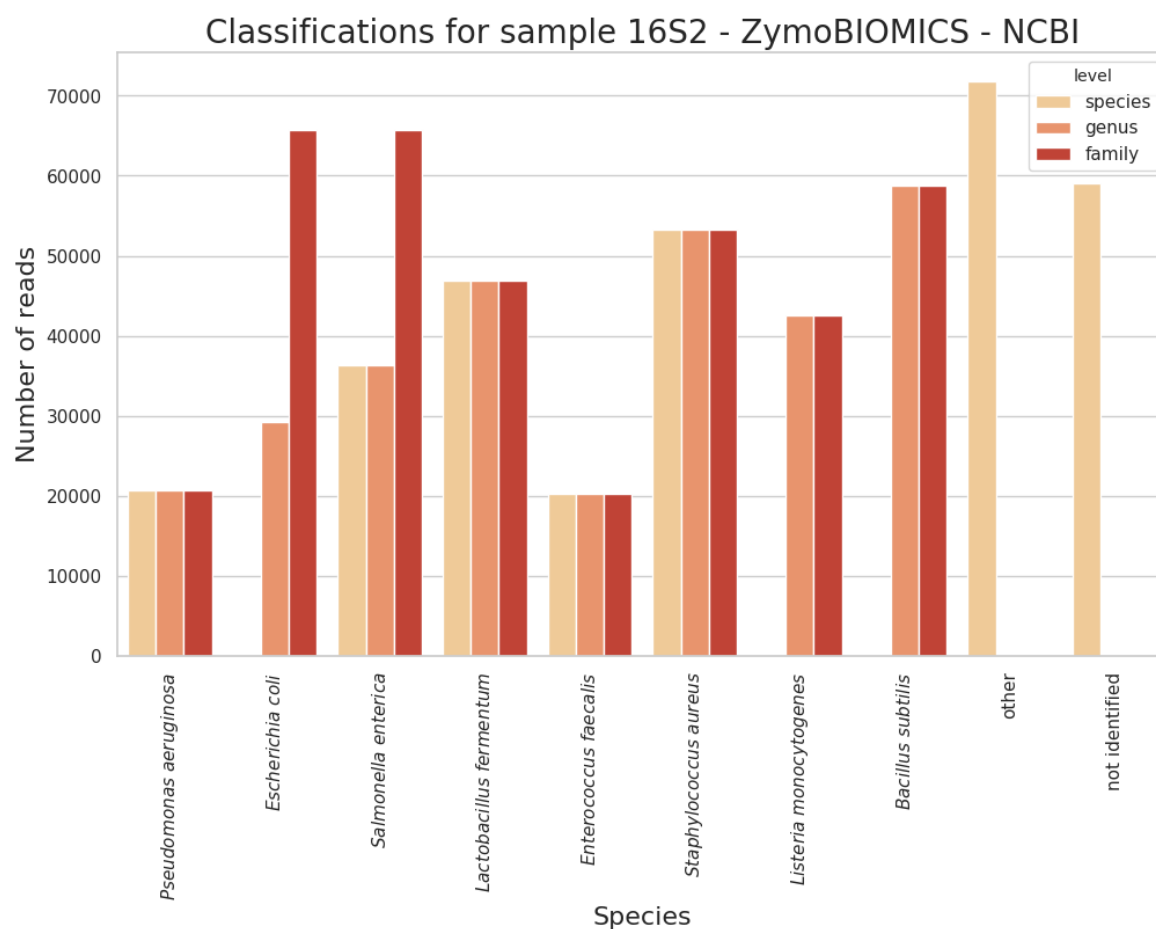


Figure S20. Overview of the number of reads that were correctly identified at the different taxonomic levels for sample 16S2 when using the NCBI 16S database for the MiSeq data. The entry "other" considers all reads that were classified but did not match any of the bacteria present in the sample. The entry "not identified" considers all reads that were not identified at the given taxonomic level but were identified at a higher taxonomic level.

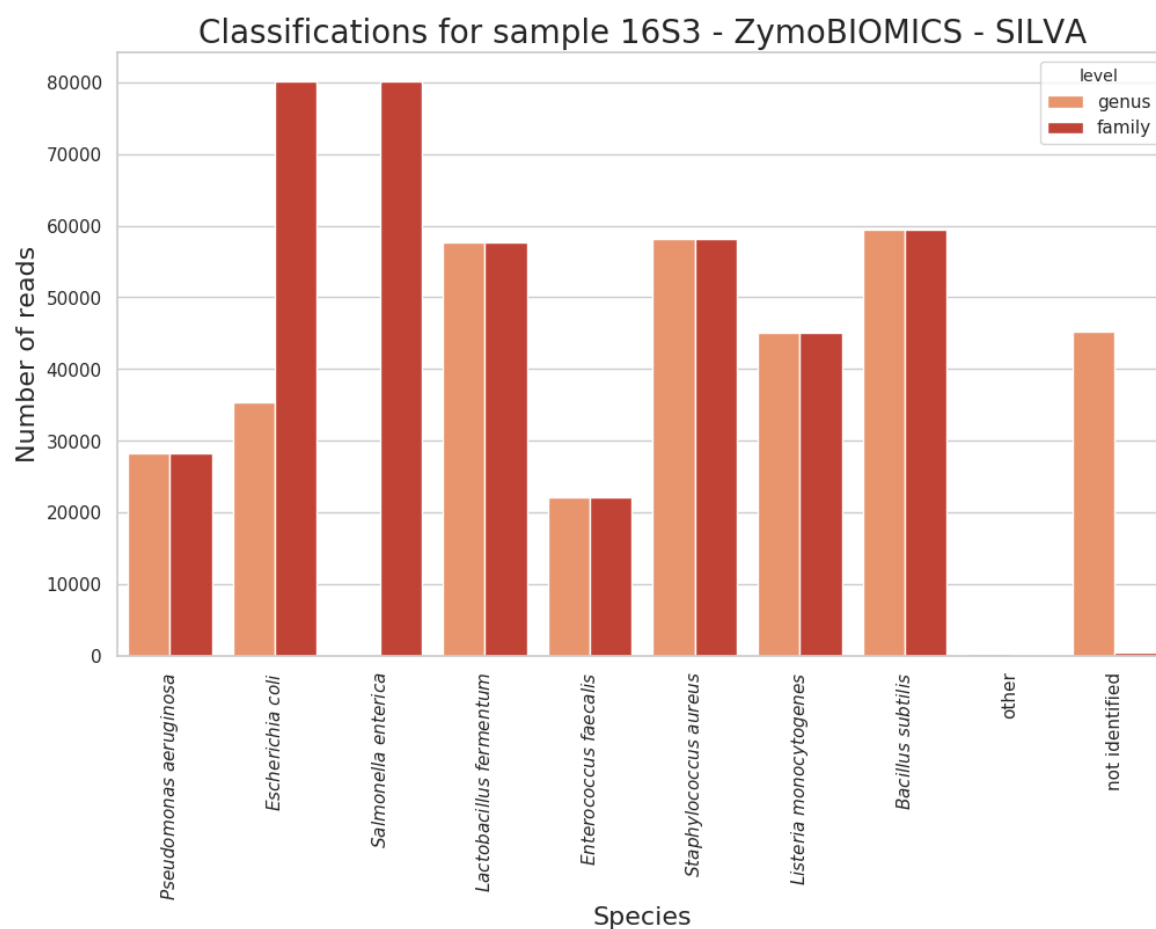


Figure S21. Overview of the number of reads that were correctly identified at the different taxonomic levels for sample 16S3 when using the SILVA database for the MiSeq data. The entry "other" considers all reads that were classified but did not match any of the bacteria present in the sample. The entry "not identified" considers all reads that were not identified at the given taxonomic level but were identified at a higher taxonomic level.

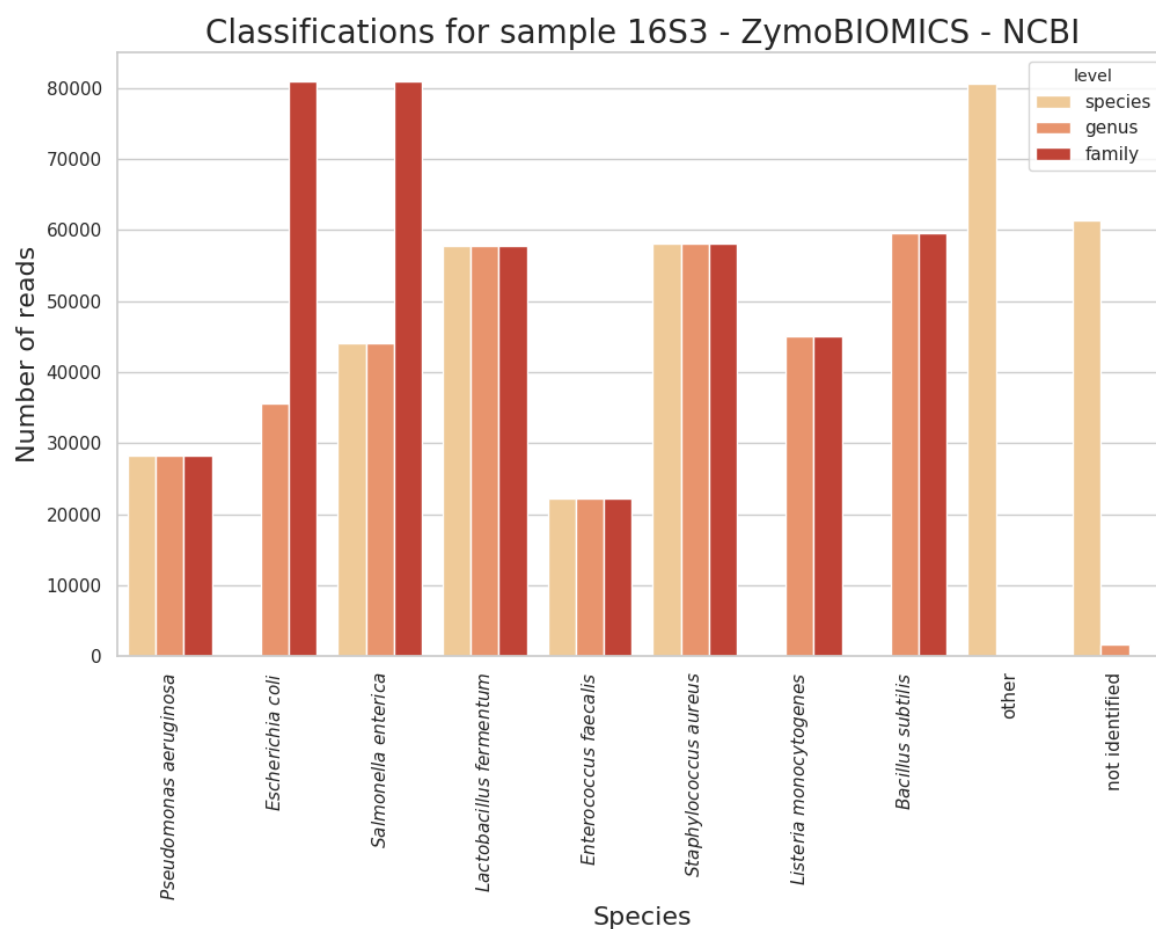


Figure S22. Overview of the number of reads that were correctly identified at the different taxonomic levels for sample 16S3 when using the NCBI 16S database for the MiSeq data. The entry "other" considers all reads that were classified but did not match any of the bacteria present in the sample. The entry "not identified" considers all reads that were not identified at the given taxonomic level but were identified at a higher taxonomic level.

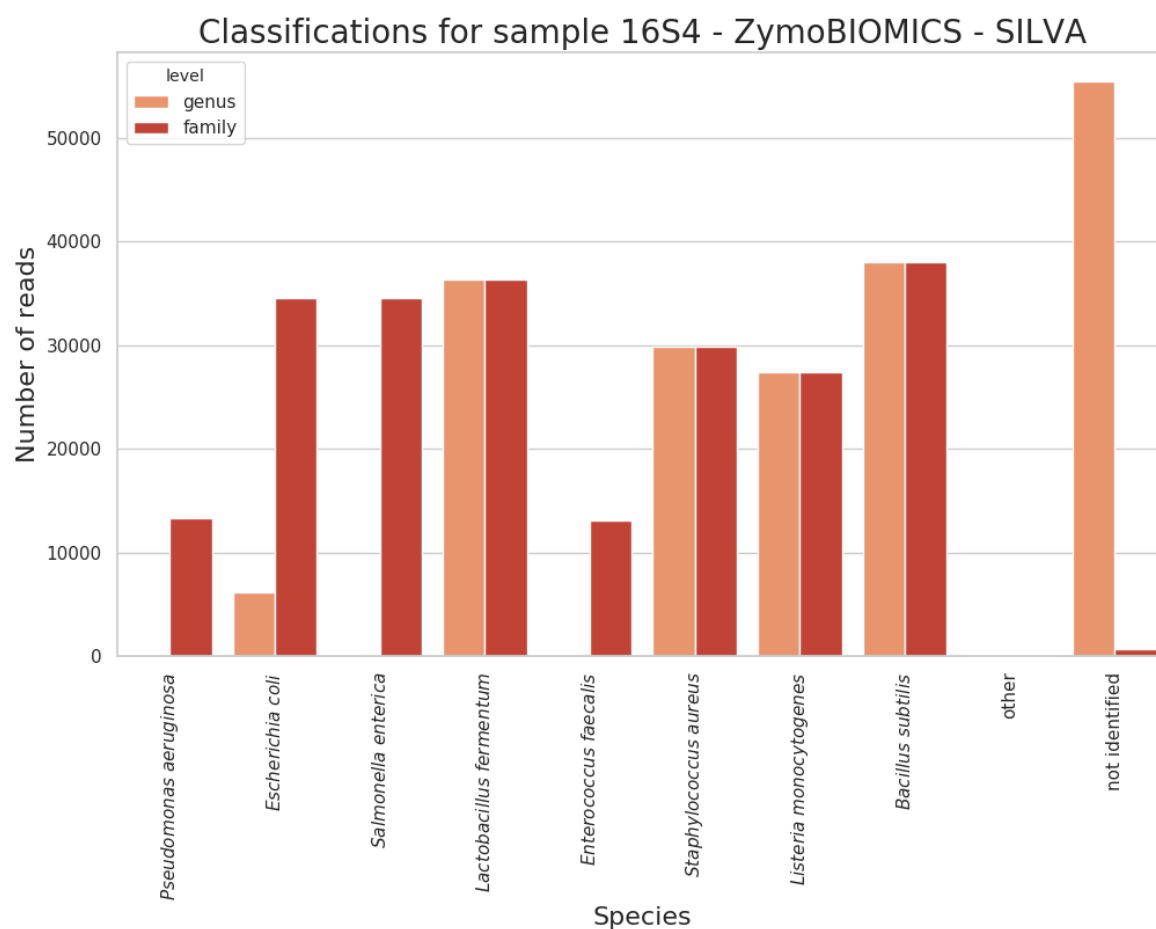


Figure S23. Overview of the number of reads that were correctly identified at the different taxonomic levels for sample 16S4 when using the SILVA database for the MiSeq data. The entry "other" considers all reads that were classified but did not match any of the bacteria present in the sample. The entry "not identified" considers all reads that were not identified at the given taxonomic level but were identified at a higher taxonomic level.



Figure S24. Overview of the number of reads that were correctly identified at the different taxonomic levels for sample 16S4 when using the NCBI 16S database for the MiSeq data. The entry "other" considers all reads that were classified but did not match any of the bacteria present in the sample. The entry "not identified" considers all reads that were not identified at the given taxonomic level but were identified at a higher taxonomic level.

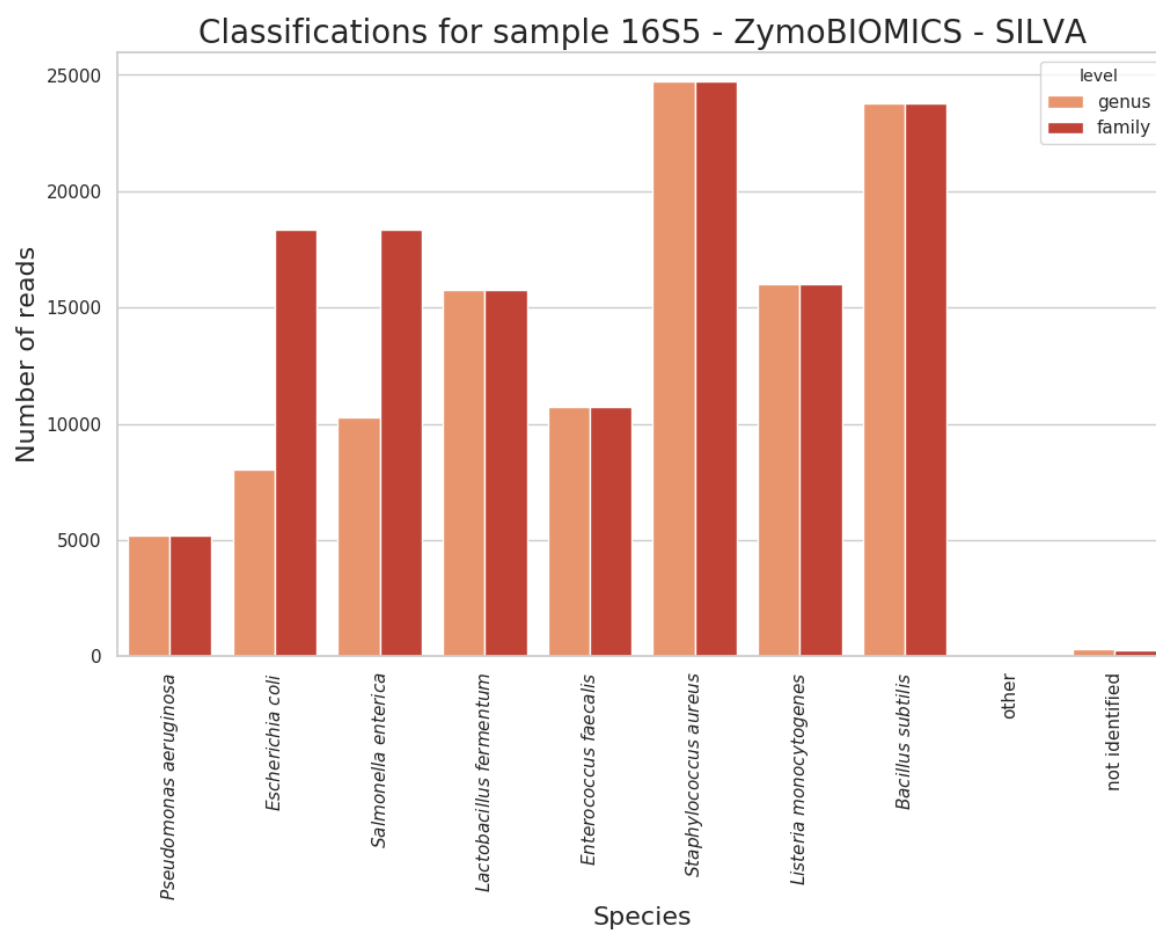


Figure S25. Overview of the number of reads that were correctly identified at the different taxonomic levels for sample 16S5 when using the SILVA database for the MiSeq data. The entry "other" considers all reads that were classified but did not match any of the bacteria present in the sample. The entry "not identified" considers all reads that were not identified at the given taxonomic level but were identified at a higher taxonomic level.

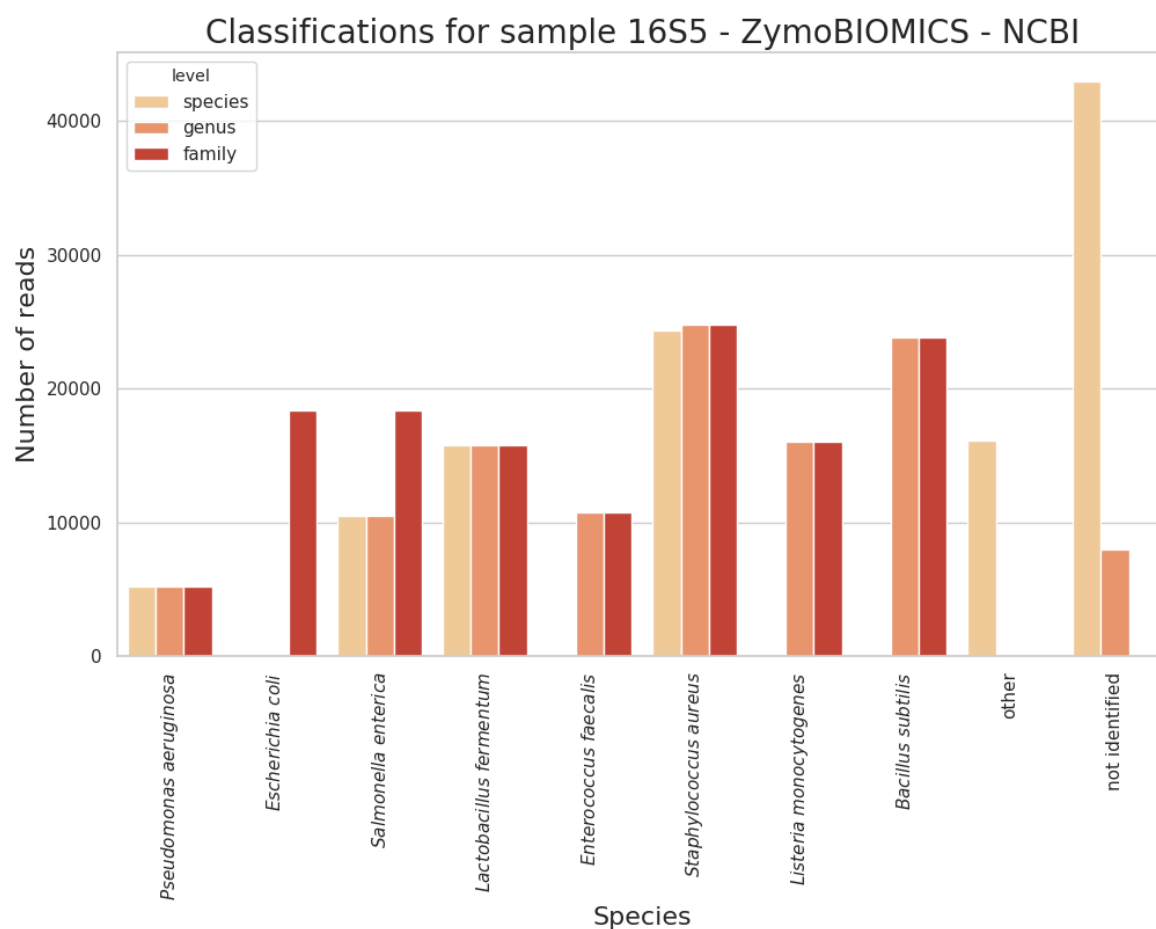


Figure S26. Overview of the number of reads that were correctly identified at the different taxonomic levels for sample 16S5 when using the NCBI 16S database for the MiSeq data. The entry "other" considers all reads that were classified but did not match any of the bacteria present in the sample. The entry "not identified" considers all reads that were not identified at the given taxonomic level but were identified at a higher taxonomic level.

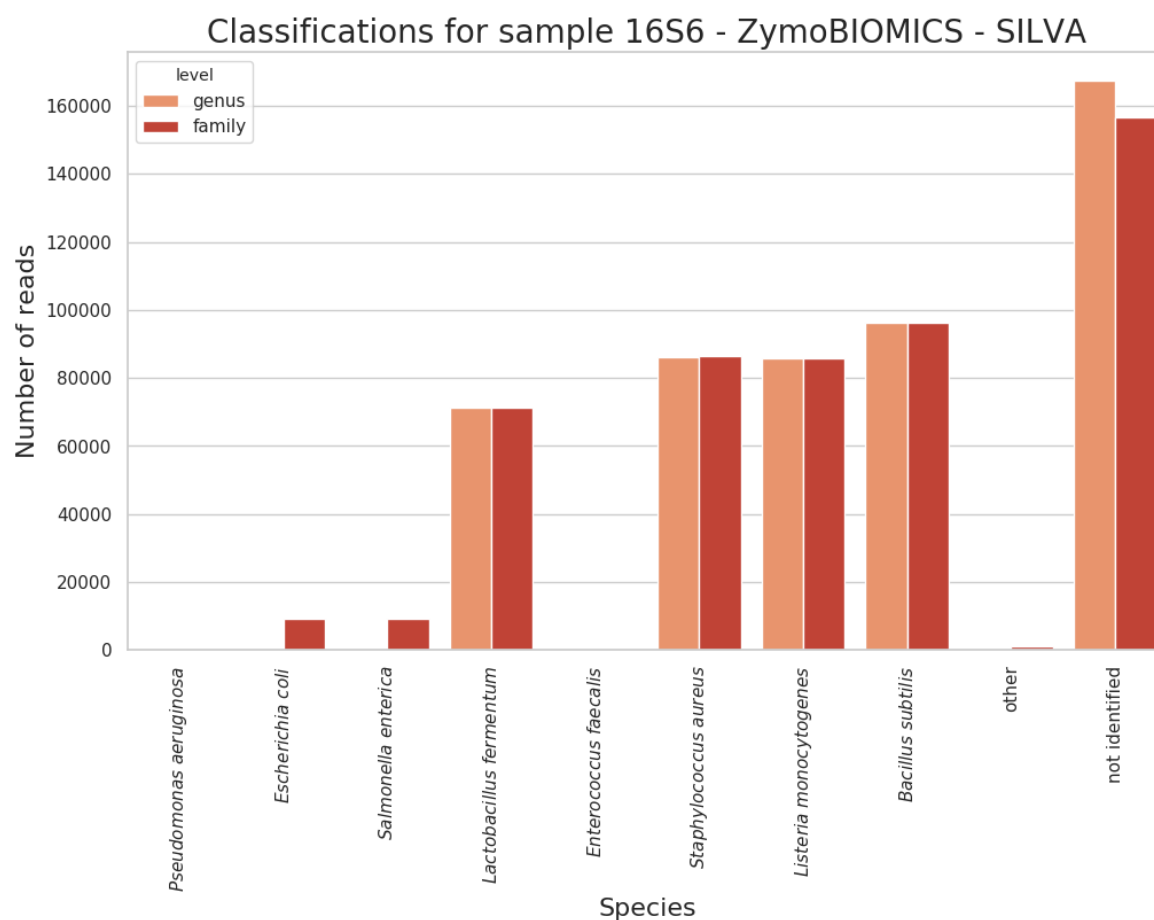


Figure S27. Overview of the number of reads that were correctly identified at the different taxonomic levels for sample 16S6 when using the SILVA database for the MiSeq data. The entry "other" considers all reads that were classified but did not match any of the bacteria present in the sample. The entry "not identified" considers all reads that were not identified at the given taxonomic level but were identified at a higher taxonomic level.

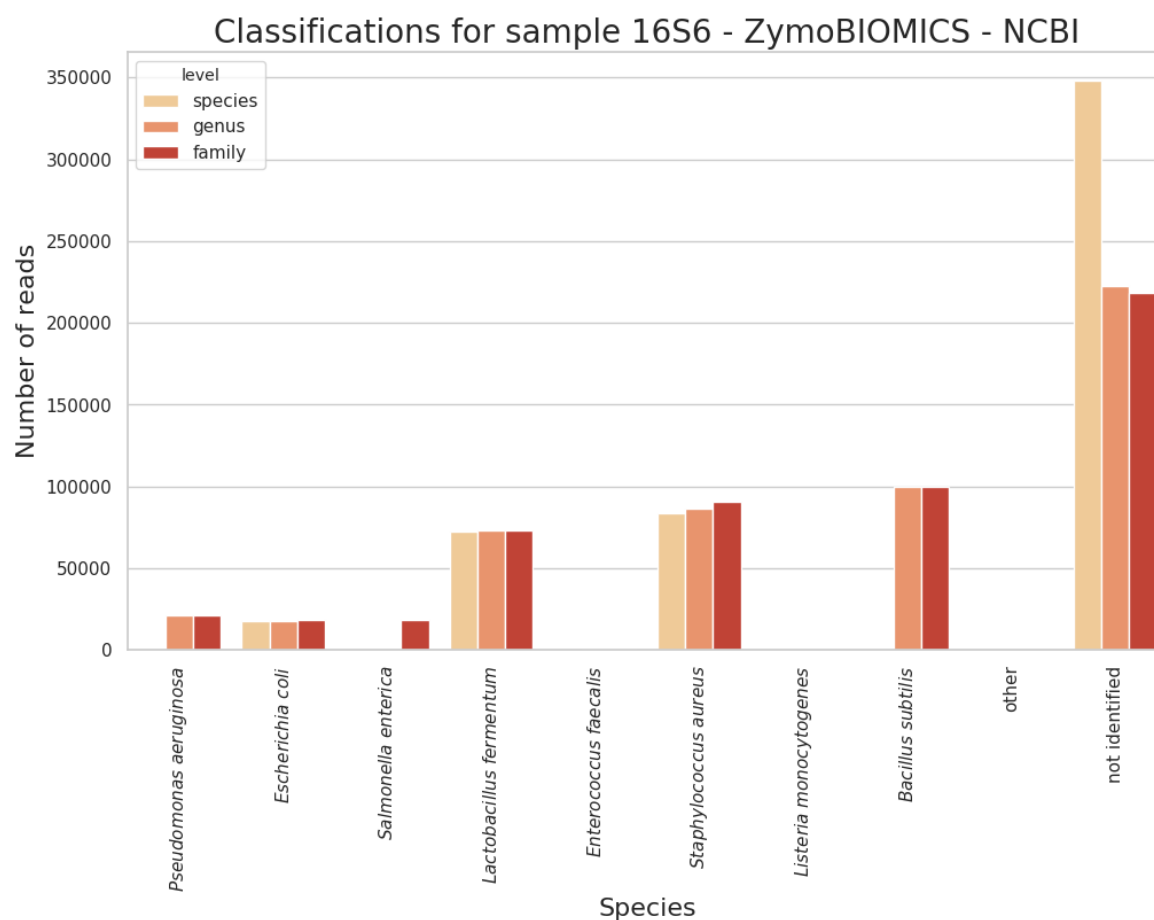


Figure S28. Overview of the number of reads that were correctly identified at the different taxonomic levels for sample 16S6 when using the NCBI 16S database for the MiSeq data. The entry "other" considers all reads that were classified but did not match any of the bacteria present in the sample. The entry "not identified" considers all reads that were not identified at the given taxonomic level but were identified at a higher taxonomic level.

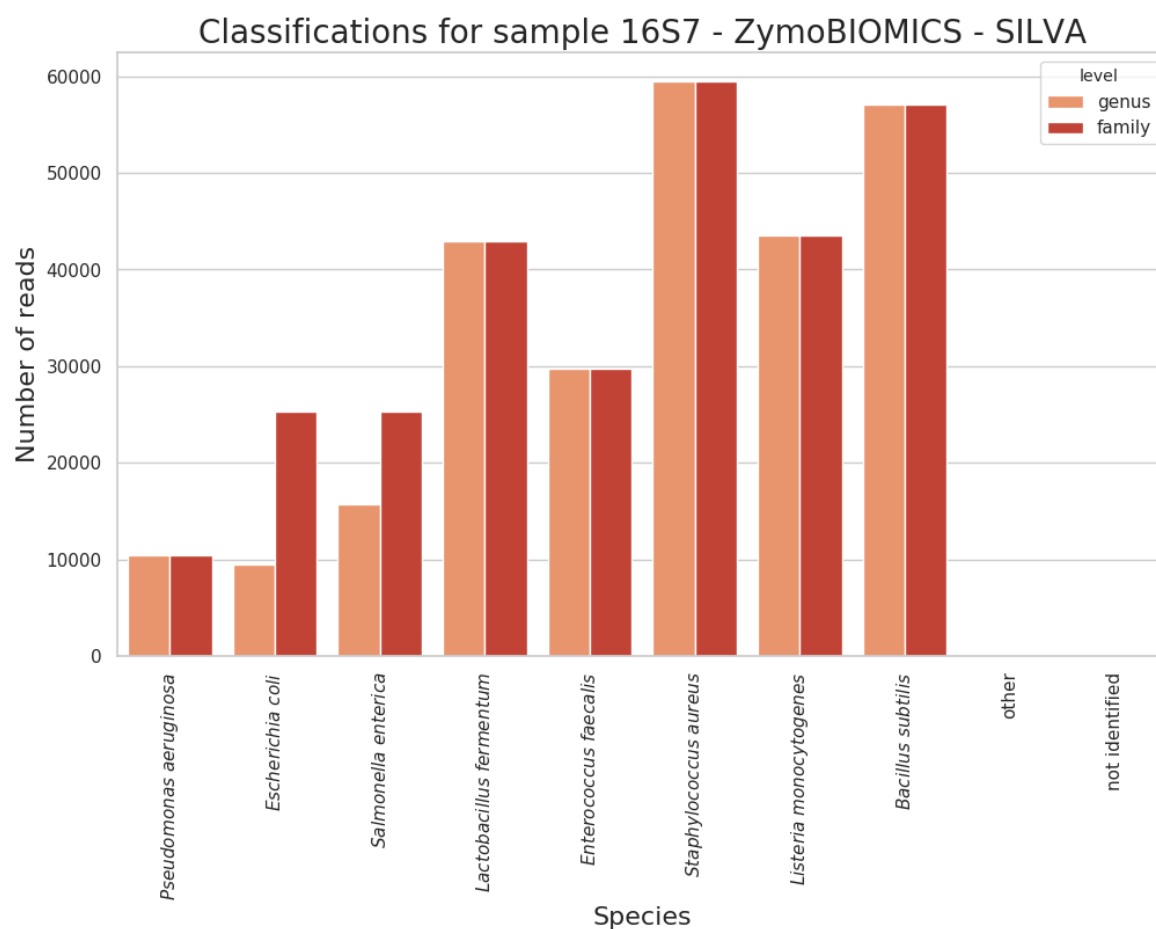


Figure S29. Overview of the number of reads that were correctly identified at the different taxonomic levels for sample 16S7 when using the SILVA database for the MiSeq data. The entry "other" considers all reads that were classified but did not match any of the bacteria present in the sample. The entry "not identified" considers all reads that were not identified at the given taxonomic level but were identified at a higher taxonomic level.

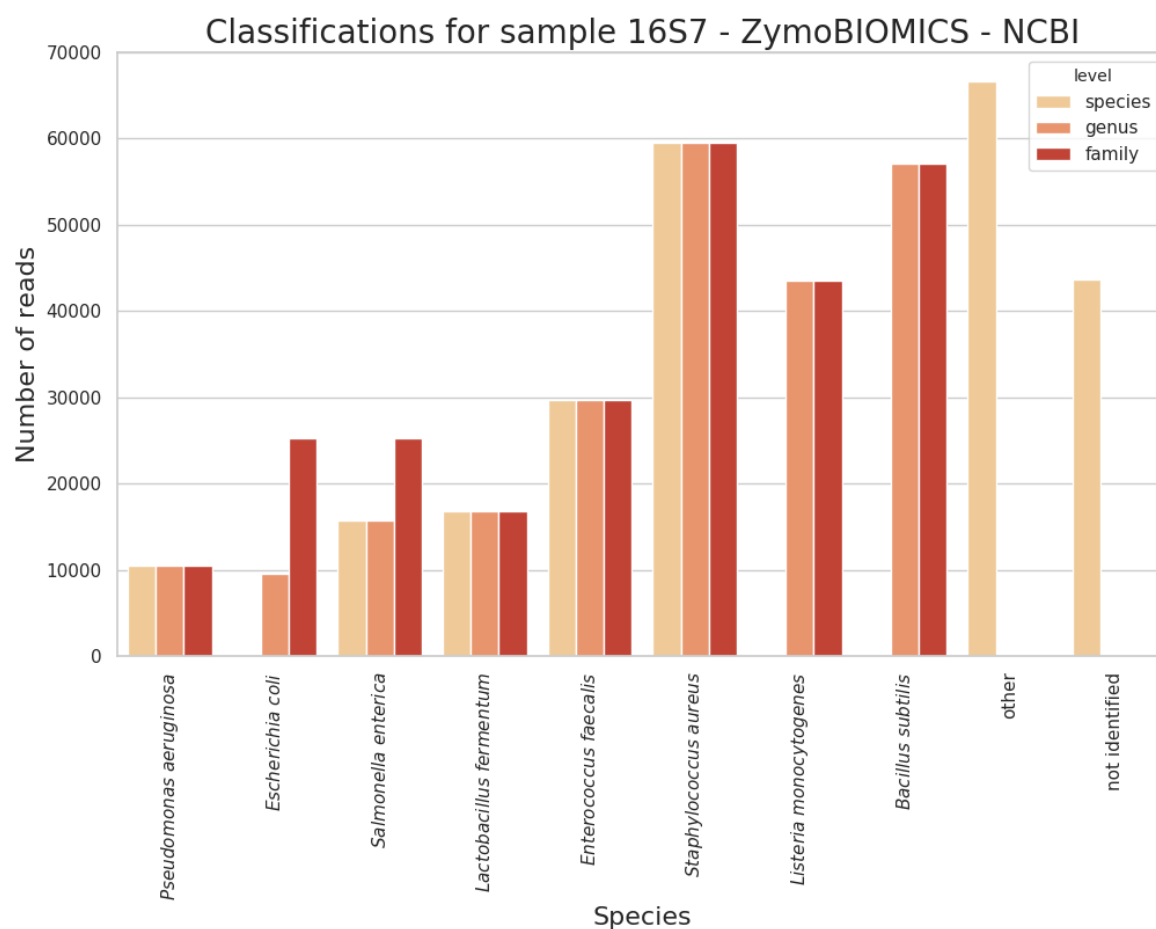


Figure S30. Overview of the number of reads that were correctly identified at the different taxonomic levels for sample 16S7 when using the NCBI 16S database for the MiSeq data. The entry "other" considers all reads that were classified but did not match any of the bacteria present in the sample. The entry "not identified" considers all reads that were not identified at the given taxonomic level but were identified at a higher taxonomic level.

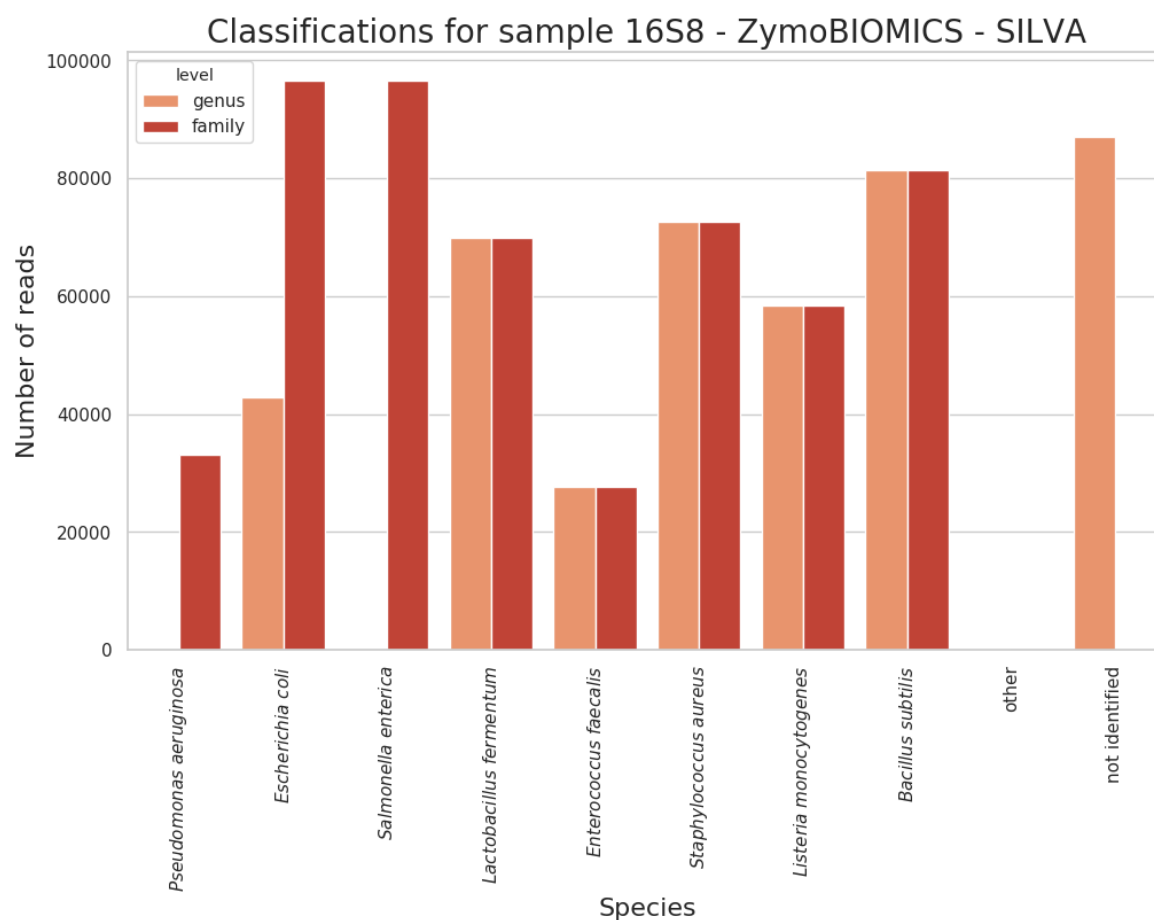


Figure S31. Overview of the number of reads that were correctly identified at the different taxonomic levels for sample 16S8 when using the SILVA database for the MiSeq data. The entry "other" considers all reads that were classified but did not match any of the bacteria present in the sample. The entry "not identified" considers all reads that were not identified at the given taxonomic level but were identified at a higher taxonomic level.

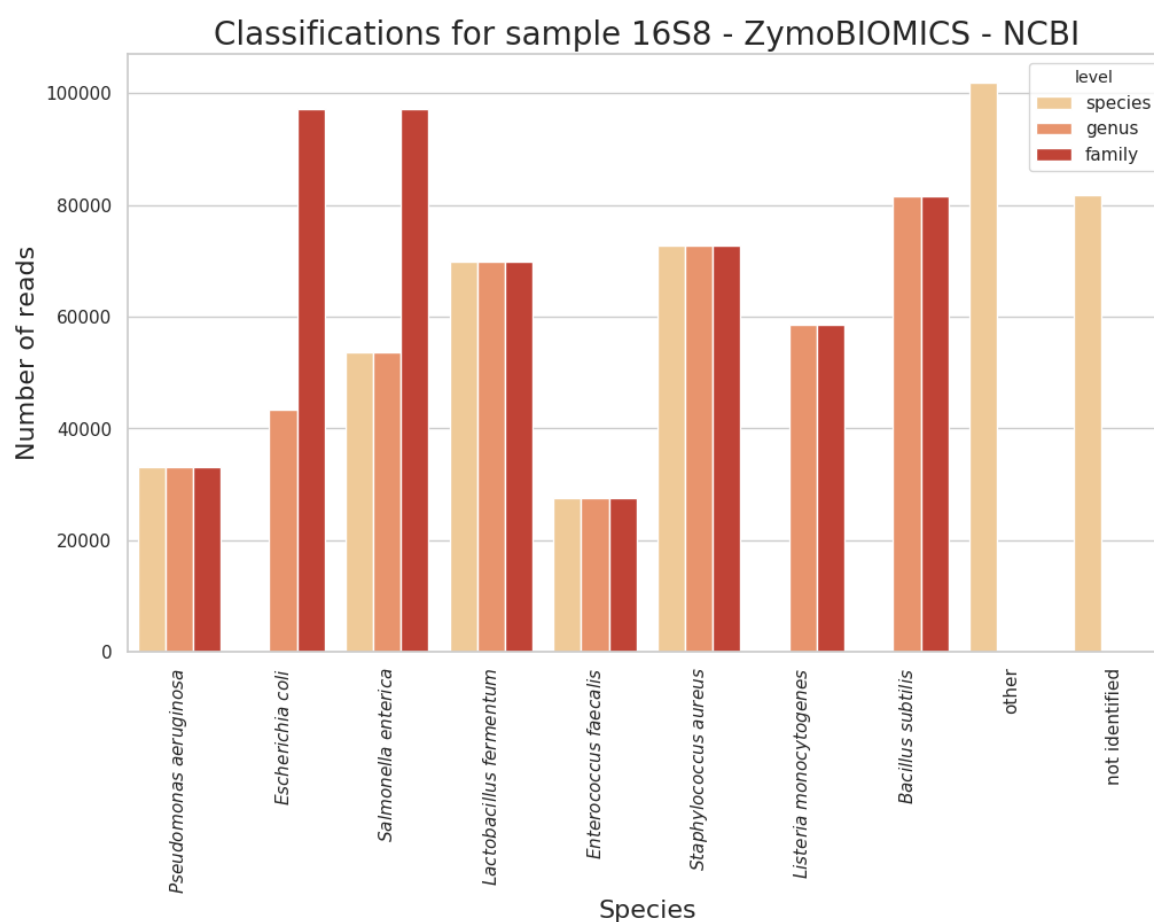


Figure S32. Overview of the number of reads that were correctly identified at the different taxonomic levels for sample 16S8 when using the NCBI 16S database for the MiSeq data. The entry "other" considers all reads that were classified but did not match any of the bacteria present in the sample. The entry "not identified" considers all reads that were not identified at the given taxonomic level but were identified at a higher taxonomic level.

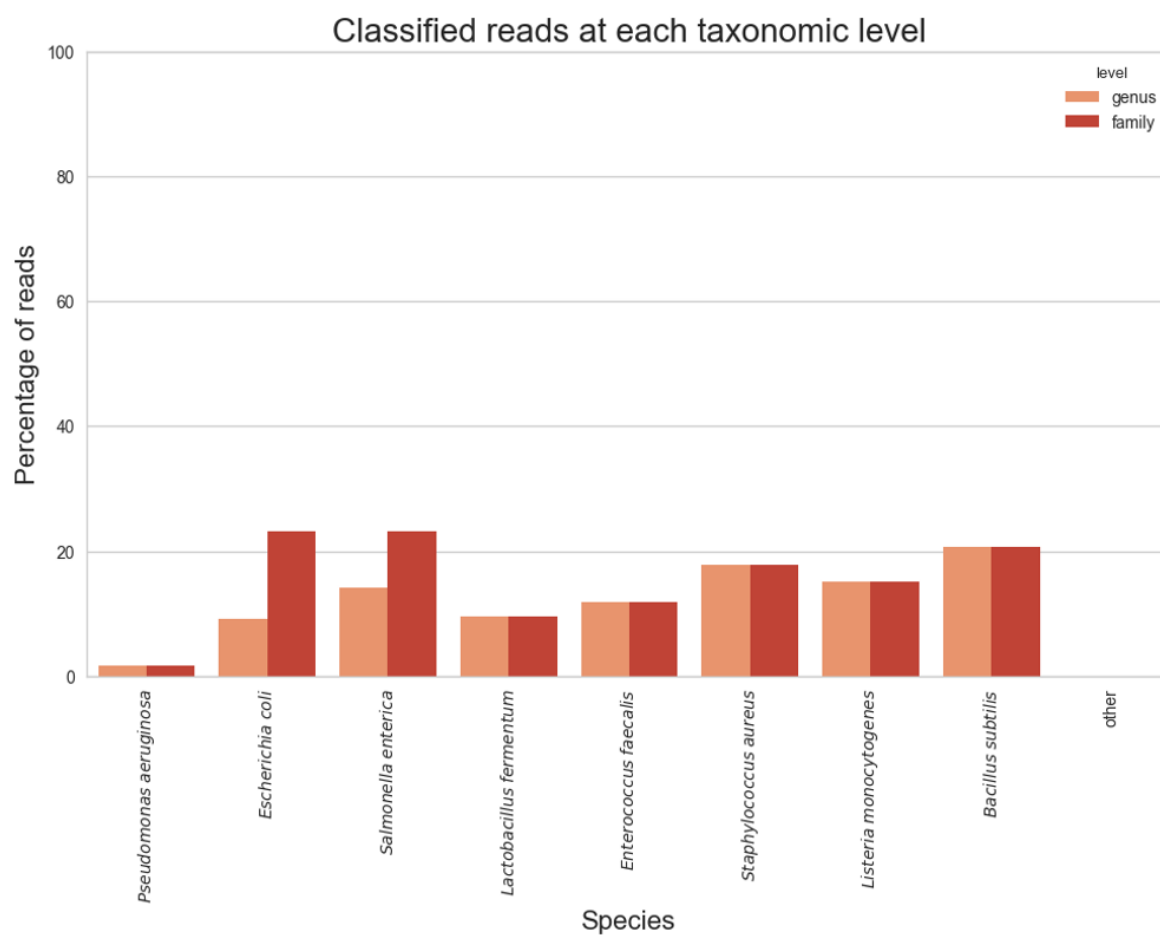


Figure S33. Overview of the percentages of reads that were correctly identified at the different taxonomic levels when using the Mothur workflow and SILVA database for the MinION data. All percentages were calculated against the number of classified reads at that level.

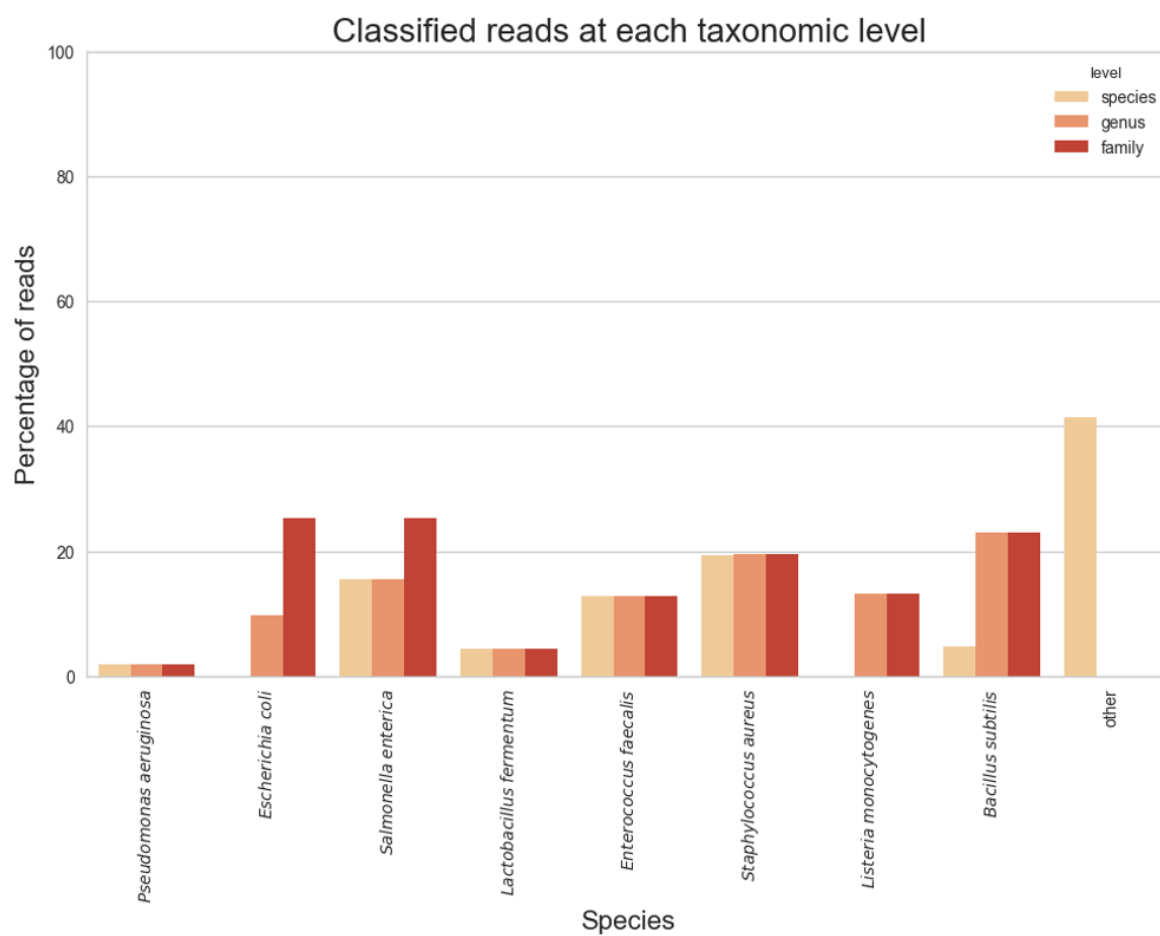


Figure S34. Overview of the percentages of reads that were correctly identified at the different taxonomic levels when using the Mothur workflow and NCBI 16S database for the MinION data. All percentages were calculated against the number of classified reads at that level.

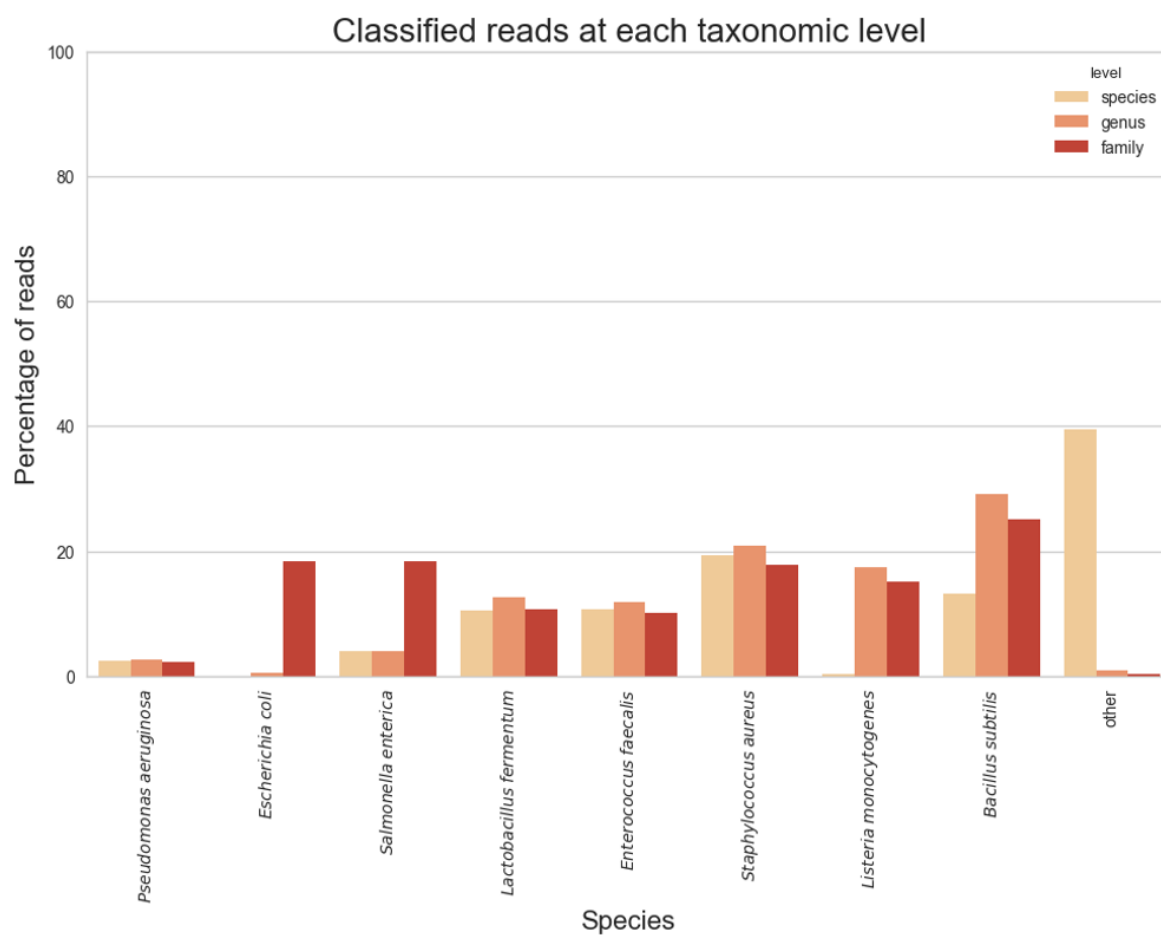


Figure S35. Overview of the percentages of reads that were correctly identified at the different taxonomic levels when using the EPI2ME workflow and NCBI 16S database for the MinION data. All percentages were calculated against the number of classified reads at that level.

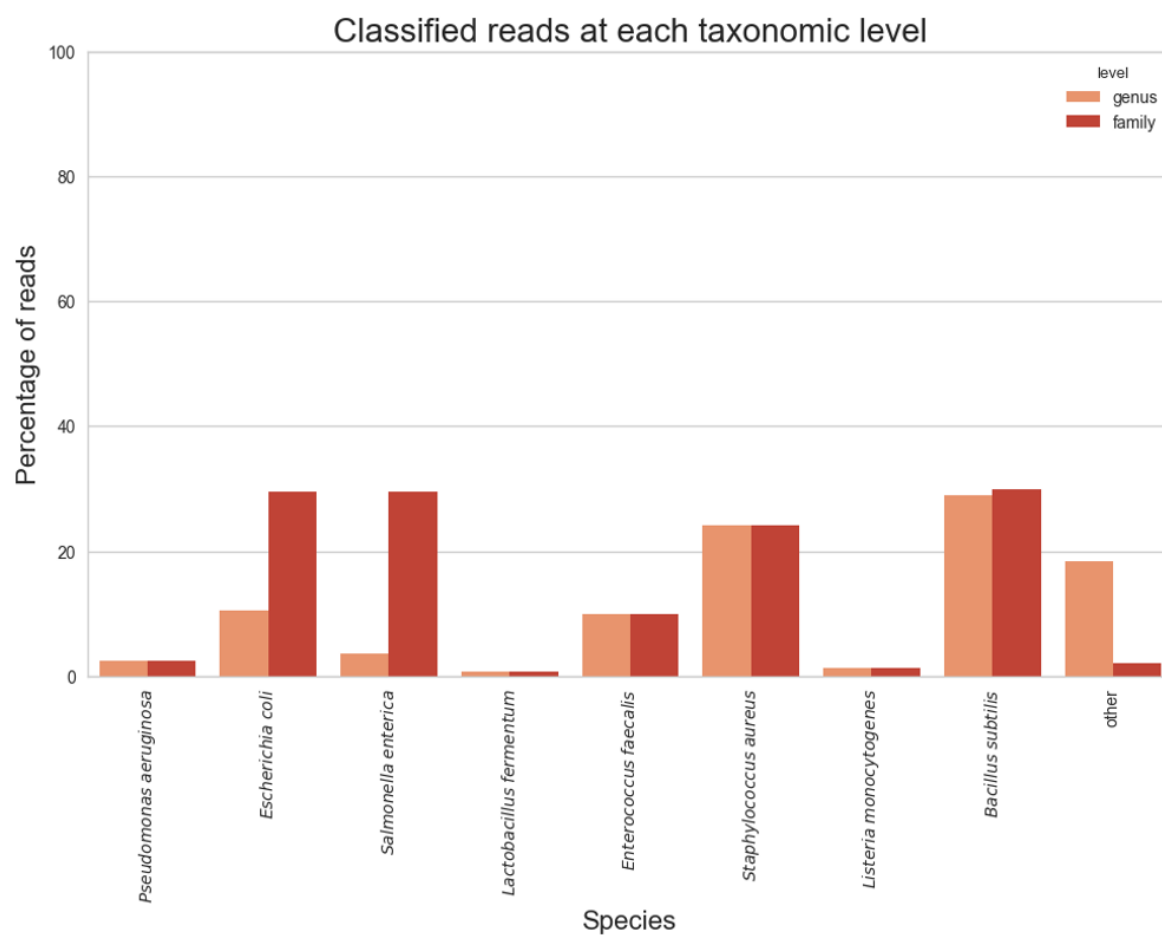


Figure S36. Overview of the percentages of reads that were correctly identified at the different taxonomic levels when using the GraphMap workflow and SILVA database for the MinION data. All percentages were calculated against the number of classified reads at that level.

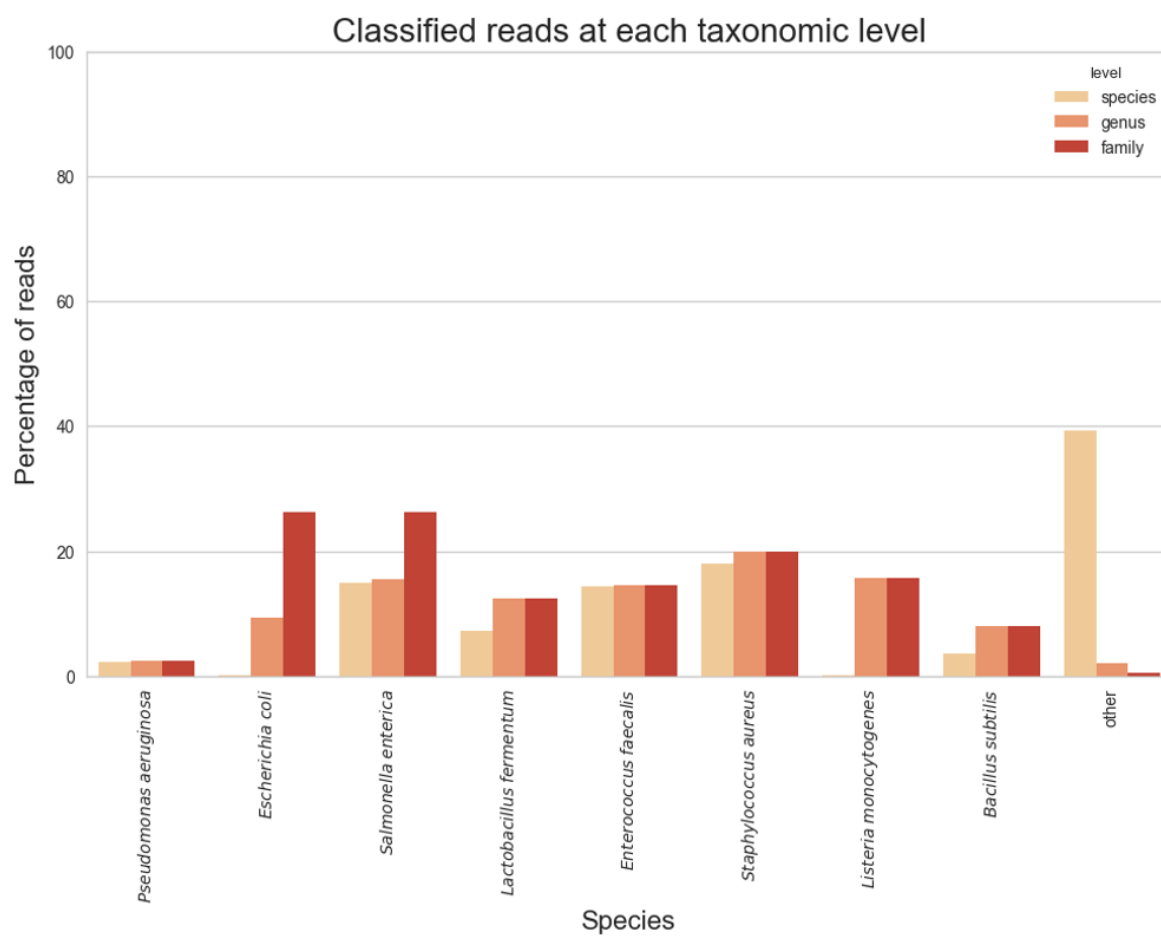


Figure S37. Overview of the percentages of reads that were correctly identified at the different taxonomic levels when using the GraphMap workflow and NCBI 16S database for the MinION data. All percentages were calculated against the number of classified reads at that level.

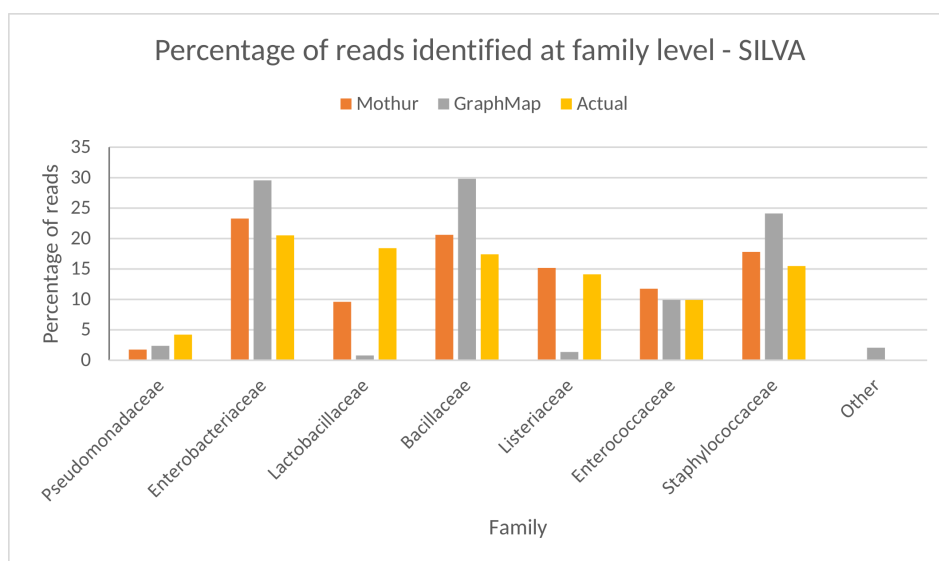


Figure S38. Overview of the percentages of reads that were correctly identified at the family level for the different families part of the mock community with the three different workflows and SILVA database for the MinION data. The "other" entry shows reads that were classified as a family not present in the mock community. The "Actual" bar shows the percentage of 16S rRNA that is present in the mock community for each species belonging that family.

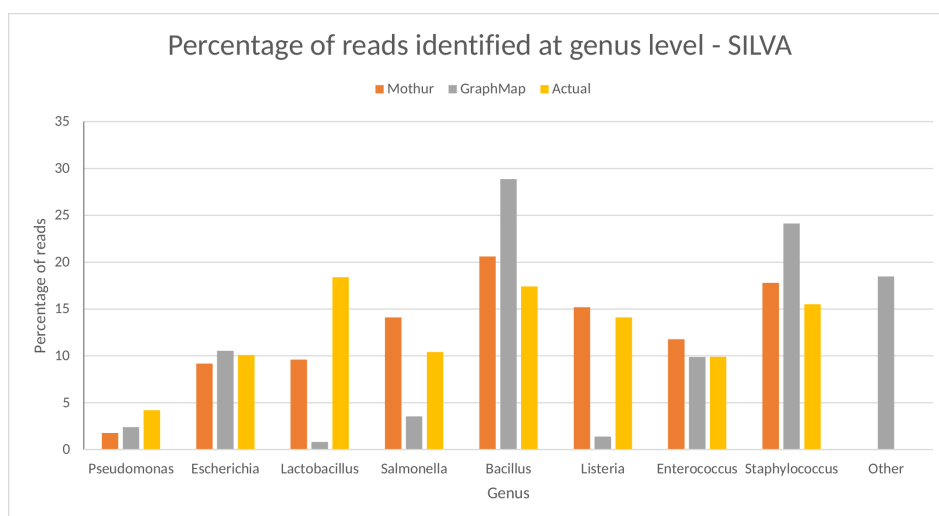


Figure S39. Overview of the percentages of reads that were correctly identified at the genus level for the different genera part of the mock community with the three different workflows and SILVA database for the MinION data. The "other" entry shows reads that were classified as a genus not present in the mock community. The "Actual" bar shows the percentage of 16S rRNA that is present in the mock community for each species belonging that genus.

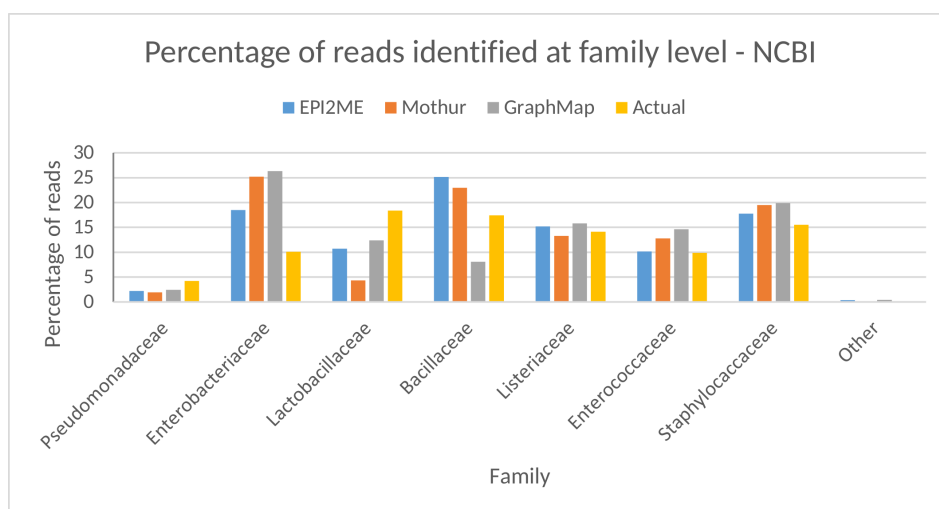


Figure S40. Overview of the percentages of reads that were correctly identified at the family level for the different families part of the mock community with the three different workflows and NCBI 16S database for the MinION data. The "other" entry shows reads that were classified as a family not present in the mock community. The "Actual" bar shows the percentage of 16S rRNA that is present in the mock community for each species belonging that family.

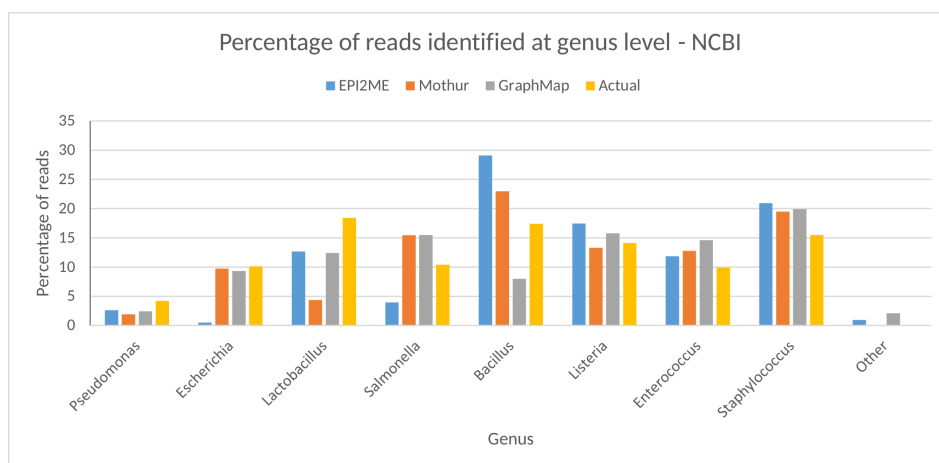


Figure S41. Overview of the percentages of reads that were correctly identified at the genus level for the different genera part of the mock community with the three different workflows and NCBI 16S database for the MinION data. The "other" entry shows reads that were classified as a genus not present in the mock community. The "Actual" bar shows the percentage of 16S rRNA that is present in the mock community for each species belonging that genus.

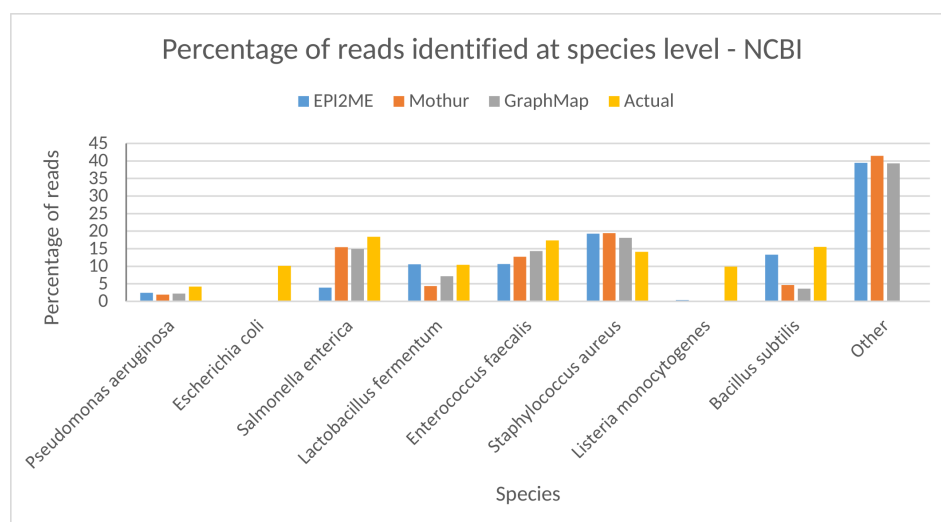


Figure S42. Overview of the percentages of reads that were correctly identified at the species level for the different species part of the mock community with the three different workflows and NCBI 16S database for the MinION data. The "other" entry shows reads that were classified as a species not present in the mock community. The "Actual" bar shows the percentage of 16S rRNA that is present in the mock community for each species.

1. Laurence, M.; Hatzis, C.; Brash, D.E. Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLoS ONE* **2014**, *9*, 1–8. doi:10.1371/journal.pone.0097876.
2. Salter, S.J.; Cox, M.J.; Turek, E.M.; Calus, S.T.; Cookson, W.O.; Moffatt, M.F.; Turner, P.; Parkhill, J.; Loman, N.J.; Walker, A.W. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology* **2014**, *12*, 87. doi:10.1186/s12915-014-0087-z.
3. Weyrich, L.; Farrer, A.G.; Eisenhofer, R.; Arriola, L.A.; Young, J.; Selway, C.A.; Handsley-Davis, M.; Adler, C.; Breen, J.; Cooper, A. Laboratory contamination over time during low-biomass sample analysis [Preprint]. *bioRxiv* **2018**.