

Supplemental Material for

Design and Evaluation of Illumina MiSeq Compatible Primers for the 18S rRNA Gene for Improved Characterization of Mixed Phototrophic Communities

Ian M. Bradley, Ameet J. Pinto, Jeremy S. Guest

**Table S1:** Algal species included in the construction of mock communities used in this study.

Classification	Description	Genus/Species	order #	Axenic (Y/N)	GC content (V4)	GC content (V8-V9)
<b>Bacillariophyceae</b>	Diatoms (Heterokonts)	<i>Thalassiosira pseudonana</i> (Hustedt) Hasle & Heimdal	1085/12 (CCAP)	Y	42.9%	46.0%
<b>Chlorophyta</b>	Green algae	<i>Chlorella vulgaris</i> ( Beijerinck)	211/11B (CCAP)	Y	49.1%	50.0%
		<i>Scenedesmus obliquus</i> (Turpin) Kützing	276/3A (CCAP)	Y	46.1%	48.7%
		<i>Trebouxia</i> sp.	213/3 (CCAP)		49.9%	50.0%
<b>Cryptophyta</b>	Cryptophytes	<i>Cryptomonas pyrenoidifera</i> Geitler emend. Hoef-Emden & Melkonian	979/61 (CCAP)	N, bacteria present	46.8%	47.0%
		<i>Rhodomonas</i> sp.	995/5 (CCAP)	N, bacteria present	45.0%	45.0%
<b>Dinoflagellata</b>	Dinoflagellates	<i>Heterocapsa niei</i> W. Blankley	1125/4 (CCAP), UTEX LB 2722	N, bacteria present	44.3%	46.0%
		<i>Symbiodinium microadriaticum</i> (UTEX name: <i>Zooxathella microadriatica</i> ) H. Ducklow, A.R. Loeblich III	UTEX LB 2281		44.8%	46.0%
<b>Haptophyta</b>	Haptophytes	<i>Prymnesium parvum</i> N. Carter	946/6 (CCAP)	Y	50.8%	49.0%
		<i>Isochrysis galbana</i> Parke	927/1 (CCAP)	Y	52.0%	50.0%
<b>Chrysophyceae</b>	Golden Algae /Heterokonts	<i>Ochromonas</i> sp. J.R. Maudsley	UTEX LB 2575	N	42.8%	46.0%
<b>Eustigmatophyceae</b>	Heterokonts	<i>Nannochloropsis oculata</i> (Droop) Hibberd	CCMP 525 (NCMA)		42.6%	46.0%

**Table S2:** Environmental samples evaluated using V4 and V8-V9 based amplicon sequencing of the 18S rRNA gene.

Environment	Sample ID Name	Description	Location
Freshwater	4F	marsh	Bray's Bayou, TX
	5F	marsh	Bray's Bayou, TX
	6F	pond	Willow Water Hole, TX
Wastewater	3W	primary effluent	Lawrence, KS WWTP
	4W	primary effluent	Algaewheel, Charleston IL WWTP
	5W	primary effluent	ISTC, IL
	6W	primary effluent	Algaewheel, UIUC
	7W	primary effluent	Raceway pond, UIUC
	8W	primary effluent	Urbana IL WWTP
	9W	secondary effluent	Urbana IL WWTP
	10W	secondary effluent	Urbana IL WWTP
Marine	1M	ocean	Revere Beach, Broadsound, MA
	2M	creek/ocean	San Juan Creek, Pacific
	4M	ocean	Dauphine Island, Lagoon
	5M	ocean	Dauphine Island, Bay
Control	NEG		
Mock Community	M11	mock community	Freshwater:Marine 100:0
	M12	mock community	100:0
	M13	mock community	100:0
	M21	mock community	100:1
	M22	mock community	100:1
	M23	mock community	100:1
	M31	mock community	100:10
	M32	mock community	100:10
	M33	mock community	100:10
	M41	mock community	1:1
	M42	mock community	1:1
	M43	mock community	1:1
	M51	mock community	10:100
	M52	mock community	10:100
	M53	mock community	10:100
	M61	mock community	1:100
	M62	mock community	1:100
	M63	mock community	1:100
	M71	mock community	0:100
	M72	mock community	0:100
M73	mock community	0:100	

**Table S3:** EukA (forward) and EukB (reverser) primers used for amplifying full-length 18S rRNA gene copies for cloning and mock community construction.

Primer #	Forward primer (5'-3')	Reverse primer (5'-3')
EukA	AACCTGGTTGATCCTGCCAGT	
EukB		TGATCCTTCTGCAGGTTACCTAC

**Table S4:** Forward and reverse dual-index primers (with heterogeneity spacers) used for the V4 region of the 18S rRNA gene.

ID	5' adapter	i5 index	Heterogeneity Spacer	pad	link	gene specific primer (V4F)
SA501	AATGATACGGCGACC ACCGAGATCTACAC	ATCGTACG		TATGGTAATT	GT	CCAGCASCYGCGGTA ATTCC
SA502	AATGATACGGCGACC ACCGAGATCTACAC	ACTATCTG	A	TATGGTAATT	GT	CCAGCASCYGCGGTA ATTCC
SA503	AATGATACGGCGACC ACCGAGATCTACAC	TAGCGAGT	GC	TATGGTAATT	GT	CCAGCASCYGCGGTA ATTCC
SA504	AATGATACGGCGACC ACCGAGATCTACAC	CTGCGTGT	CGC	TATGGTAATT	GT	CCAGCASCYGCGGTA ATTCC
SA505	AATGATACGGCGACC ACCGAGATCTACAC	TCATCGAG	ATGA	TATGGTAATT	GT	CCAGCASCYGCGGTA ATTCC
SA506	AATGATACGGCGACC ACCGAGATCTACAC	CGTGAGTG	TGCCA	TATGGTAATT	GT	CCAGCASCYGCGGTA ATTCC
SA507	AATGATACGGCGACC ACCGAGATCTACAC	GGATATCT	GAGCC	TATGGTAATT	GT	CCAGCASCYGCGGTA ATTCC
SA508	AATGATACGGCGACC ACCGAGATCTACAC	GACACCGT	CCAGCAC	TATGGTAATT	GT	CCAGCASCYGCGGTA ATTCC

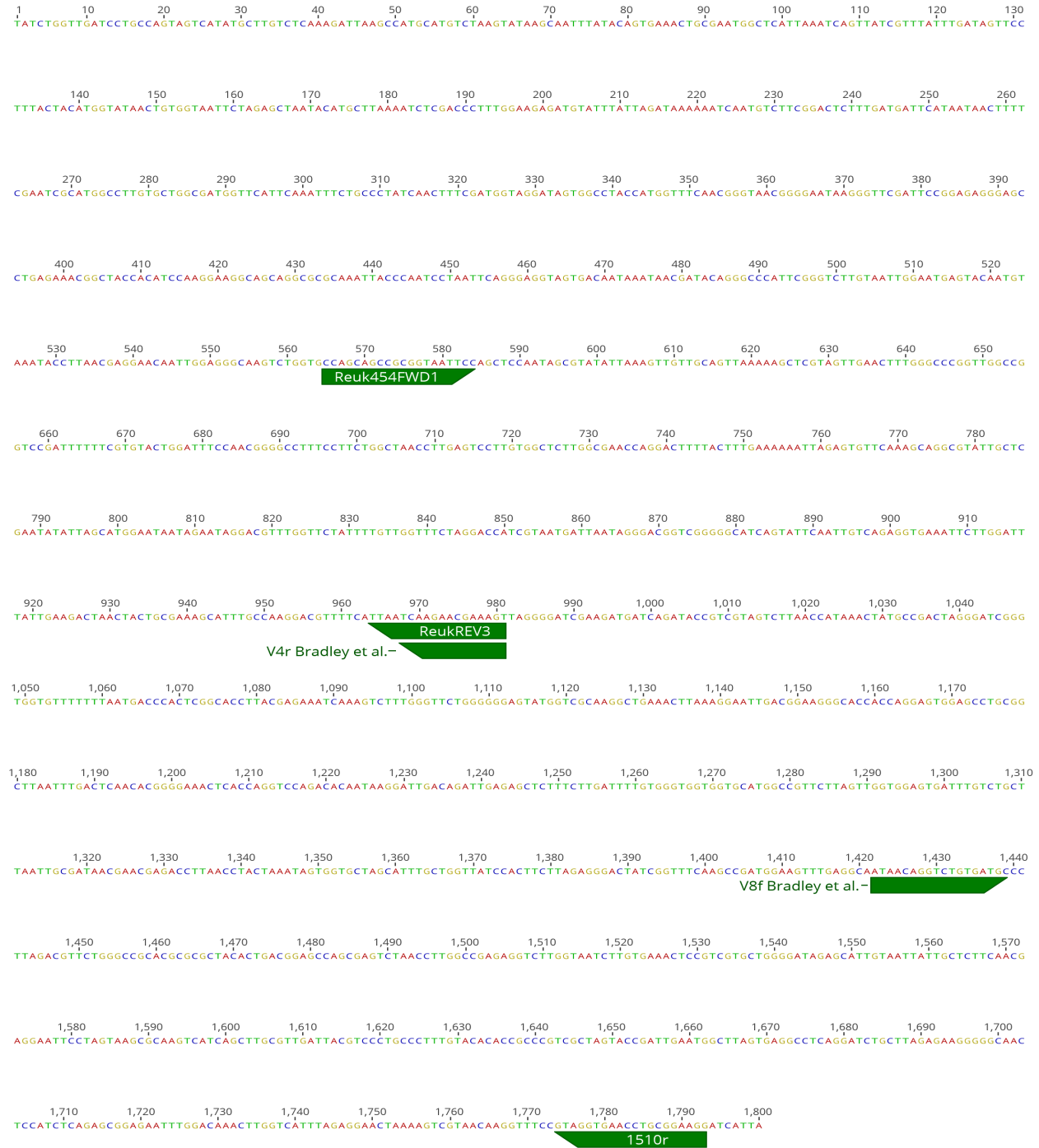
ID	3' adapter	i7 index	Heterogeneity Spacer	pad/linker		gene specific (V4R)
SA701	CAAGCAGAAGACGGC ATACGAGAT	AACTCTCG		AGTCAGTCAG	CC	ACTTTCGTTCTTGAT
SA702	CAAGCAGAAGACGGC ATACGAGAT	ACTATGTC	T	AGTCAGTCAG	CC	ACTTTCGTTCTTGAT
SA703	CAAGCAGAAGACGGC ATACGAGAT	AGTAGCGT	GT	AGTCAGTCAG	CC	ACTTTCGTTCTTGAT
SA704	CAAGCAGAAGACGGC ATACGAGAT	CAGTGAGT	CCA	AGTCAGTCAG	CC	ACTTTCGTTCTTGAT
SA705	CAAGCAGAAGACGGC ATACGAGAT	CGTACTCA	ATCA	AGTCAGTCAG	CC	ACTTTCGTTCTTGAT
SA706	CAAGCAGAAGACGGC ATACGAGAT	CTACGCAG	TGCCT	AGTCAGTCAG	CC	ACTTTCGTTCTTGAT
SA707	CAAGCAGAAGACGGC ATACGAGAT	GGAGACTA	GACTGC	AGTCAGTCAG	CC	ACTTTCGTTCTTGAT
SA708	CAAGCAGAAGACGGC ATACGAGAT	GTCGCTCG	CCTGCTC	AGTCAGTCAG	CC	ACTTTCGTTCTTGAT
SA709	CAAGCAGAAGACGGC ATACGAGAT	GTCGTAGT		AGTCAGTCAG	CC	ACTTTCGTTCTTGAT
SA710	CAAGCAGAAGACGGC ATACGAGAT	TAGCAGAC	T	AGTCAGTCAG	CC	ACTTTCGTTCTTGAT
SA711	CAAGCAGAAGACGGC ATACGAGAT	TCATAGAC	GT	AGTCAGTCAG	CC	ACTTTCGTTCTTGAT
SA712	CAAGCAGAAGACGGC ATACGAGAT	TCGCTATA	CCG	AGTCAGTCAG	CC	ACTTTCGTTCTTGAT

**Table S5:** Forward and reverse dual-index primers (with heterogeneity spacers) used for the V8-V9 region of the 18S rRNA gene.

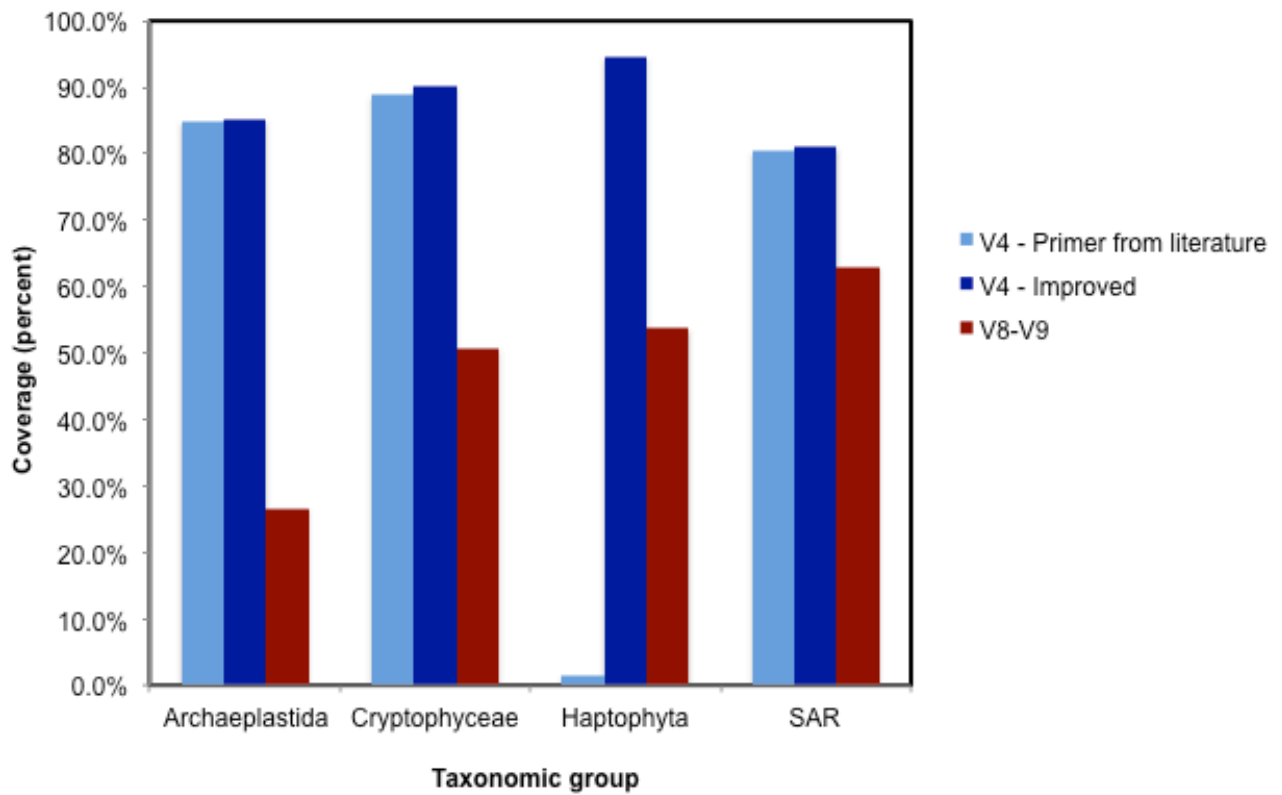
	5' adapter	i5 index	Heterogeneity Spacer	pad	link	gene specific primer (V8F)
SB501	AATGATACGGCGACC ACCGAGATCTACAC	CTACTATA		CTACGTACAG	CC	ATAACAGGTCTGTGAT GCCCT
SB502	AATGATACGGCGACC ACCGAGATCTACAC	CGTTACTA	T	CTACGTACAG	CC	ATAACAGGTCTGTGAT GCCCT
SB503	AATGATACGGCGACC ACCGAGATCTACAC	AGAGTCAC	GA	CTACGTACAG	CC	ATAACAGGTCTGTGAT GCCCT
SB504	AATGATACGGCGACC ACCGAGATCTACAC	TACGAGAC	CGA	CTACGTACAG	CC	ATAACAGGTCTGTGAT GCCCT
SB505	AATGATACGGCGACC ACCGAGATCTACAC	ACGTCTCG	ATGA	CTACGTACAG	CC	ATAACAGGTCTGTGAT GCCCT
SB506	AATGATACGGCGACC ACCGAGATCTACAC	TCGACGAG	TGCGA	CTACGTACAG	CC	ATAACAGGTCTGTGAT GCCCT
SB507	AATGATACGGCGACC ACCGAGATCTACAC	GATCGTGT	GAGTGA	CTACGTACAG	CC	ATAACAGGTCTGTGAT GCCCT
SB508	AATGATACGGCGACC ACCGAGATCTACAC	GTCAGATA	ACTGTGG	CTACGTACAG	CC	ATAACAGGTCTGTGAT GCCCT

	3' adapter	i7 index	Heterogeneity Spacer	pad/linker	link	gene specific (V9R)
SB701	CAAGCAGAAGACGGC ATACGAGAT	AAGTCGAG		GCATGTGCAT	GG	CCTTCYGCAGGTTCA CCTAC
SB702	CAAGCAGAAGACGGC ATACGAGAT	ATACTTCG	T	GCATGTACGT	GG	CCTTCYGCAGGTTCA CCTAC
SB703	CAAGCAGAAGACGGC ATACGAGAT	AGCTGCTA	GT	GCATGTACGT	GG	CCTTCYGCAGGTTCA CCTAC
SB704	CAAGCAGAAGACGGC ATACGAGAT	CATAGAGA	CAA	GCATGTACGT	GG	CCTTCYGCAGGTTCA CCTAC
SB705	CAAGCAGAAGACGGC ATACGAGAT	CGTAGATC	ATCC	GCATGTACGT	GG	CCTTCYGCAGGTTCA CCTAC
SB706	CAAGCAGAAGACGGC ATACGAGAT	CTCGTTAC	TATGA	GCATGTACGT	GG	CCTTCYGCAGGTTCA CCTAC
SB707	CAAGCAGAAGACGGC ATACGAGAT	GCGCACGT	CTGTCC	GCATGTACGT	GG	CCTTCYGCAGGTTCA CCTAC
SB708	CAAGCAGAAGACGGC ATACGAGAT	GGTACTAT	ACTATCC	GCATGTACGT	GG	CCTTCYGCAGGTTCA CCTAC
SB709	CAAGCAGAAGACGGC ATACGAGAT	GTATACGC		GCATGTACGT	GG	CCTTCYGCAGGTTCA CCTAC
SB710	CAAGCAGAAGACGGC ATACGAGAT	TACGAGCA	T	GCATGTACGT	GG	CCTTCYGCAGGTTCA CCTAC
SB711	CAAGCAGAAGACGGC ATACGAGAT	TCAGCGTT	CA	GCATGTACGT	GG	CCTTCYGCAGGTTCA CCTAC
SB712	CAAGCAGAAGACGGC ATACGAGAT	TCGCTACG	AGT	GCATGTACGT	GG	CCTTCYGCAGGTTCA CCTAC

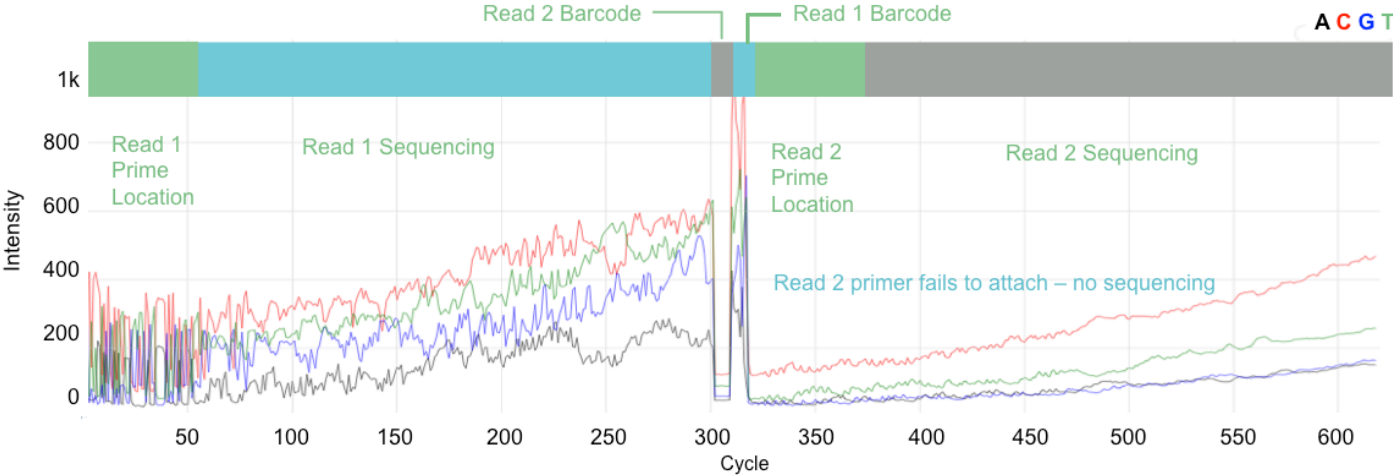
**Figure S1:** Priming locations for the 5 primers used in this study, aligned to the 18S rRNA gene from *S. cerevisiae*.



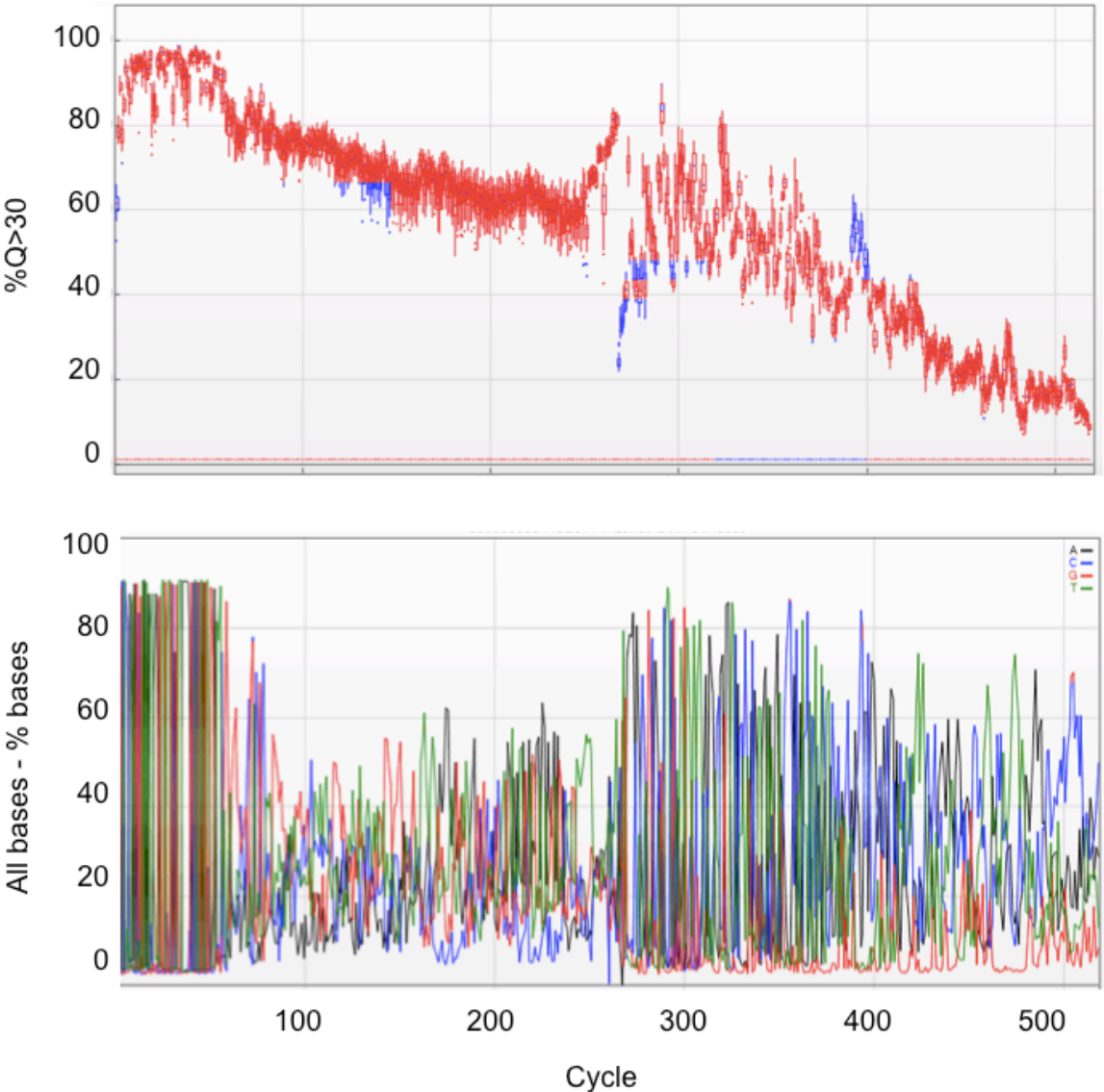
**Figure S2:** SILVA TestPrime results from *in silico* PCR using the original V4 primer set, modified V4 primer set, and the V8-V9 set. The low coverage for the V8-V9 region is due to a large portion of the database missing the V9 primer region.



**Figure S3:** Fluorescent intensities per cycle for the four nucleotides during the first MiSeq run.

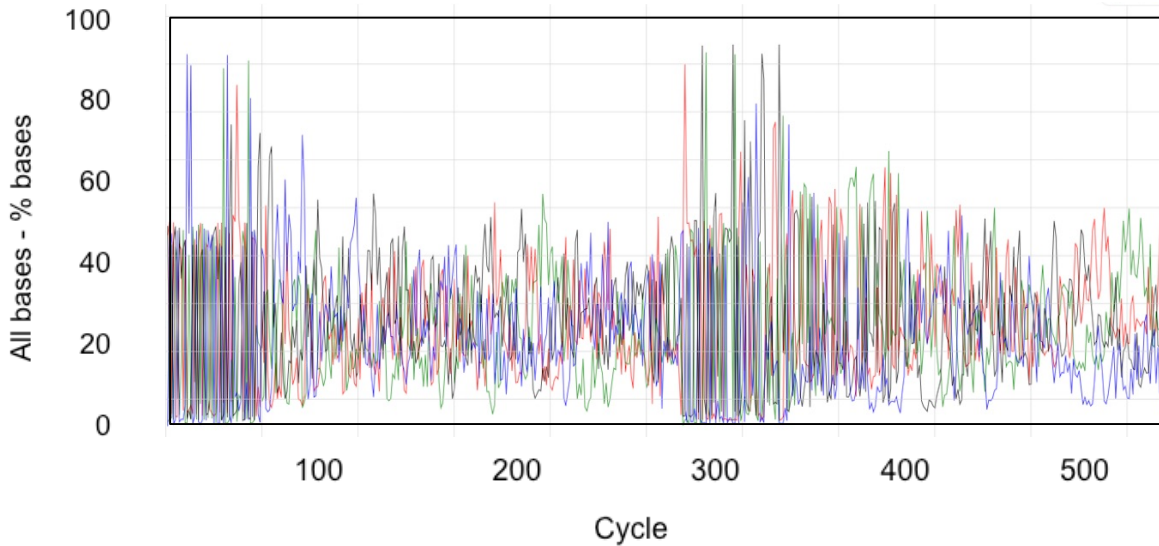


**Figure S4:** Proportion of reads with greater than phred score 30 (top) and proportion of the four nucleotides detected (bottom) per cycle during the second sequencing run.

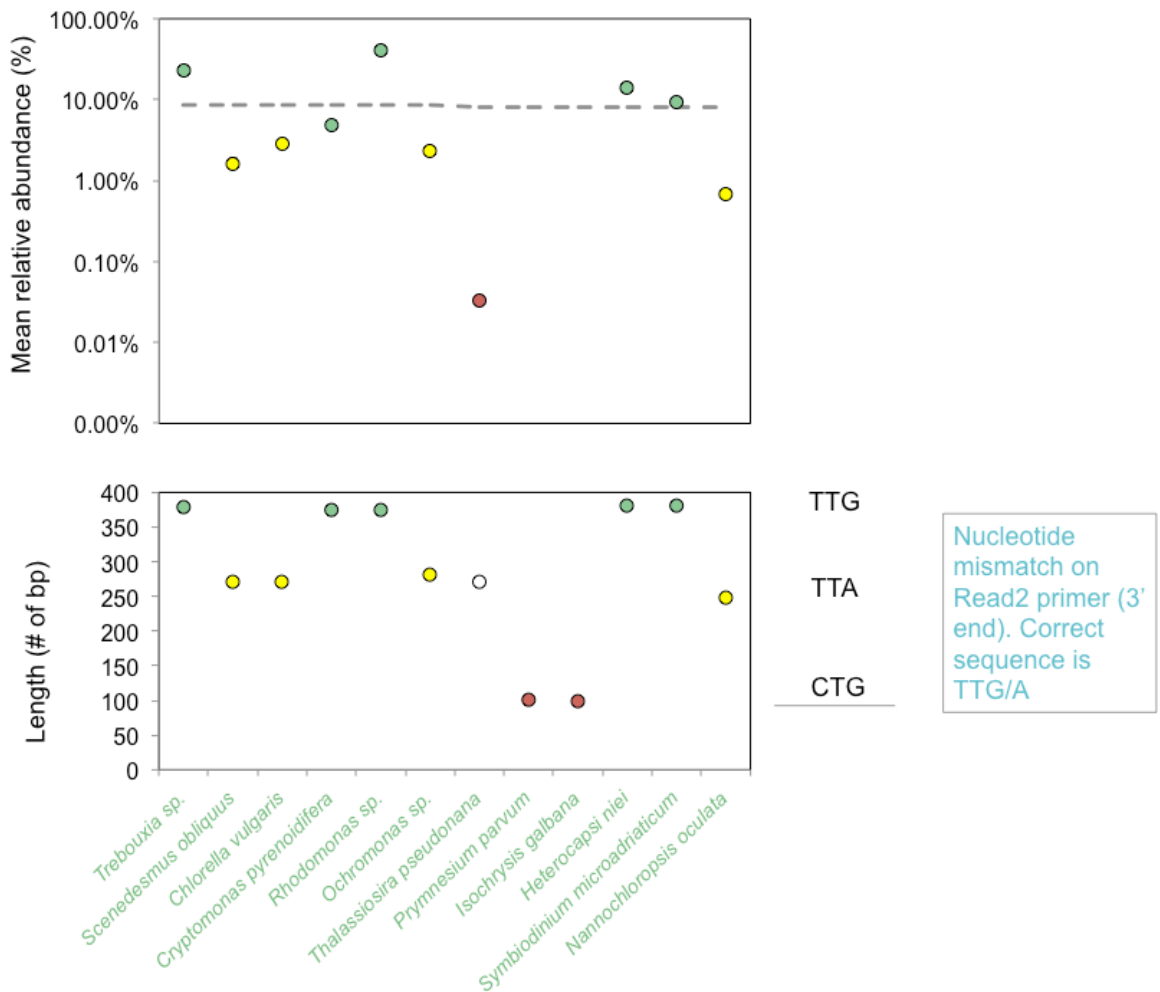




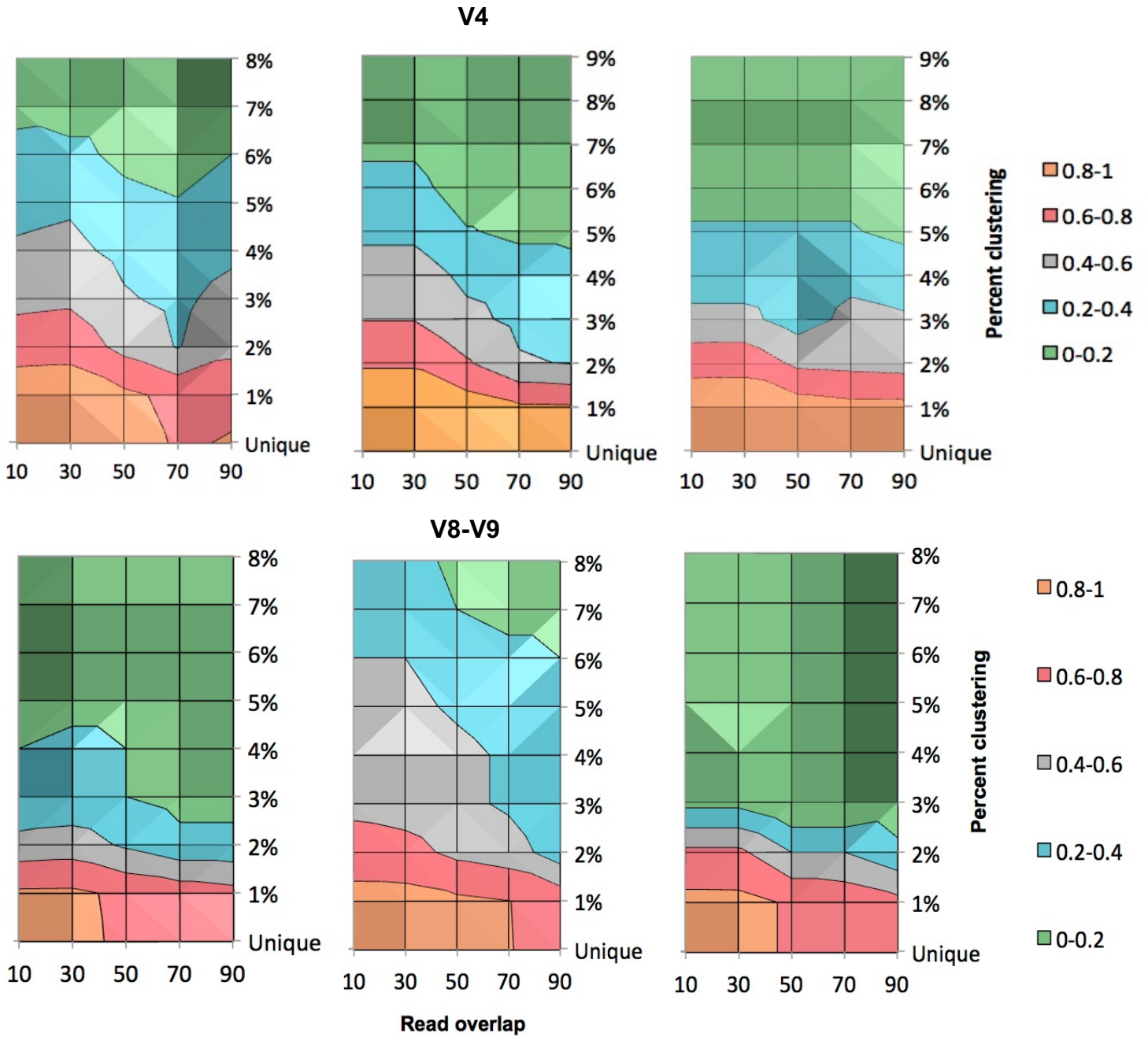
**Figure S5:** The distribution of the four nucleotides per sequencing run was greatly improved with the inclusion of heterogeneity spacers for the third sequencing run.



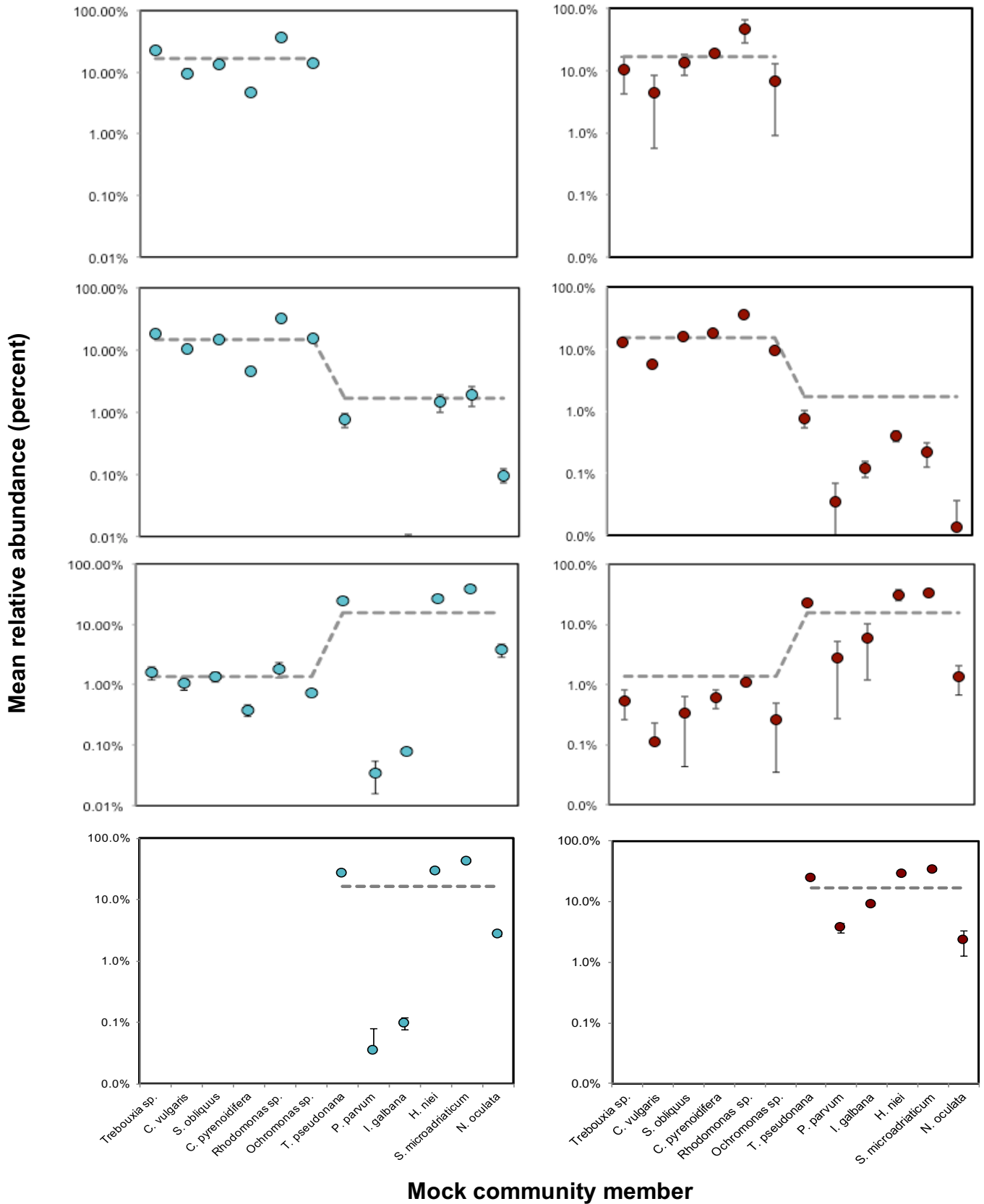
**Figure S6:** Mean relative abundance of mock community obtained from sequencing run 2 (V4 primers – Reuk454FWD1 and ReukRev3) for the even community, MC4.



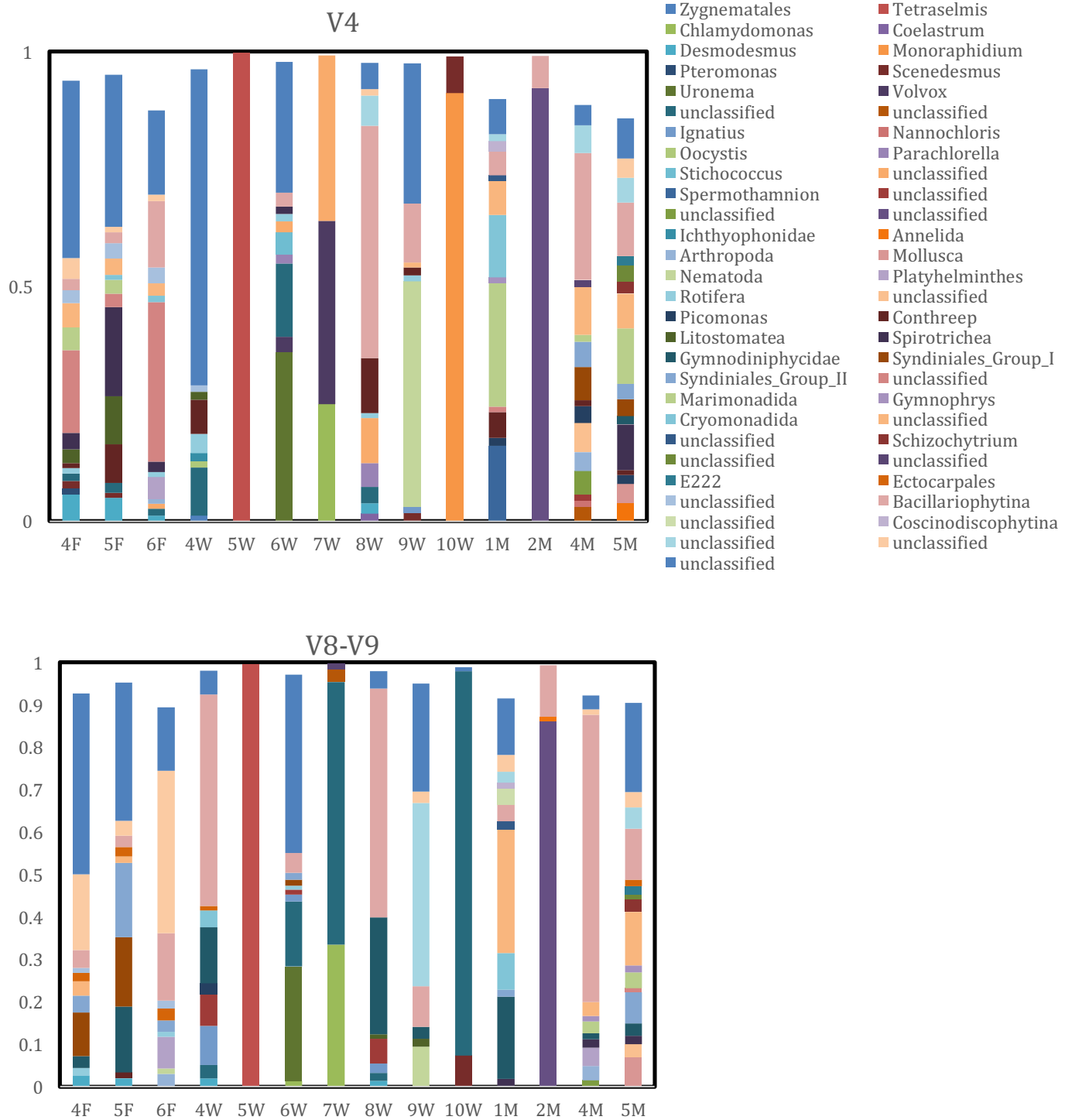
**Figure S7:** Read overlap versus similarity cutoff for V4 (top) and V8-V9 (bottom) primer sets and the effect on Jaccard distance (dissimilarity). Each set shows 1 of 3 replicates for the even mock community (MC4).



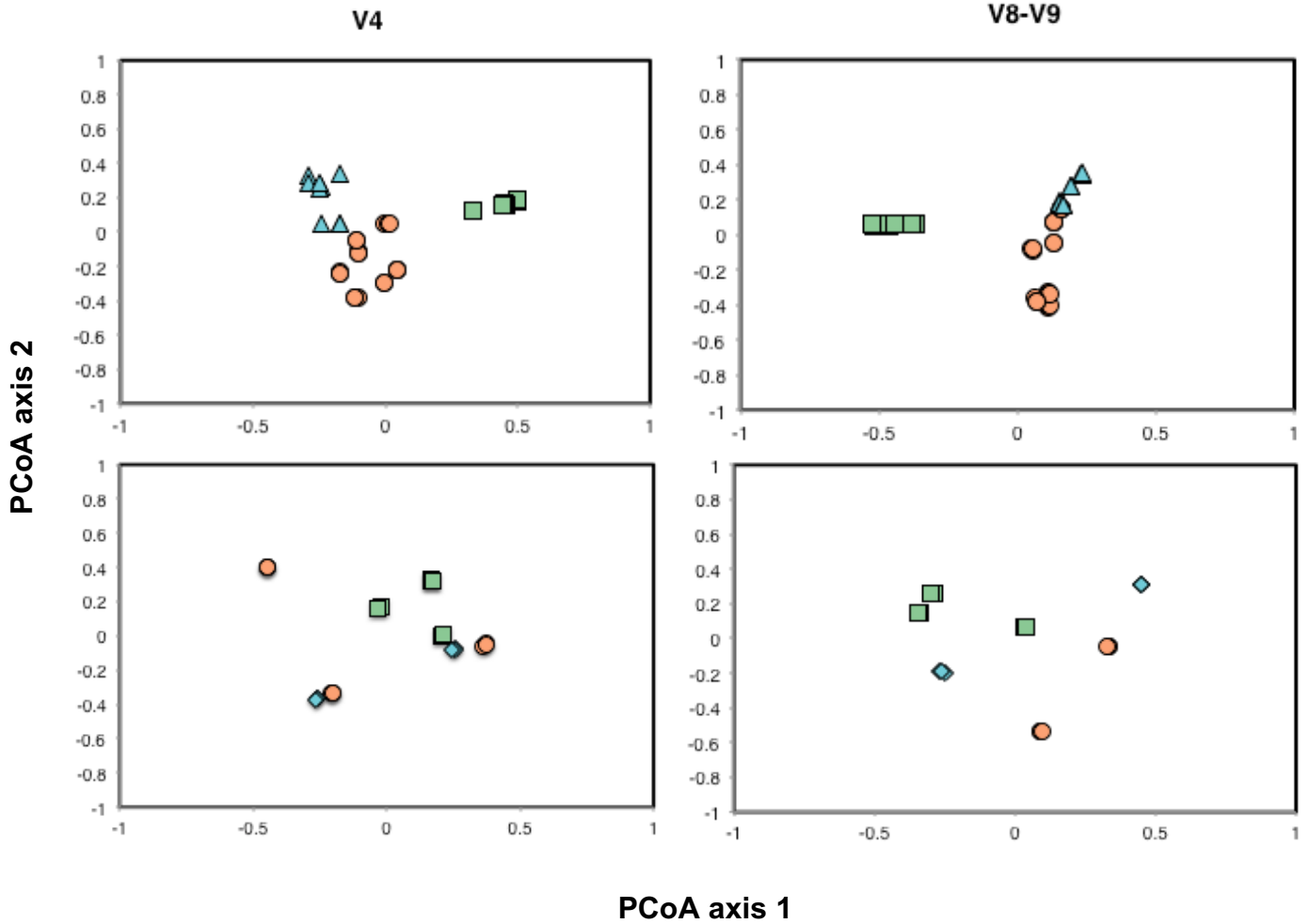
**Figure S8:** Mean relative abundance of each mock community member for MC1, MC3, MC5, and MC7 (in order from top to bottom) determined by targeting the V4 (blue) and V8-V9 (red) hypervariable region of the 18S rRNA gene.



**Figure S9:** Community composition for the V4 (top) and V8-V9 (bottom) regions at the genus level (taxonomy level = 6 in mothur) across all environmental samples. Only genus with relative abundance >0.01% are shown for clarity. Samples containing the letters F, W, and M, are freshwater, wastewater, and marine coastal samples, respectively.



**Figure S10:** PCoA plot of environmental samples using Bray-Curtis metrics for V4 (left) and V8-V9 (right). The top graphs show freshwater (green), marine (blue), and wastewater (orange) samples. The bottom graphs show samples within the wastewater treatment process (orange – primary clarification, green – treatment, blue – secondary clarification).



**Table S6:** Results for Analysis of molecular variance (AMOVA) test comparing sequencing libraries for environmental samples for the V4 and V8-V9 hypervariable region of the 18S rRNA gene.

<b>Groups</b>	<b>V4</b>	<b>V8-V9</b>
<b>Environmental samples</b>		
freshwater-marine-wastewater	p-value <0.001	p-value <0.001
freshwater-marine	p-value <0.001	p-value <0.001
freshwater-wastewater	p-value <0.001	p-value <0.001
marine-wastewater	p-value <0.001	p-value <0.001
<b>Wastewater samples</b>		
primary-secondary-treatment	p-value <0.001	p-value <0.001
primary-secondary clarification	p-value 0.004	p-value 0.005
primary clarification-treatment	p-value 0.002	p-value 0.002
secondary clarification-treatment	p-value <0.001	p-value <0.001

**Table S7:** Mean relative abundance of environmental samples for the V4 and V8-V9 regions at the class level (taxonomy level = 3 in mothur). These percentages correlate to the bar graphs displayed in Figure 8.

	V4													
	4F	5F	6F	4W	5W	6W	7W	8W	9W	10W	1M	2M	4M	5M
Chloroplastida	12%	10%	8%	34%	100%	64%	99%	24%	4%	99%	0%	0%	7%	2%
Rhodophyceae	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	17%	0%	0%	0%
Cryptomonadales	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Goniomonas	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Kathablepharidae	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	1%
Pavlovophyceae	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	92%	5%	0%
Holozoa	2%	1%	7%	1%	0%	2%	0%	1%	50%	0%	2%	0%	13%	9%
Nucleomycea	1%	1%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
unclassified														
Opisthokonta	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Picomonadea	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	4%	2%
Alveolata	25%	40%	38%	4%	0%	2%	0%	12%	2%	0%	9%	0%	15%	22%
Rhizaria	11%	8%	5%	11%	0%	0%	0%	0%	1%	0%	48%	0%	13%	21%
Stramenopiles	6%	7%	22%	36%	0%	3%	1%	56%	13%	0%	13%	7%	37%	30%
unclassified SAR	4%	1%	1%	0%	0%	0%	0%	1%	0%	1%	1%	0%	0%	4%
unclassified														
Eukaryota	38%	32%	18%	13%	0%	28%	0%	5%	30%	0%	7%	0%	4%	9%
	V8-V9													
	4F	5F	6F	4W	5W	6W	7W	8W	9W	10W	1M	2M	4M	5M
Chloroplastida	6%	4%	3%	22%	100%	47%	98%	12%	2%	99%	0%	0%	1%	1%
Rhodophyceae	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Cryptomonadales	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Goniomonas	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Kathablepharidae	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Pavlovophyceae	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	86%	2%	0%
Holozoa	3%	1%	14%	3%	0%	1%	0%	1%	11%	0%	2%	1%	9%	12%
Nucleomycea	1%	2%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%
unclassified														
Opisthokonta	0%	0%	0%	0%	0%	1%	0%	1%	2%	0%	0%	0%	0%	0%
Picomonadea	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	2%	2%
Alveolata	17%	50%	4%	13%	0%	4%	0%	28%	3%	0%	24%	0%	6%	18%
Rhizaria	4%	2%	1%	5%	0%	0%	0%	0%	0%	0%	38%	0%	3%	13%
Stramenopiles	9%	6%	24%	51%	0%	6%	1%	54%	53%	0%	17%	13%	71%	27%
unclassified SAR	18%	3%	38%	0%	0%	0%	0%	0%	3%	0%	4%	0%	1%	4%
unclassified														
Eukaryota	43%	33%	15%	6%	0%	42%	0%	4%	25%	1%	13%	0%	3%	21%

**Figure S11:** Principle coordinate analyses using Bray-Curtis distance metrics indicate that amplicon data clusters according to sample type (freshwater, green; wastewater orange; coastal, grey). The V4 samples (circles) and the V8-V9 samples (square) cluster together for each sample type.

