SUPPLEMENTARY METHODS:

Preparation and Sanger sequencing of controls— We prepared four control PCR products with unique barcoded primers by PCR amplifying plasmid DNA cloned from four different bacterial *nirS* genes. Partial *nirS* genes from a *Pseudomonas aeruginosa* strain, an environmental isolate from the Arabian Sea, SK209-23-5, and two clones (SHC2C and SHCSF) from environmental DNA extracted from sediments of the Patten's Cove salt marsh (Boston MA), were PCR amplified using 0.4 μM of primers nirS1F and nirS6R (Braker *et al.*, 1998), together with Phusion® High-Fidelity DNA Polymerase (New England BioLabs, Ipswich, MA), 1× Phusion HF buffer, 1.5 mM $MgCl_2$, 3% DMSO, 1.6 mM total dNTP (Roche Applied Science, Indianapolis, IN), and 400 μg ml^{-1} non-acetylated BSA (Sigma-Aldrich, St. Louis, MO) on an S1000™ Thermal Cycler (Bio-Rad, Hercules, CA). An initial denaturing step at 98°C for 2 min was followed by 30 amplification cycles consisting of 10 s at 98°C, 30 s at 61°C, and 1 min at 72°C, and by a final elongation step at 72°C for 5 min. PCR fragments of ca. 890 bp, visualized on a 1.5% agarose gel (1×TAE buffer), were excised and purified using the QIAquick Gel Extraction Kit (QIAGEN, Valencia, CA). A-overhangs were added to purified PCR products following the TOPO® cloning manual. Clone libraries were made using the TOPO® TA Cloning® Kit with pCR®2.1-TOPO vector (Life Technologies, Grand Island, NY) to obtain a unique *nirS* gene clone from each purified PCR product. Plasmid DNA was purified using the AxyPrep Plasmid Miniprep Kit (Axygen, Union City, CA) and Sanger sequenced bi-directionally using M13F and M13R primers on an ABI 3730XL DNA Analyzer (Applied Biosystems, Foster City, CA) at Massachusetts General Hospital DNA Core Facility (Cambridge, MA).

Pyrosequencing of controls—For pyrosequencing of control sequences we used 38 pg μl^{-1} of the purified plasmid DNA from the four controls, along with 0.1 μM of primers nirS3F/nirS6R (Braker *et al.*, 1998), and the same concentrations of other reagents as above, but with slightly more $MgCl_2$ (1.75 mM) and with an annealing temperature of 56°C. PCR amplicons of ca. 650 bp were excised and purified. These primers were selected because they yielded fragments of the appropriate length for pyrosequencing on the Roche FLX sequencer with titanium chemistry and are nested within the primer set most frequently used in marine environments (Braker *et al.*, 1998). The primers were constructed with the required sequence adaptors for the FLX sequencing protocol along with a multiplex identifier (MID) that was unique to each forward primer (Table S1). As with all amplicon sequencing, any bacteria containing homologs of *nirS* that are not detected by these degenerate primers will result in an underestimate of the overall diversity of this gene in the environment.

Site description and collection of environmental samples— We collected samples from eight salt marsh plots at Great Sippewissett Marsh, Falmouth MA in September 2009. Samples are designated as marsh plots 1-9 (mp1-mp9; marsh plot 4 was not used in this analysis). We collected surface sediments (top 0.5 cm) using a sterile 25 cc syringe corer. All cores were collected from the rooting zone of pure or nearly pure stands of *Spartina alterniflora*, taking care to avoid the edge of the plots that are in close contact with creek banks. All cores were immediately frozen in liquid nitrogen and returned to the laboratory where they were stored at -80°C until DNA extraction.

Pyrosequencing of environmental samples— Environmental DNA from the salt marsh plots were extracted in triplicate using the PowerSoil® DNA Isolation Kit from MoBio Laboratories (Carlsbad, CA). DNA products were pooled and amplified with primers nirS1F and nirS6R, in three independent PCR runs. PCR products were agarose gel verified and pooled. Nested PCR was performed on 0.15% (v/v in final PCR reaction mixture) of the pooled PCR products using the adapted nirS3F and nirS6R primers (Table S1) in three independent PCR runs for each

sample. Gel purified PCR amplicons were quantified via Quant-IT™ Picogreen® Reagent from Invitrogen (now Life Technologies, Grand Island, NY) on an Agilent MX3005p qPCR system (Santa Clara, CA). Both control and environmental samples for pyrosequencing were eluted in 10 mM Tris-HCl buffer (pH 8) to a concentration of 10 ng μl^{-1} and combined in equal ratios for pyrosequencing on Roche's 454 FLX Genome Sequencer (Branford, CT) at the Josephine Bay Paul Center for Comparative Molecular Biology and Genomics at the Marine Biological Laboratory in Woods Hole, MA. Pyrosequencing was performed using the 454 Titanium Chemistry following manufacturers instructions.

SUPPLEMENTARY REFERENCE

Braker G, Fesefeldt A, Witzel K. (1998). Development of PCR primer systems for amplification of nitrite reductase genes (*nirK* and *nirS*) to detect denitrifying bacteria in environmental samples. *Appl Environ Microb* **64**: 3769–3775.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1: Overview of the data flow in FunFrame. Red vertices are user-provided input files; green represents intermediate files and outputs; edge labels indicate programs. Blue vertices and dashed lines indicate optional Qiime processing for UniFrac distances and sub-sampled alpha diversity estimation.

Figure S2: Heat map of Bray-Curtis dissimilarity values calculated from the *nirS* community composition of each of the salt marsh plots. A dissimilarity value of 1.00 indicates complete divergence among the samples. These data make ecological sense, as marsh plots 2, 6, and 8 are all plots that have more similar elevations above mean sea level than the other cluster of marsh plots.

Figure S3: Principle coordinates analysis of unweighted UniFrac similarities calculated within FunFrame. The UniFrac results mirror the Bray-Curtis results in Fig. S2, with the higher elevation salt marsh plots 2, 6, and 8 clustering together.

SUPPLEMENTARY TABLES

Table S1: Primer sequences, including pyrosequencing adaptors and multiplex identifiers (MID) for each sample. All samples used the same reverse primer (nirS3R: 5'-CTATGCGCCTTGCCAGCCCGC-TCAG-CGTTGAACTTRCCGGT-3'

Primer name	MID	Sample ID	Full primer sequence
NirS3Fa	TCGAG	mp2	5'-CGTATCGCCTCCCTCGCGCCA-TCAG-TCGAG-TTCCTBCAYGACGGYGGC-3'
NirS3Fb	TAGTG	mp1	5'-CGTATCGCCTCCCTCGCGCCA-TCAG-TAGTG-TTCCTBCAYGACGGYGGC-3'
NirS3Fd	ACTCG	mp3	5'-CGTATCGCCTCCCTCGCGCCA-TCAG-ACTCG-TTCCTBCAYGACGGYGGC-3'
NirS3Fr	CGACG	mp5	5'-CGTATCGCCTCCCTCGCGCCA-TCAG-CGACG-TTCCTBCAYGACGGYGGC-3'
NirS3Ft	TGACT	mp6	5'-CGTATCGCCTCCCTCGCGCCA-TCAG-TGACT-TTCCTBCAYGACGGYGGC-3'
NirS3Fu	ACTGC	mp7	5'-CGTATCGCCTCCCTCGCGCCA-TCAG-ACTGC-TTCCTBCAYGACGGYGGC-3'
NirS3Fv	GTCAC	mp8	5'-CGTATCGCCTCCCTCGCGCCA-TCAG-GTCAC-TTCCTBCAYGACGGYGGC-3'
NirS3Fw	GACAG	mp9	5'-CGTATCGCCTCCCTCGCGCCA-TCAG-GACAG-TTCCTBCAYGACGGYGGC-3'
NirS3Fx	ATGCT	*P. aeruginosa*	5'-CGTATCGCCTCCCTCGCGCCA-TCAG-ATGCT-TTCCTBCAYGACGGYGGC-3'
NirS3Fy	GATCT	SK209235	5'-CGTATCGCCTCCCTCGCGCCA-TCAG-GATCT-TTCCTBCAYGACGGYGGC-3'
NirS3Fz	ACACT	SHC2C	5'-CGTATCGCCTCCCTCGCGCCA-TCAG-ACACT-TTCCTBCAYGACGGYGGC-3'
NirS3Faa	AGCTA	SHCSF	5'-CGTATCGCCTCCCTCGCGCCA-TCAG-AGCTA-TTCCTBCAYGACGGYGGC-3'

Table S2: Analysis of pyrosequencing error of four clones containing the *nirS* gene fragment and assessment of the data pipeline we developed to remove sequencing noise.

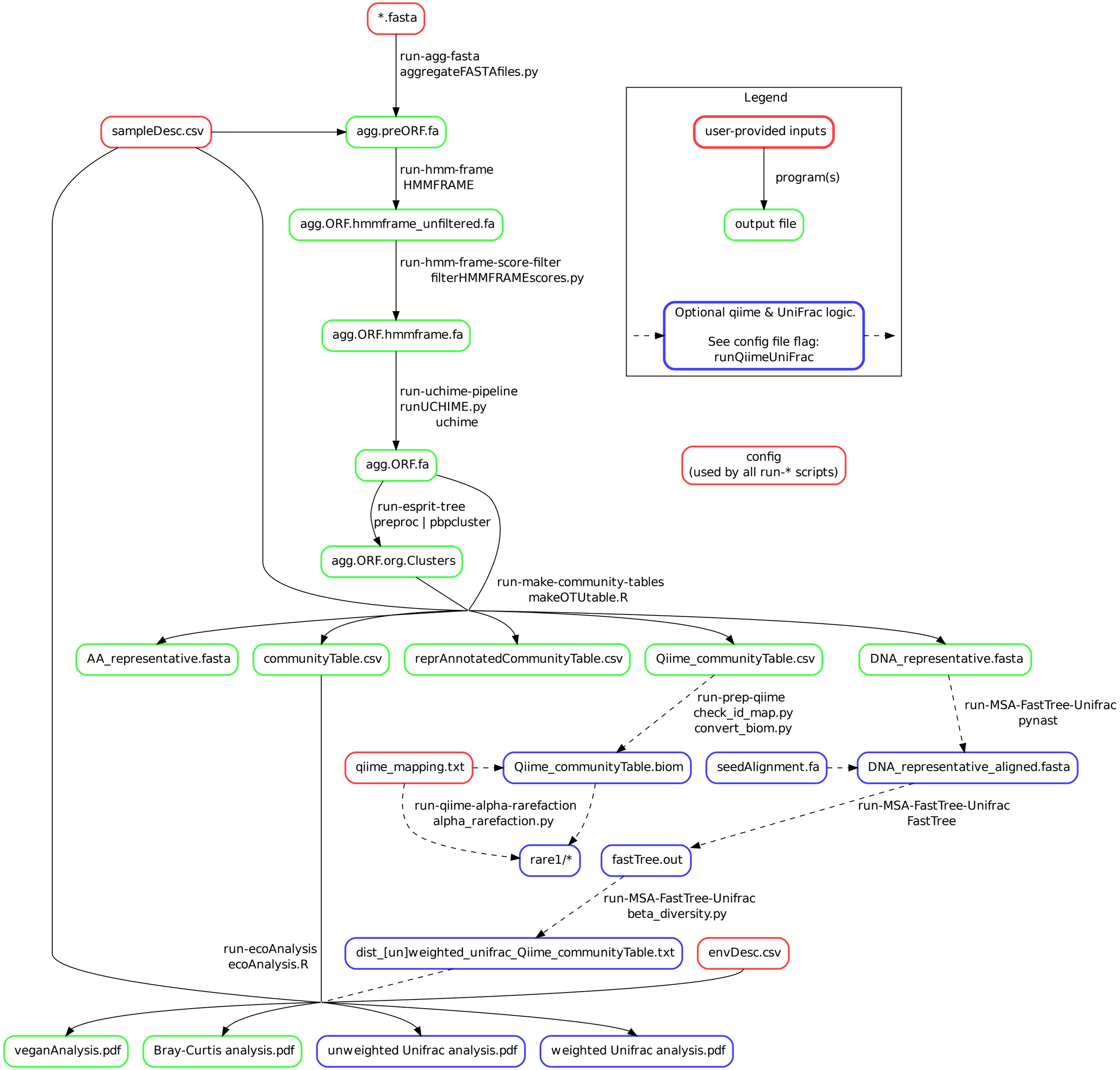
	Clone identification			
	P. aeruginosa	SHC2C	SHCSF	SK209235
Before pipeline:				
Total number of sequences	9084	5780	11,382	14,494
Number of exact matches	5057	1618	5496	6077
Number of OTUs	2332	1733	2091	3681
Number of singletons	2062	1515	1778	3245
Number of doubletons	132	90	152	199
Calculated error rate	44.3%	72.0%	51.7%	58.07%
After FunFrame pipeline:				
Sequences removed	7	320	710	70
Remaining sequence numbers	9077	5460	10,672	14,424
Number of exact matches	8958	5456	10,670	14,424
Number of OTUs	9	3	3	1
Calculated error rate	0.18%[1]	0.073%	0.019%	0.00%
After stop-codon removal				
Sequences removed	110	3889	238	1771
Remaining sequence numbers	8974	1891	11,144	12,723
Number of exact matches	8847	1887	11,142	12,721
Number of OTUs	11	4	3	2
Calculated error rate	1.4%	0.21%	0.017%	0.016%

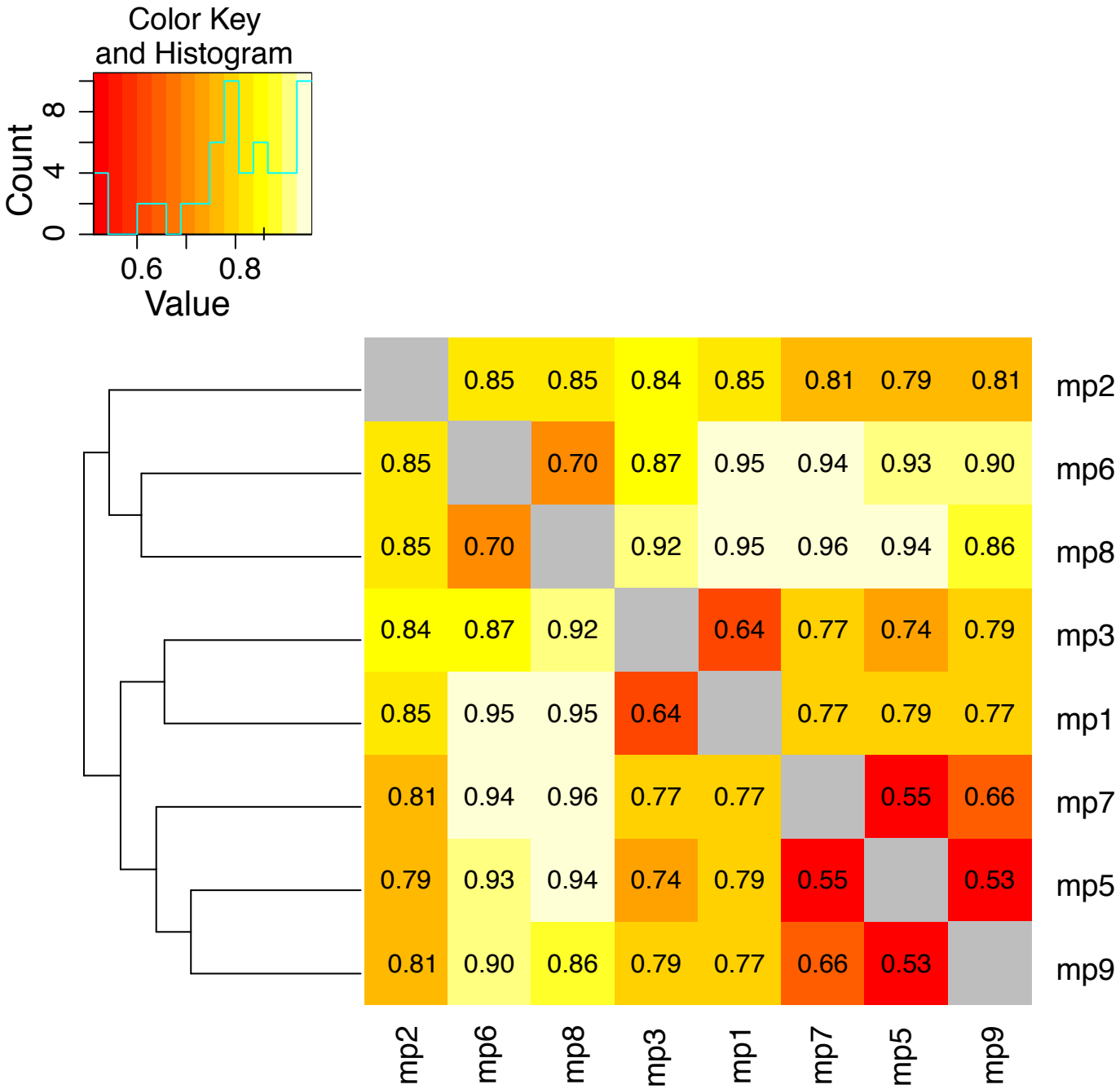
[1]This error rate includes the 8958 sequences that were an exact match to the parent sequence and another 103 sequences that had 98% identity to the parent sequence with overlapping, but not identical, coverage. That is: 100-[(8958+103)/9077*100].

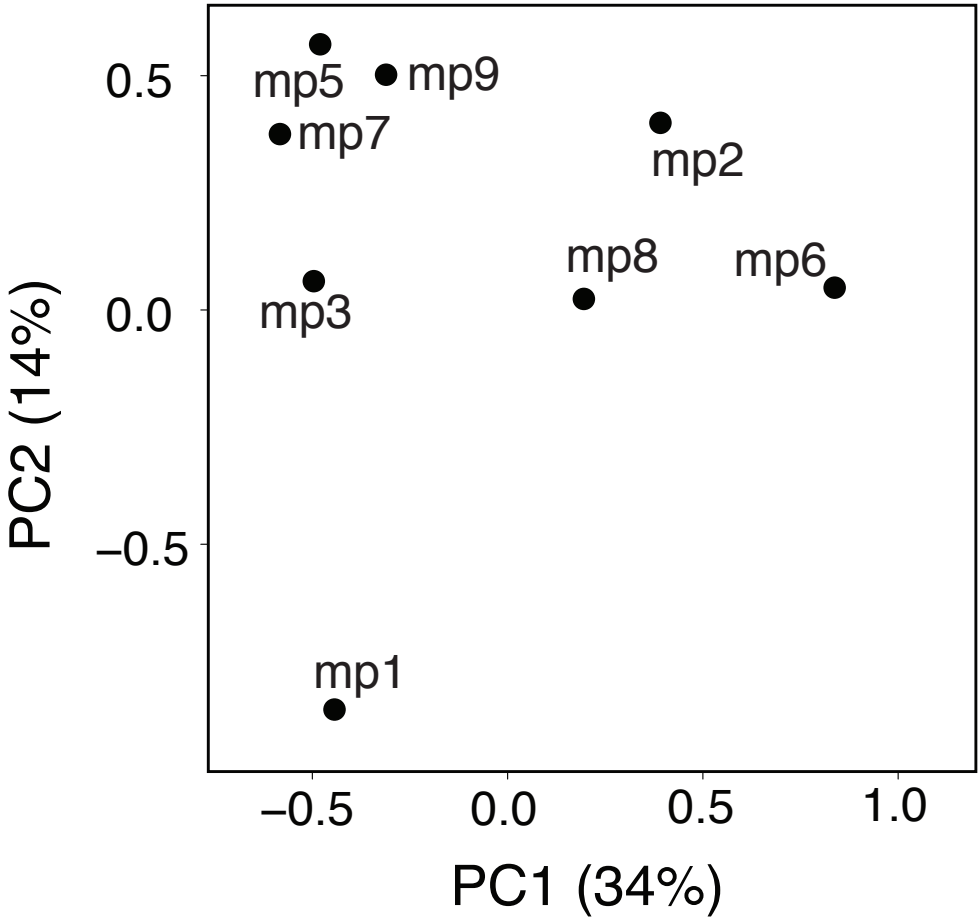
Table S3: Numbers of reads removed throughout the pipeline for each of the marsh sediment samples.

	Number of sequences			
Plot ID	Prior to HMM-Frame	After HMM-Frame and Prior to UCHIME	After UCHIME	% remaining
FunFrame pipeline:				
P. aeruginosa	9084	9078	9077	99.9
SHC2C	5780	5480	5460	94.5
SHCSF	11,382	10,672	10,672	93.8
SK209235	14,494	14,424	14,424	99.5
mp1	9179	9134	8217	89.5
mp2	5829	5621	4662	80.0
mp3	9244	9152	8513	92.1
mp5	11,469	11,411	10,036	87.5
mp6	12,531	12,379	11,096	88.5
mp7	10,214	10,172	8800	86.2
mp8	5797	5751	5248	90.5
mp9	14,660	14,385	12,398	84.6
Sum	**119,663**	**117,659**	**108,603**	**90.8**

	Prior to stop-codon filtering	After stop-codon filtering and Prior to UCHIME	After UCHIME	% remaining
Stop codon filtering				
P. aeruginosa	9084	8975	8974	98.8
SHC2C	5780	1891	1891	32.7
SHCSF	11,382	11,144	11,144	97.6
SK209235	14,494	12,723	12,723	87.8
mp1	9179	8472	7646	83.3
mp2	5829	5309	4218	72.4
mp3	9244	8735	8144	88.1
mp5	11,469	9951	8612	75.1
mp6	12,531	10,952	9815	78.3
mp7	10,214	9268	7954	77.9
mp8	5797	4186	3836	66.2
mp9	14,660	12,741	10,945	74.7
Sum	**119,663**	**104,347**	**95,902**	**80.1**


*.fasta
run-agg-fasta
aggregateFASTAfiles.py
sampleDesc.csv
agg.preORF.fa
run-hmm-frame
HMMFRAME
agg.ORF.hmmframe_unfiltered.fa
run-hmm-frame-score-filter
filterHMMFRAMEscores.py
agg.ORF.hmmframe.fa
run-uchime-pipeline
runUCHIME.py
uchime
agg.ORF.fa
run-esprit-tree
preproc | pbpcluster
agg.ORF.org.Clusters
run-make-community-tables
makeOTUtable.R
AA_representative.fasta
communityTable.csv
reprAnnotatedCommunityTable.csv
Qiime_communityTable.csv
DNA_representative.fasta
run-prep-qiime
check_id_map.py
convert_biom.py
run-MSA-FastTree-Unifrac
pynast
qiime_mapping.txt
Qiime_communityTable.biom
seedAlignment.fa
DNA_representative_aligned.fasta
run-qiime-alpha-rarefaction
alpha_rarefaction.py
run-MSA-FastTree-Unifrac
FastTree
rare1/*
fastTree.out
run-MSA-FastTree-Unifrac
beta_diversity.py
run-ecoAnalysis
ecoAnalysis.R
dist_[un]weighted_unifrac_Qiime_communityTable.txt
envDesc.csv
veganAnalysis.pdf
Bray-Curtis analysis.pdf
unweighted Unifrac analysis.pdf
weighted Unifrac analysis.pdf
Legend
user-provided inputs
program(s)
output file
Optional qiime & UniFrac logic.
See config file flag:
runQiimeUniFrac
config
(used by all run-* scripts)


Color Key
and Histogram
Count
0
4
8
0.6
0.8
Value
0.85 0.85 0.84 0.85 0.81 0.79 0.81
0.85 0.70 0.87 0.95 0.94 0.93 0.90
0.85 0.70 0.92 0.95 0.96 0.94 0.86
0.84 0.87 0.92 0.64 0.77 0.74 0.79
0.85 0.95 0.95 0.64 0.77 0.79 0.77
0.81 0.94 0.96 0.77 0.77 0.55 0.66
0.79 0.93 0.94 0.74 0.79 0.55 0.53
0.81 0.90 0.86 0.79 0.77 0.66 0.53
mp2
mp6
mp8
mp3
mp1
mp7
mp5
mp9
mp2
mp6
mp8
mp3
mp1
mp7
mp5
mp9


mp5
mp9
mp7
mp2
mp8
mp6
mp3
mp1
0.5
0.0
−0.5
−0.5
0.0
0.5
1.0
PC2 (14%)
PC1 (34%)