Salt marsh sediment diversity: a test of the variability of the rare biosphere over small spatial scales

Jennifer L. Bowen[a,1], Hilary G. Morrison[b], John E. Hobbie[a] and Mitchell L. Sogin[b]

Running title: salt marsh sediment microbial diversity

Subject category: Microbial population and community ecology

[a]The Ecosystems Center, Marine Biological Laboratory, Woods Hole, MA, USA, 02543

[c]Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA, USA, 02543

[1]Corresponding author. Current address: Biology Department, University of Massachusetts Boston, 100 Morrissey Blvd., Boston MA 02125, jennifer.bowen@umb.edu

Abstract

Much of the phylogenetic diversity in microbial systems arises from rare taxa that comprise the long tail of taxon rank distribution curves. This vast diversity presents a challenge to testing hypotheses about the effects of perturbations on microbial community composition because within site variability of the rare taxa may be sufficiently large that it would require a prohibitive degree of sequencing to discern differences among samples. In this study we used pyrosequencing of 16S rRNA tags to examine the diversity and local-scale variability of salt marsh sediment bacteria. Our goal was to determine whether pyrosequencing could produce similar patterns in community composition among replicate environmental samples from the same location. We hypothesized that repeated sampling from the same location would produce different snapshots of the rare community due to incomplete sequencing of the taxonomically rich rare biosphere. The concern was that variation resulting from incomplete sequencing could mask subtle community shifts caused by environmental perturbations. Our data indicate that salt marsh sediments contain a remarkably diverse array of bacterial taxa and, in contrast to our hypothesis, repeated sampling from within the same site produces reliably similar patterns in bacterial community composition, even among rare organisms. These results demonstrate that deep sequencing of 16s tags is well suited to distinguish site-specific similarities and differences among rare taxa and is a valuable tool for hypothesis testing in microbial ecology.

Key words: pyrosequencing/salt marshes/microbial diversity/rare biosphere

1 ## Introduction

2 For decades microbial ecologists faced the challenge of inferring microbial community
3 composition from modest-sized ribosomal RNA (rRNA) data sets that represented amplicon
4 libraries from environmental DNA. The larger amplicon libraries (>1000 sequences) represented
5 only a very small fraction of the taxa present in most source communities (Whitman *et al.*, 1998)
6 although a few studies have collected on the order of 70,000 sequences (Ley *et al.*, 2006). As a
7 result, a number of mathematical models have been proposed to extrapolate taxonomic richness
8 of microbes based on relatively small sample sizes (summarized in Schloss, 2008; Lozupone and
9 Knight, 2008).

10 Since the first next-generation sequencer became commercially available in 2005,
11 pyrosequencing has become the preferred tool for examining microbial community composition
12 because it allows researchers to sequence much more deeply into a community than had
13 previously been possible with the time and cost constraints of Sanger sequencing (Margulies *et*
14 *al.,* 2005; Sogin *et al*., 2006). One result of this tremendous advance in sequencing capability is
15 the recognition, for the first time, of the vast diversity of low abundance microbial taxa that exist
16 in surface and deep sea waters (Sogin *et al.,* 2006; Huber *et al.,* 2007), soil (Roesch *et al.,* 2007),
17 and human gut (Turnbaugh *et al*., 2009) ecosystems. Kunin et al. (2010) suggest that much of
18 the diversity described in these initial studies are a result of sequencing error, however the error
19 rate of these methods after appropriate quality control procedures do not exceed one error for
20 every 500 bases. The common use of complete linkage algorithms for clustering sequences into
21 OTUs artificially inflates diversity estimates for very simple and complex communities. The
22 single linkage pre-clustering algorithm used here corrects for sequencing error, and provides
23 compelling evidence for the presence of the rare biosphere (Huse et al. 2010). New research is

1 needed to understand the ecological and evolutionary role of the rare biosphere, though evidence
2 already suggests that these rare organisms do display biogeography (Galand *et al.,* 2009) and that
3 they provide a source pool of diversity that allows microbial communities to respond to
4 environmental change (Brazelton *et al.,* 2010).

5 The ability to detect how environmental perturbation alters low abundance microbial taxa
6 (defined operationally as sequences present on average less than five times in 20,000-25,000 tag
7 sequences, or less than 0.025% of the time) requires that the variability of the rare biosphere
8 within a particular site be sufficiently small that differences among sites can be inferred. If low
9 abundance taxa represent a universal source pool of bacteria (the "everything" in Baas Becking's
10 (1934) axiom "everything is everywhere"), we hypothesize that it would be challenging to infer
11 meaningful differences between the rare biospheres of two different samples, even with the depth
12 of sequencing currently possible. However, if there is some sort of environmentally driven
13 functional selection acting on the rare members of the microbial community then, assuming
14 sufficient sampling depth, there should be greater similarity in the rare biospheres of
15 environmental replicates than from samples taken from two different locations.

16 The logic of this argument is as follows: if the rare biosphere represents a source pool of
17 microbes that results from universal dispersal then repeated samples taken from the same site,
18 when not sequenced to completion, will display a snapshot of the rare biosphere that is selected
19 at random from all the low abundance taxa present. Any similarity that happens to exist among
20 the community composition of low abundance taxa in repeated samples would be a result of the
21 chance sequencing of the same equally rare organisms. If this source pool does represent a
22 cosmopolitan distribution of organisms, then a snapshot of the rare biosphere taken from two
23 replicate samples should be roughly as dissimilar as the snapshot taken between two different

samples because in all cases we are subsampling from the same universal source pool. If this hypothesis is true it creates a challenge for testing hypotheses on the role that environmental perturbation plays in determining the distribution of low abundance taxa. It becomes impossible to differentiate whether differences in the distribution of rare organisms in two different samples (for example, samples taken before and after a disturbance event) are real differences or if, instead, the differences are an artifact of incomplete sequencing.

In light of these considerations, we assessed the variability of microbial community compositions in replicate environmental samples taken over very small spatial scales in salt marsh sediments. Salt marshes are critically important marine habitats that are thought to harbor tremendous microbial diversity (Lozupone and Knight, 2007). Salt marshes play a key role in protecting adjacent coastal habitats from human-derived influence (Valiela and Cole, 2002) and because marshes are precariously located between terrestrial uplands and marine waters, they are vulnerable to environmental perturbations from both environments. Many of the ecosystem services provided by salt marshes are microbially mediated, yet little is known about the extent of diversity in these key habitats. Achieving a comprehensive understanding of the role that this microbial diversity plays in ecosystem-scale processes in salt marshes first requires an understanding of the within-site variability in the microbial community. Only if a repeatable baseline community structure can be established will it be possible to assess how human perturbations are altering the ecological subsidy provided by the marsh microbial community.

The objectives of this study were three-fold. First, we documented the extent of bacterial diversity in salt marsh sediments. Second, by examining diversity in both individual and pooled samples taken from the same location in the marsh we assessed the local-scale variability in the sediment microbial community. We hypothesized that pooling and homogenizing sediments

1 from a number of samples, and taking a subsample of the pool would decrease within-site
2 variability and lead to more repeatable patterns in community composition because it would
3 integrate over the inherent patchiness of the system. Finally, we compared individual and pooled
4 marsh sediments with an outgroup sample from the water column of an adjacent marsh creek to
5 test whether pyrosequencing could be used to distinguish the rare biospheres of two different
6 samples. Results from these three objectives provide convincing evidence that in salt marsh
7 sediments the community composition of the rare biosphere is sufficiently similar that a baseline
8 community can be described, a necessary first step for testing hypotheses regarding the role of
9 human disturbance in structuring microbial communities.

10 **Methods**

11 *Sample collection*

12 We collected samples from the tall form *Spartina alterniflora* habitat of the Great
13 Sippewissett Salt Marsh in Falmouth, MA (41° 34.58 N, 70° 38.23 W) on 10 September 2008
14 from within a 100 cm^2 area of unvegetated marsh sediments. A sterile 5 cc syringe core was
15 used to sample the top 1 cm of marsh sediment. Six individual samples were taken and extruded
16 immediately into separate 2 mL cryovials that were stored on ice and then transferred to a -80°C
17 freezer at the Marine Biological Laboratory in Woods Hole, MA. An additional 12 sediment
18 cores were also taken from the same 100 cm^2 area; six of the 12 cores were pooled in a sterile 20
19 mL scintillation vial and the remaining six were extruded into a second scintillation vial. These
20 vials were stored on ice and returned to the lab where they were homogenized with a sterile
21 spatula. Subsamples from each of the pooled and homogenized cores were removed and stored
22 at -80°C in 2 mL cryovials. The microbial community from the water column of a creek
23 draining the adjacent Little Sippewissett Salt Marsh that was sampled on 10 July 2007 served as

1 an outgroup. One liter of water was collected in a triple rinsed Nalgene bottle and returned on
2 ice to the lab for filtration. The 1 L sample was vacuum filtered through a Sterivex filter, lysis
3 buffer was added, and the filter unit was stored at -80°C until DNA extraction.

4 *DNA extraction and amplification*

5 DNA from 0.5 grams of marsh sediment was extracted using the PowerSoil™ DNA
6 Isolation kit (MoBio Laboratories, Carlsbad, CA, USA) following manufacturer's instructions.
7 DNA from the water column sample was extracted using the Gentra PureGene DNA extraction
8 kit (Qiagen) also following the manufacturer's instructions. The hypervariable V6 region of the
9 bacterial 16S rRNA gene was amplified using a cocktail of five forward and four reverse primers
10 that amplify the vast majority of known bacteria (Huber *et al.,* 2007). The primers contain the
11 Roche A- and B- adapters fused to a 5-nucleotide multiplex identifier (MID) and terminated by
12 19 bp that complement conserved regions flanking the bacterial 16S rRNA genes. The MID
13 allows the bioinformatic identification of pyrosequencing reads from multiple samples in a single
14 pyrosequencing analysis (Huber *et al.,* 2007). Amplified DNA was purified using a MinElute
15 PCR Purification kit (Qiagen, Valencia, CA) and quantified on a Bioanalyzer 2100 (Agilent,
16 Palo Alto, CA) prior to sequencing on a Roche GSFLX pyrosequencer. Further details on these
17 methods have been published elsewhere (Sogin *et al*., 2006; Huber *et al.,* 2007; Huse *et al.,*
18 2007; Huse *et al*., 2008; Huse *et al*., 2010).

19 *Data analysis*

20 After sequencing, data were subjected to rigorous quality control checks as described
21 previously (Huse *et al.,* 2007; Huse *et al*., 2008; Huse *et al*., 2010). These quality control
22 measures included the removal of all reads that had any ambiguous base calls, that had read
23 lengths longer than the typical distribution of sequence lengths, or that had inexact matches to

1 the initial primers. With these quality checks in place the read error rate associated with
2 pyrosequencing was reduced to less than 0.2% (Huse *et al.*, 2007). Sequences that passed
3 quality checks were trimmed to remove both primers and were then assigned taxonomy using
4 GAST (Huse et al. 2008). The single linkage preclustering algorithm (SLP; Huse et al. 2010)
5 used nearest neighboring on rank abundance sorted sequences to identify 2% preclusters, and
6 average neighboring in mothur (Schloss et al. 2009) to identify 3%, 6% and 10% clusters
7 (OTUs). All data were normalized to the sample that contained the highest number of sequence
8 tags (ENV 1: 24,675 (range: 20,783-24,675). The CatchAll software program (Bunge et al.
9 2010) calculated non-parameteric ACE and the Chao1 richness indices. We used EstimateS
10 (Version 8.0.0, R. K. Colwell, http://purl.oclc.org/estimates) to calculate similarity matrices
11 using the Bray Curtis similarity index [C_N=2jN/(aN +bN), where aN = total number of
12 individuals in site A, bN = total number of individuals in site B, and jN = the sum of the lower of
13 the two abundances in both samples]. The Vegdist program in R was used to calculate
14 dissimilarties and to construct phenograms using average linkage clustering, which is an
15 Unweighted Pair Group Mean (UPGMA) method of analysis. The cumulative frequency
16 histograms were calculated on natural log transformed abundance data using the GraphPad
17 Software (La Jolla, CA) statistical package Prism. Curve fit parameters were determined in
18 Prism by fitting Gaussian curves to the data using a least squares fit.

19 **Results and discussion**

20 *Salt marsh microbial diversity*

21 Of the 43 phyla recognized in these analyses all but one, Caldiserica, was present at least
22 one time in our salt marsh samples (Table S1). Marsh sediments were dominated by the
23 Proteobacteria, but had considerable contributions from Bacteroidetes, Acidobacteria,

1 Chloroflexi, Gemmatimonadetes, Planctomycetes, and Verrucomicrobia. By contrast, the water
2 column sample used as an outgroup was >90% Proteobacteria, with a minor contribution from
3 Bacteroidetes (7%) and Cyanobacteria (1%). The remaining 27 phyla present accounted for less
4 than 2% of the organisms sequenced in the water column sample (Table S1). We examined the
5 distribution of orders within the Proteobacteria to further describe the community composition of
6 the sediment samples. Within the Proteobacteria there were 47 identified orders of which 39
7 were present in the marsh sediment samples (Table S2). The most abundant orders were roughly
8 evenly split among Rhodobacterales (12%) Myxococcales (13%), unidentified δ-proteobacteria
9 (10%), and Xanthomonadales (14%). Of these dominant orders, only Rhodobacterales was also
10 numerically important in the water column outgroup sample. The other two orders that
11 dominated the water column sample were Rickettsiales, of which the ubiquitous pelagic bacteria
12 SAR11 is a member, and Alteromonadales (Table S2).

13 Analysis of samples at the phylum and order levels indicated remarkable similarity
14 among the sediment samples and at both levels of biological organization the sediments were
15 quite different than the water column outgroup (Tables S1, S2). These results suggest a degree
16 of functional selection acting within the sediments, but the dominant members of the community
17 drive these conclusions. A higher resolution analysis of the microbial community composition
18 can be performed at the species level (Fig. 1). Using the software present in the Visualization
19 and Analysis of Microbial Population Structure analysis pipeline (http://vamps.mbl.edu/) we
20 plotted the relative abundance of bacterial species in each of the sediment samples and in the
21 water column outgroup sample (LSM). When all species were included in the analysis (Fig. 1A)
22 there were clear similarities among all sediment samples and they were distinctly different than
23 the water column sample. Taking the analysis one step further, we examined the distribution of

1 just those species that accounted for less than 1% of all species present in the sample (that is, we
2 removed all the abundant taxa to look more closely to the rare species). Although the microbial
3 community composition of the sediments shows a greater variability when the most abundant
4 taxa are removed (Fig. 1B) there is still a relatively greater degree of cohesion among the
5 sediment samples than between the sediment and water samples. This apparent cohesion
6 suggests that there is some environmentally induced functional selection acting on the sediment
7 samples that maintains similarity in community composition even among the low abundance
8 taxa.
9 Next, we used the clustering methodology described by Huse et al. (2010) to calculate
10 rarefaction curves and different estimators of diversity (Fig. 2, Table 1), for each of the six
11 sediment samples and the water column outgroup sample at three different degrees of clustering,
12 3% (Fig. 2A), 6% (Fig. 2B) and 10% (Fig. 2C). In all cases there were no differences between
13 the individual (ENV1-ENV6) and pooled (HOM1 and HOM2) samples, though all sediment
14 samples had considerably higher richness than the water column sample. Furthermore, even at
15 the 10% clustering level the slope of the sediment rarefaction curves remain curvilinear,
16 indicating that there was likely considerable diversity yet to be sequenced.
17 Additional estimators of taxonomic richness, the Chao and Ace estimators and the
18 Shannon Diversity index were calculated based on tags clustered at 3%, 6%, and 10% sequence
19 divergence (Table 1). These estimators are sensitive to the depth of sampling, averaging an
20 inflation of one OTU for every 1000 sequence reads (Mark Welch et al. in prep), but the samples
21 included in these analyses were sequenced to roughly the same depth so we take these indicators
22 to provide a good relative estimation of taxa richness and diversity. At the 3% clustering level,
23 each of the sediment samples contained twice as many observed OTUs (~4100) as did the water

1 column sample (~1850 OTUs). Chao and ACE estimators tend to underestimate actual richness
2 due to their extrapolation from small sample sizes (Hong *et al.*, 2006, Quince *et al.*, 2008).
3 However, as a minimum estimate these estimators indicated that there are between 7000-10,000
4 bacterial OTUs in the sediments when clustered at 3% sequence divergence (Table 1). This
5 surpasses the Chao estimates of richness for 3% clusters in the water column, but the ACE
6 estimator of bacterial 3% OTUs in the water column sample was roughly equivalent to the
7 sediment sample estimates. When clustered at the 6% and 10% sequence divergence levels both
8 richness metrics indicated that the estimated taxonomic richness in the water column sample was
9 considerably lower than the estimated richness of the sediment samples (Table 1). Shannon
10 Diversity indices calculated at the standard 3% level of sequence divergence also suggest greater
11 diversity in the sediments than in the water column (Table 1). These estimates of diversity and
12 richness are within those reported for other soils (Roesch *et al.* 2007, Morales *et al.* 2009).
13 Several factors may have contributed to the tremendous bacterial diversity found in these
14 salt marsh sediments. Located between terrestrial uplands and marine waters, salt marshes are
15 strongly influenced by both habitats (Valiela and Teal, 1979) and may retain legacies of both
16 microbial source communities. Steep and fluctuating redox gradients in salt marshes (Howes *et*
17 *al.*, 1981) also suggest a wide range of substrates amenable to microbial metabolisms. Different
18 mineral fractions of soils have distinct bacterial communities (Carson *et al.*, 2009), so variations
19 in mineral content of the marsh could increase microbial diversity. Furthermore, organic matter
20 has tremendous spatial complexity at small scales (Lehman *et al.*, 2008) so organic rich salt
21 marsh sediments likely have considerable diversity associated with niche differentiation around
22 organic aggregates.
23 *Comparison of individual and homogenized samples*

1 The factors that promote diversity in marsh sediments also act to promote patchiness
2 within those sediments. We hypothesized that this patchiness would lead to very high within-site
3 variability that would make it difficult to establish a baseline community composition from
4 which to discern differences among treatments in an experimental perturbation. The data,
5 however, demonstrate remarkable similarity in community composition among multiple samples
6 collected from within the same region of the marsh (Fig. 1, Tables S1, S2), suggesting that
7 within-site variability is small. Further evidence that within-site variability is small can be
8 gleaned from a comparison of the individual samples with the pooled samples. We hypothesized
9 that pooling multiple sediment cores and sequencing a subsample from the pooled and
10 homogenized sediments would produce a snapshot of the community that would be more
11 representative than any single snapshot from individual samples. By sequencing the pooled
12 subsample to the same depth as each of the individual samples the data would be skewed toward
13 those taxa that were present in multiple subsamples. This would decrease the importance of
14 patchy taxa and of the very minor constituents of the rare community that were only present in
15 one or two of the subsamples. The result would be a repeatable assessment of within-site
16 variability, a necessary step for subsequent hypothesis testing.
17 Remarkably, the community composition in the pooled samples (HOM1 and HOM2) does
18 not appear to be tremendously different from the individual samples (ENV1-ENV6, Fig. 1). The
19 only plausible explanation for this similarity is that the community composition of each of the
20 pooled subsamples was roughly similar to each of the individual samples. If there were patches
21 of different microbes that were locally abundant (present in one or two subsamples but not in all
22 six) this would skew the taxa abundances in the homogenized samples such that they would be
23 different than the individual samples. That the data do not demonstrate this skew in either

1 homogenized sample lends further support to the conclusion that the within-site variability in
2 these sediments is small. Creating a mechanism to quantify within-site variability will allow for
3 the identification of a baseline community so that deviations from the baseline can be observed.
4 *Quantifying similarities among samples*
5 If pyrosequencing is to be effectively used to examine differences in microbial
6 communities either along environmental gradients, or that result from environmental
7 perturbations, within-site variability must be quantified sufficiently well that a different site (or a
8 post-disturbance community within the same site) can be distinguished. If sites were entirely
9 dominated by a few numerically abundant taxa that differ from location to location this would be
10 a relatively simple statistical test. Most pyrosequencing data, however, suggest the presence of a
11 long tail of low abundance taxa that exist in many habitats (*e.g.* Sogin *et al.*, 2006). It is
12 therefore not sufficient to examine differences among dominant taxa; it must also be possible to
13 quantify similarities and differences among the rare members of the microbial community.
14 As a first step we quantified the differences in bacterial community composition of the
15 individual and homogenized sediment samples using the Bray-Curtis similarity index (Magurran,
16 1988). We then calculated similarities between the sediment samples and the water column
17 outgroup sample. The input data for these analysis came from the GAST taxon assignments
18 generated via the Marine Biological Laboratory's VAMPS pipeline (http://vamps.mbl.edu/). We
19 compared similarities across the entire community of microbes and also among the most
20 abundant taxa (operationally defined as those taxa present, on average, more than 100 times per
21 sediment sample), the rare taxa (operationally defined as those present, on average, fewer than 5
22 times per sediment sample), and the various clusters in between those two extremes (Fig. 3).

We hypothesized that environmental selection within sediment samples would lead to considerable similarities among the most abundant taxa but that even among the most abundant taxa the sediments would have little similarity with the water column outgroup. Furthermore, when comparing taxa with low abundances the community similarity in replicate sediment samples would go down because incomplete sequencing would lead to a snapshot of taxa selected at random from all the low abundance taxa present in each sample. We feared that this stochastic element would increase dissimilarity among sediments and would make interpreting results of experimental perturbations difficult. If the dissimilarity created by incomplete sequencing of replicate samples was sufficiently large, there would be as much dissimilarity among the replicate sediment samples as there would be between the sediment samples and the outgroup water sample.

We were correct that the abundant taxa in the sediment samples were similar to one another both within the individual environmental replicates (Fig. 3, blue columns) and between the individual and homogenized samples (Fig. 3, red columns), though the extent of the similarity (>80%) was a surprise (Fig. 3). It was also not surprising that the dominant members of the sediment bacterial community were considerably different than the dominant members of the bacterial community from the water column sample (Fig. 3, green columns). The more surprising feature of these data is evident when examining the similarities and differences among the rare members of the community. Although similarity among sediment samples did decrease as the number of sequences per tag decreased, even among those tags present fewer than five times in over 20,000 sequences per sample, there was a remarkable degree of similarity (~44%). If variability within the community composition of the rare sediment microbes were large then the chance sequencing of identical rare tags would be low, resulting in low similarity among

1 replicate samples. That the data indicates a similarity of ~44% among rare taxa in the sediment
2 replicates suggests that there is functional selection acting among the rare members of the
3 community and that there is considerably greater similarity among the rare biosphere of
4 environmental replicates than between the rare biosphere of the sediments and the water column
5 outgroup.

6 When including all the taxonomic data, an unweighted Pair Group Mean Analysis
7 (UPGMA) phenogram shows one cluster of sediment samples that are only 20-30% dissimilar
8 but that is more than 80% dissimilar to the outgroup water column sample (Fig. 4A). As a
9 further test of whether the rare biosphere of similar samples could be distinguished from the rare
10 biosphere of a different sample, we also performed the UPGMA on taxa present fewer than five
11 times (Fig. 4B). The UPGMA clusters of the rare taxa show a slightly different order of
12 clustering than when all sequences were considered (Fig. 4A), but nonetheless all sediment
13 samples cluster together and are far removed from the outgroup. This provides further evidence
14 that environmental replicates display similar community compositions, even among the rare
15 members of the consortia.

16 Microbial communities that have fundamentally different structures would not only
17 cluster differently from one another they would likely have different cumulative frequency
18 distributions. While it is possible that two samples could have different community
19 compositions but similar frequency distributions, the inverse is not, that is, communities that
20 have different cumulative frequency distributions cannot have the same community structure.
21 Quantifying the shape of the frequency distribution can thus provide a mechanism for confirming
22 differences in community compositions that may result from environmental perturbation. We
23 characterized the frequency distribution of the sediment samples by fitting Gaussian curves to

the data (Fig. 5). The amplitude, mean, and standard deviations of these curves can then be used to compare among replicates and to contrast with the outgroup sample. The sediment replicates had similarly shaped curves and overlapping 95% confidence intervals (Table 2). Averaged across all the sediment samples the amplitude of the Gaussian curves indicates that the sediment samples had approximately 4000 OTUs (4168 ± 314) compared to 1056 OTUS in the outgroup, thus confirming our previous conclusion that these sediment samples harbor considerably greater diversity than was found in the water column draining an adjacent marsh.

The mean and standard deviation of the Gaussian curve fits, indicators of the number of sequences per tag and the spread of the data, respectively, were higher in the water column outgroup than in the sediment samples (Table 2, Fig. 5). This would be expected from a sample that is dominated by a handful of very abundant taxa. The sediment samples, however, contain fewer very high abundance tags; rather, they have a more even distribution of less abundant taxa. This is evident by the different extent of the curves along the x-axis (Fig. 5). In the sediments it takes 250-300 of the most abundant tags to account for 50% of all the sequences; in the water column just the two most dominant tags account for 50%.

Both the sediment samples and the water column outgroup sample demonstrate a long tail of low abundance taxa, but this tail is considerably longer in the sediment samples. This is indicated both by the overall taxonomic richness (Table 1) as well as by the Gaussian curve fits. The location of the y-intercept on each of the curves indicates the number of sequences that occur only one time (Fig. 5). This particular water column sample had 625 tags that occurred once, compared to between 1750 and 2250 tags in the sediment samples. Furthermore, the initial slope of the curves suggest that there are many more tags in the sediments that are present between 2-10 times as compared to the water column sample. This analysis underscores both the

vast richness of the microbial reservoir in marine sediments and the similar composition of the communities among environmental replicates.

Conclusions

The development of pyrosequencing as a technique for deep sequencing of microbial communities has contributed a tremendous amount of new information to our knowledge of the diversity of these systems. Microbial ecologists are now able to use this technology to begin asking questions about the role that diversity plays in understanding ecosystem function. However, the interpretability of these data depends on the magnitude of the variability at local scales. The data presented here indicate that over small spatial scales, at least in salt marsh sediments, the microbial community is surprisingly homogeneous. Individually collected sediment cores had similar estimates of richness and diversity, and similarity indices calculated from sequence information from all the individually collected sediments were of the same magnitude. Furthermore, homogenizing multiple sediment samples in an effort to decrease the variability among individual samples proved unnecessary. The highly similar community structure of the environmental replicates stands in contrast to the wide divergence seen between the sediment samples and an outgroup sample collected from a nearby water column. The pyrosequencing method was able to easily differentiate this outgroup from the sediment samples and provides strong justification for the use of pyrosequencing to assess changes in the diversity of microbial communities along environmental gradients or as a result of environmental perturbation.

Acknowledgements

Funding for this research came from the National Science Foundation (DEB- 0717155 to JEH and HGM, and award DBI-0400819 to JLB). Support for the sequencing facility comes

1 from the National Institutes of Health and National Science Foundation (NIH/NIEHS 1 P50
2 ES012742-01 and NSF/OCE 0430724-J Stegeman PI to HGM and MLS) and a grant from the
3 WM Keck Foundation (to MLS). Salary support was contributed by a fellowship from the
4 Princeton University Council on Science and Technology to JLB. Anonymous reviewers
5 improved the quality of this manuscript.

6 **References Cited**

7 Baas Becking LGM. (1934). Geobiologie of Inleiding tot de Milieukunde. The Hague, The
8 Netherlands: W.P. Van Stockum & Zoon.
9 Bowen JL, Ward BB, Morrison HG, Hobbie JE, Valiela I, Deegan LA, Sogin ML. (2011).
10 Microbial community composition in sediments resists perturbation by nutrient
11 enrichment. ISME J.
12 Brazelton WJ, Ludwig KA, Sogin ML, Andreishcheva EN, Kelley DS, Shen C-C, Edwards RL,
13 Baross JA. (2010). Archaea and bacteria with surprising microdiversity show shifts in
14 dominance over 1000-year time scales in hydrothermal chimneys. Proc Natl Acad Sci
15 USA **107**: 1612-1617.
16 Bunge JA, Woodard L, Connolly S (2010) CatchAll: Parametric and nonparametric estimation of
17 species richness and population size. Manuscript in preparation.
18 Carson JK, Campbell L, Rooney D, Clipson N, Gleeson DB. (2009). Minerals in soil select
19 distinct bacterial communities in their microhabitats. FEMS Microbiol Ecol **67**: 381-388.
20 Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. (2009). The bacterial species challenge:
21 making sense of genetic and ecological diversity. Science **323**: 741-746.
22 Galand PE, Casamayor EO, Kirchman DL, Lovejoy C. (2009). Ecology of the rare microbial
23 biosphere of the Arctic Ocean. Proc Natl Acad Sci USA **106**: 22427-22432.

1 Hong S-H, Bunge J, Jeon S-O, Epstein SS (2006) Predicting microbial species richness. Proc
2 Natl Acad Sci USA **103**: 117-122.
3 Howes BL, Howarth RW, Teal JM, Valiela I. (1981). Oxidation-reduction potentials in a salt
4 marsh: spatial patterns and interactions with primary production. Limnol Oceanogr **26**:
5 350-360.
6 Huber JA, Mark Welch DB, Morrison HG, Huse SM, Neal PR, Sogin ML. (2007). Microbial
7 population structures in the deep marine biosphere. Science **318**: 97-100.
8 Huse SM, Huber JA, Morrison HG, Sogin ML, Mark Welch DB. (2007). Accuracy and quality
9 of massively parallel DNA pyrosequencing. Gen Biol **8**: R143.
10 Huse SM, Dethlefsen L, Huber JA, Mark Welsh DB, Relman DA, Sogin ML. (2008). Exploring
11 microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS
12 Genetics 4(11): e1000255.
13 Huse SM, Mark Welch DB, Morrison HG, Sogin ML. (2010). Ironing out the wrinkles in the
14 rare biosphere through improved OTU clustering. Environ Microbiol
15 DOI:10.1111/j.1463-2920.2010.02193.x.
16 Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. (2009). Wrinkles in the rare biosphere:
17 pyrosequencing errors lead to artificial inflation of diversity estimates. Environ
18 Microbiol **12**: 118-123.
19 Lehmann J, Solomon D, Kinyangi J, Dathe L, Wirick S, Jacobsen C. (2008). Spatial complexity
20 of soil organic matter forms at nanometer scales. Nature Geosci **1**: 238-242.
21 Ley RE, Harris JK, Wilcox J, Spear JR, Miller SR, Bebout BM, Maresca JA, Bryant DA, Sogin
22 ML, Pace NR. (2006). Unexpected diversity and complexity from the Guerrero Negro
23 hypersaline microbial mat. Appl Environ Microbiol **72:** 3685-3695.

1 Lozupone CA, Knight R. (2007). Global patterns in bacterial diversity. Proc Natl Acad Sci
2 USA. **104**: 11436-11440.
3 Magurran AE. (1988). Ecological Diversity and its Measurement. Princeton, New Jersey:
4 Princeton University Press.
5 Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. (2005). Genome
6 sequencing in microfabricated high-density picolitre reactors. Nature **437**: 376-380.
7 Morales SE, Cosart TF, Johnson JV, Holben WE. (2009). Extensive phylogenetic analysis of a
8 soil bacterial community illustrates extreme taxon evenness and the effects of amplicon
9 length, degree of coverage, and DNA fractionation on classification and ecological
10 parameters. Appl Environ Microbiol **75**: 668-675.
11 Quince C, Curtis TP, Sloan WT. (2008). The rational exploration of microbial diversity. ISME
12 J **10**: 997-1006.
13 Roesch LFW, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, Kent AD, et al. (2007).
14 Pyrosequencing enumerates and contrasts soil microbial diversity ISME J **1**: 283-290.
15 Shloss PD Handelsman J. (2005). Introducing DOTUR, a computer program for defining
16 operational taxonomic units and estimating species richness. Appl Environ Microbiol **71**:
17 1501-1506.
18 Schloss PD. (2008). Evaluating different approaches that test whether microbial communities
19 have the same structure. ISME J **2**: 265-275.
20 Schloss PD, Wescott SL, Ryabin T, Hall JR, Hartmann M, et al. (2009). Introducing
21 MOTHUR: open-source, platform-independent, community supported software for
22 describing complex microbial communities. Appl Environ Microbiol **75**: 7537-7541.

1 Sogin ML, Morrison HG, Huber JA, Mark Welch DB, Huse SM, Neal PR, Arrieto JM, Herndl
2 GJ. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere".
3 Proc Natl Acad Sci USA. **103**: 12115-12120.
4 Turnbaugh, PJ, Hamaday M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones
5 WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AE, Knight R, Gordon JI.
6 (2009). A core gut microbiome in obese and lean twins. Nature **457**: 480-484.
7 Valiela I, Teal JM. (1979). The nitrogen budget of a salt marsh. Nature **280**: 652-656.
8 Valiela I, Cole ML. (2002). Comparative evidence that salt marshes and mangroves may protect
9 seagrass meadows from land-derived nitrogen loads. Ecosystems **5**: 92-102.
10 Whitman WB, Coleman DC, Wiebe WJ. (1998). Prokaryotes: The unseen majority. Proc Natl
11 Acad Sci USA. **95**: 6578-658.
12

1 Table 1. Salt marsh bacterial diversity and richness derived from multiple diversity estimators for

2 individual sediment samples, homogenized samples, and a representative water column sample

3 from an adjacent marsh.

	Individual sediment samples		Homogenized sediment samples		Water column sample
	Mean	Std. dev.	Mean	Std. dev.	
3% clusters					
Shannon (H)	7.09	0.07	7.09	0.17	4.21
Observed OTUs	4086	333	4206	202	1841
Chao	7244	643	7733	71	5049
Ace	9474	915	10,027	485	10,357
6% clusters					
Observed OTUs	3277	253	3306	180	1166
Chao	5125	451	5274	157	2216
Ace	6181	594	6321	72	3457
10% clusters					
Observed OTUs	2404	183	2433	175	800
Chao	3336	298	3309	159	1254
Ace	3299	309	3307	140	1592

4

5

Table 2: Best fit and 95% confidence intervals describing the Gaussian curves fit to frequency histograms describing the six replicate environmental samples (ENV1-ENV6), the two homogenized samples (HOM1-HOM2) and the outgroup sample from the adjacent marsh.

	ENV1	ENV2	ENV3	ENV4	ENV5	ENV6	HOM1	HOM2	LSM
Best fit values:									
Amplitude	4580	3719	4077	4428	4263	3753	4406	4114	1056
Mean	4.68	4.57	4.67	4.65	4.66	4.60	4.55	4.82	5.93
Standard Deviation	4.60	4.22	4.46	4.55	4.50	4.36	4.38	4.82	7.40
95% confidence interval:									
Amplitude	4561-4599	3706-3732	4059-4095	4408-4448	4244-4282	3738-3769	4385-4428	4097-4132	1047-1066
Mean	4.57-4.78	4.48-4.67	4.56-4.78	4.54-4.77	4.54-4.77	4.49-4.71	4.43-4.66	4.70-4.94	5.57-6.28
Standard Deviation	4.41-4.79	4.06-4.37	4.27-4.64	4.35-4.75	4.31-4.70	4.27-4.64	4.18-4.58	4.61-5.02	6.70-8.09

Figure Legends

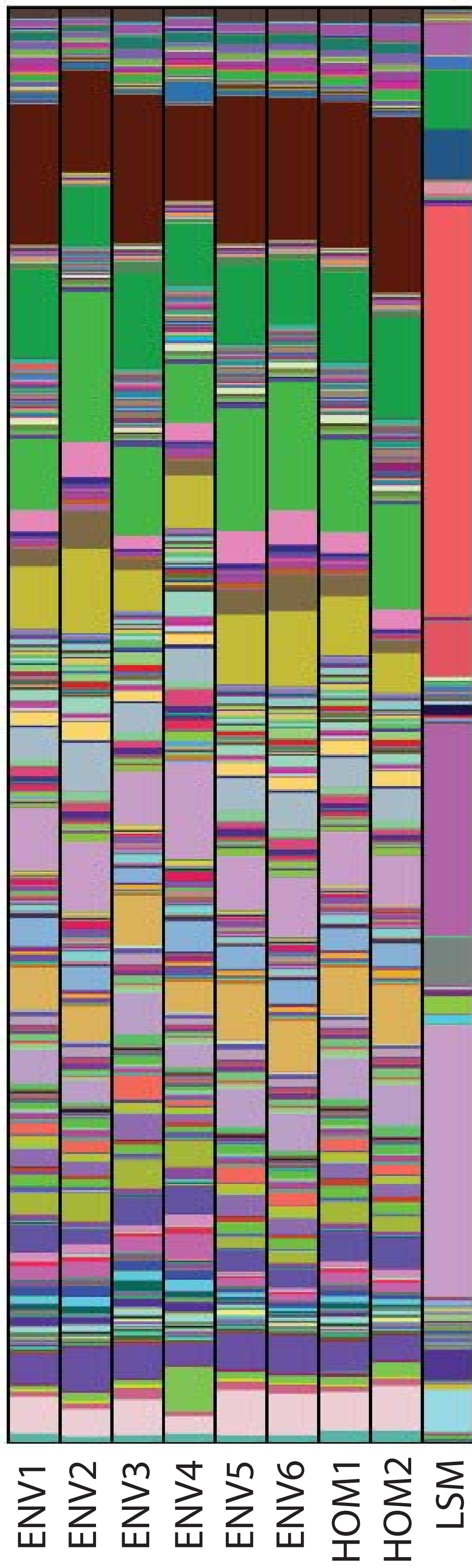
Figure 1. Stacked bar plots of the bacterial species present in sediment samples and in a water column outgroup sample. Fig. 1A represents the relative abundance of all species present and Fig. 1B represents the relative abundance of those species present less than one percent of the time. There are too many species in each sample to make a legend decipherable but the species data are publically available at **www.vamps.mbl.edu**. The data include six sediment samples that were collected individually (ENV1-ENV6), two that were subsampled from pooled sediments (HOM1 and HOM2) and one from the water column draining an adjacent marsh (LSM).

Figure 2. Rarefaction curves for OTUs clustered at 3% (2A), 6% (2B) and 10% (2C) sequence divergence. ENV = individual samples, HOM = homogenized samples, LSM = water column outgroup.

Figure 3. Comparison of Bray Curtis similarity values among individual samples, between individual samples and homogenized samples and between sediment samples and the water column outgroup.

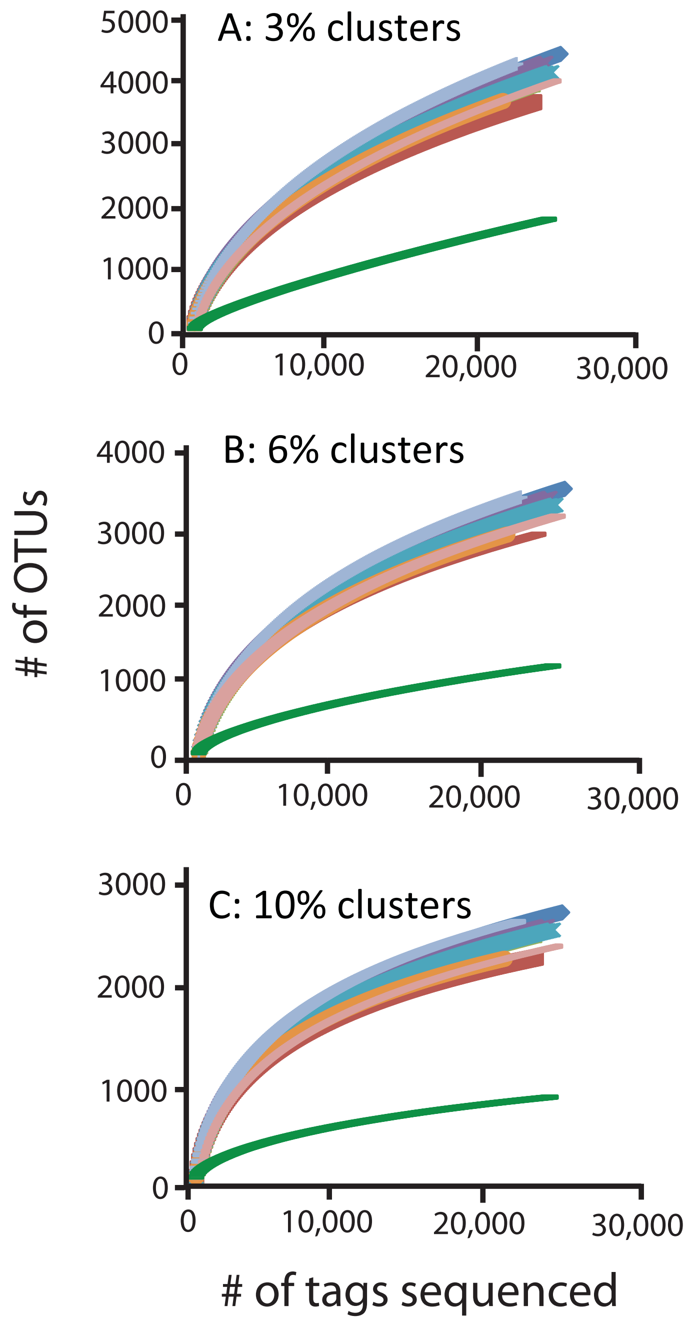
Figure 4. UPGMA determined clustering of sediment environmental replicates compared to the water column outgroup sample. Analysis was performed with all data (5A) and with just those taxa that were present fewer than five times per sample (5B).

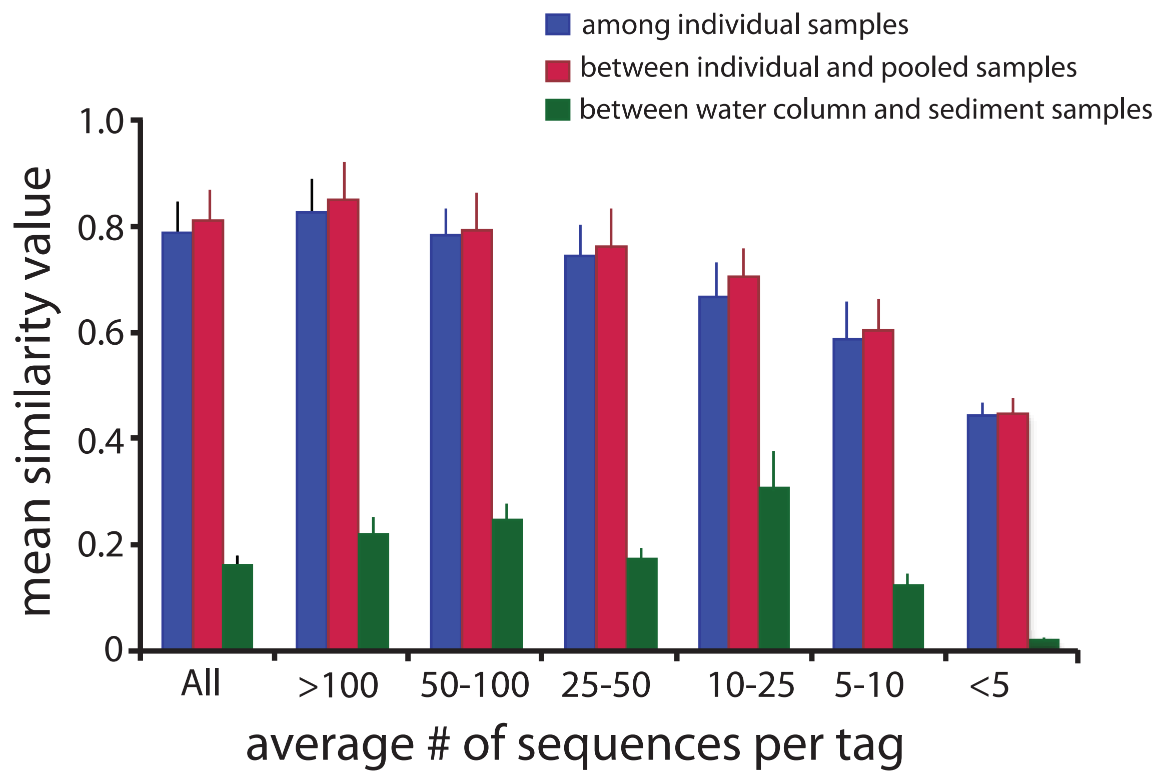
Figure 5. Cumulative frequency of OTUs plotted against the log abundance of sequences per OTU. ENV = individual samples, HOM = homogenized samples, LSM = water column outgroup.


A: all species
ENV1
ENV2
ENV3
ENV4
ENV5
ENV6
HOM1
HOM2
LSM

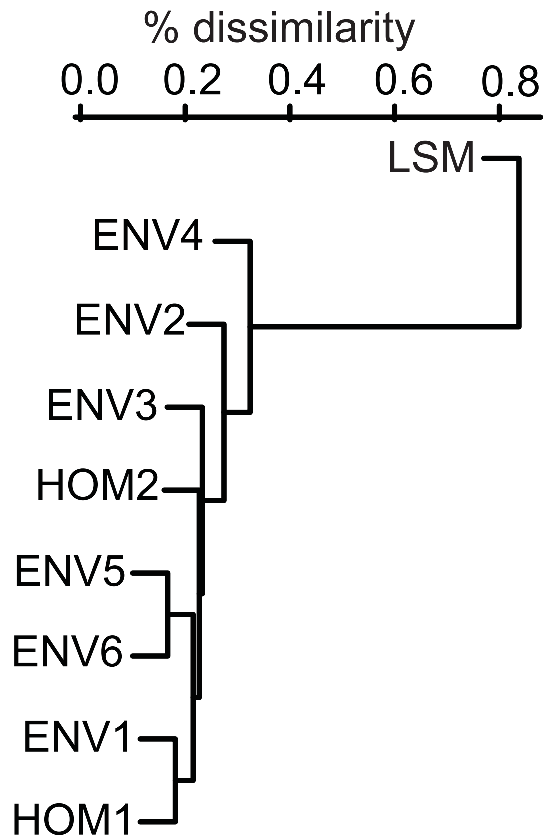
B: species <1% of abundance
ENV1
ENV2
ENV3
ENV4
ENV5
ENV6
HOM1
HOM2
LSM

ENV1 ENV2 ENV3 ENV4 ENV5 ENV6 HOM1 HOM2 LSM

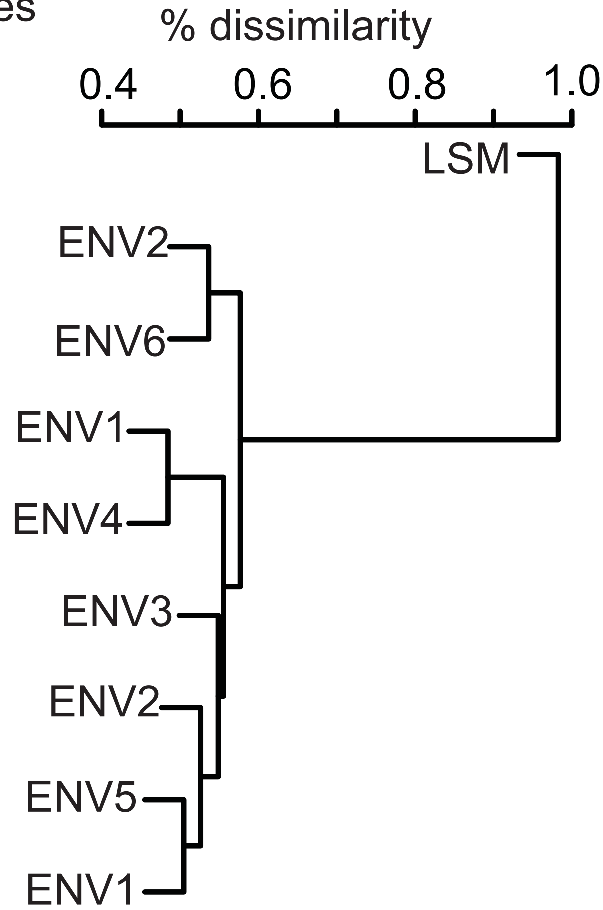


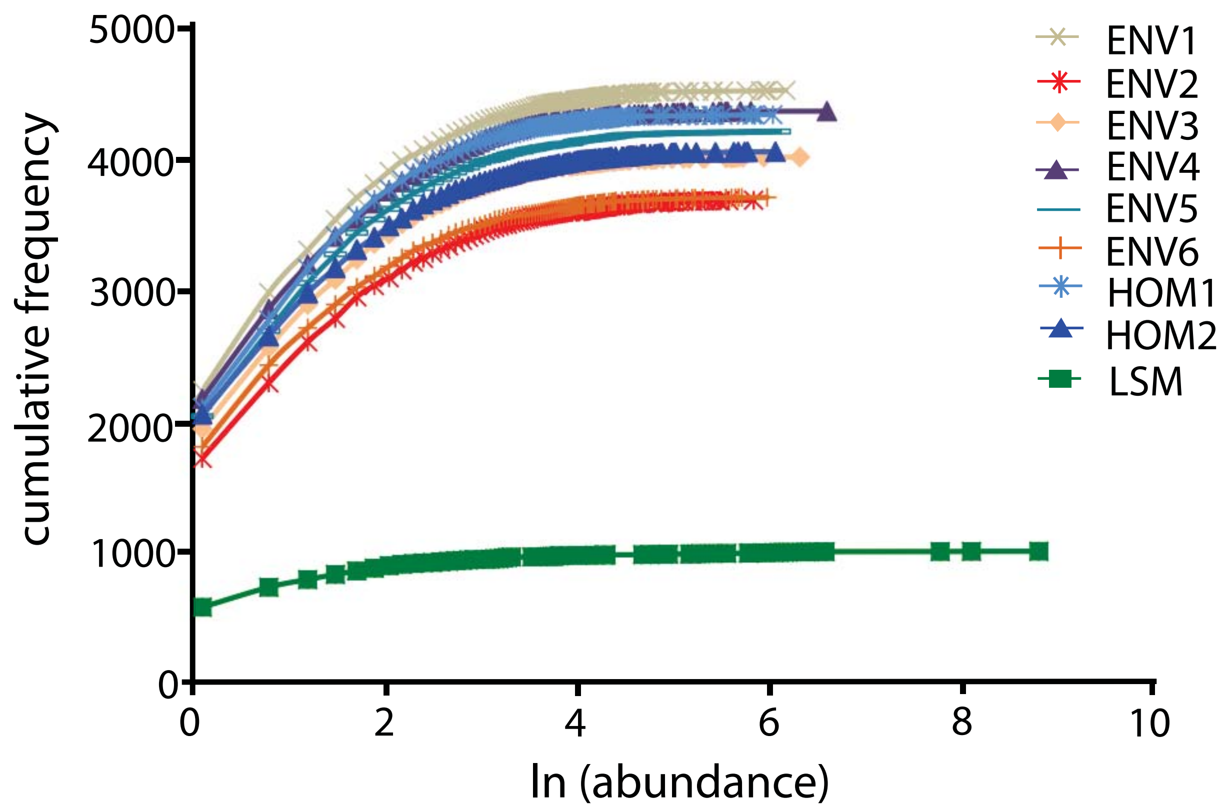

among individual samples
between individual and pooled samples
between water column and sediment samples
1.0
0.8
0.6
0.4
0.2
0
mean similarity value
All
>100
50-100
25-50
10-25
5-10
<5
average # of sequences per tag

A. all sequences



B. present <5 times





5000
4000
3000
2000
1000
0
cumulative frequency
0
2
4
6
8
10
ln (abundance)
ENV1
ENV2
ENV3
ENV4
ENV5
ENV6
HOM1
HOM2
LSM