Supplemental Text

TRFLP Analysis and Quality Control

For fragment analysis at Laragen, an Applied Biosystems size standard ranging from 50-1500 bp and labeled with a proprietary red dye (Rox^{TM}) was added to each sample before injection into the genetic analyzer capillary. Raw fragment data were analyzed with Peak Scanner Software (Applied Biosystems). In order to distinguish signal from noise, a fixed detection threshold of 50 fluorescence units (FU) was used in the Peak Scanner software. This level was chosen to find a balance between reducing the number of peaks from electronic noise in the analysis (Osborn *et al.*, 2000), while not excluding the small reproducible peaks (Dunbar *et al.*, 2001). Additional steps were taken to ensure reproducibility of low-level peaks during the binning step.

After data were exported from Peak Scanner, peak area was normalized to the total area within each sample to allow for comparisons of relative abundance of fragments between samples (Kaplan & Kitts 2004). Peaks were then binned using a dynamic binning script (Ruan *et al.*, 2006). For binning, replicates were combined, and only peaks that occurred in both replicates with an area greater than 0.1% of the total area in each sample were accepted as 'real'. Using this technique, 34 samples were analyzed for variations in microbial eukaryote community composition within and across seep habitats.

Diversity estimates from TRFLP

In many cases, different OTUs can have the same fragment length after digestion; therefore, the discrepancy between known OTUs (based on full-length sequence identity)

and the number of unique fragment sizes after *in silico* digestion can be used to estimate the diversity captured by each enzyme. Comparing the number of unique fragment sizes after *in silico* digestion of our 18S rRNA gene clones to the known number of OTUs in our libraries independently determined if the restriction enzymes used in this study over- or underestimated our 18S rRNA gene sequence diversity estimates. Based on *in silico* digestion, diversity estimates using the restriction enzyme *Hae*III were 30% higher than estimates based on digestion with the restriction enzyme *Bst*UI. This finding was consistent with the empirical TRFLP fingerprint data where the number of unique fragments in the *Hae*III digested samples were approximately 30% higher than the number of unique fragments in the *Bst*UI digested samples. However, even though *Hae*III provided a higher estimate of diversity than *Bst*UI, it still underestimated the diversity in our 18S rRNA clone libraries by approximately 25-30%. Since the patterns in community composition revealed by both enzymes were similar but *Hae*III provided a better estimate of 18S rRNA gene diversity, the data shown and statistical analyses are based on the *Hae*III digested 18S rRNA gene fragments.

We observed good consistency between predicted and actual fragment lengths within a few base pairs (S3). These data, in conjunction with our *in silico* digestion data (not shown), suggest that TRFLP can distinguish between microbial eukaryotes to at least the class level, and in some cases even between clades of closely related organisms within a class. However, due to the potential for run time errors (Kaplan & Kitts 2003) and the fact that even phylogenetically distinct organisms can have similar fragment lengths, we did not attempt to assign peaks to any particular taxa. Rather the TRFLP

fragment patterns were examined for the whole community, and these fingerprints were compared across samples.

Carbonate samples were excluded from TRFLP analysis because the corresponding clone library for these samples was comprised predominantly of metazoan sequences (>90%; data not shown). Therefore, unlike the sediment TRFLP, which could be interpreted as a microbial eukaryote fingerprint (90% microbial eukaryote sequences), the carbonate TRFLP could not. We were only able to include carbonates in the iTAG analysis, because this type of sequencing enabled greater sequencing depth and separation of microbial eukaryote sequences from metazoan sequences.

Comparison of predicted and actual fragment

Since discrepancies between true and observed fragment lengths can vary for a number of reasons (Kaplan & Kitts, 2003; Schutte *et al.*, 2008), 10 clones, representative of the most abundant classes of ciliates recovered in the clone libraries as well as one metazoan clone and one non-ciliate protist clone, were digested with both *Hae*III and *Bst*UI and analyzed in the same manner as the rest of the samples to confirm that their actual base pair location was consistent with *in silico* digestion. There was good consistency between predicted and actual fragment lengths within a few base pairs. The observed offsets (S3.1) may occur as a result of run-time errors and may be related to fragment length and purine content (Kaplan & Kitts 2003).

References:

Dunbar J, Ticknor LO, Kuske CR. Phylogenetic specificity and reproducibility and new method for analysis of terminal restriction fragment profiles of 16S rRNA genes from bacteria communities. *Appl Environ Microbiol* 2001:67:190-197.

Kaplan CW, Kitts CL. Bacterial succession in petroleum land treatment unit. *Appl Environ Microbiol* 2004:70:1777-1786.

Kaplan CW, Kitts CL. Variation between observed and true Terminal Restriction Fragment length is dependent on true TRF length and purine content. *J Microbiol Methods* 2003:54:121-125.

Osborn AM, Moore ERB, Timmis KN. An evaluation of terminal restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. *Environ Microbiol* 2000:2:39-50.

Ruan Q, Steele JA, Schwalbach MS, *et al.* A dynamic programming algorithm for binning microbial community profiles. *Bioinformatics.* 2006:22:1508-1514.

Schutte UME, Abdo Z, Bent SJ, *et al.* Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial *communities. Appl Microbiol Biotechnol* 2008:80:365-380.