Supplemental Text

Limitations of 18S rRNA sequencing

As discussed in the main text, fundamental problems arise from 18S rRNA environmental surveys because of different copy numbers of SSU rRNA genes between taxa (Medinger *et al*., 2010; Gong *et al.,* 2015), extraction methods (Santos *et al.,* 2015), PCR biases (e.g., primer biases, preferential amplification of some taxa) (Medinger *et al.,* 2010; Stoeck *et al.,* 2010; 2014; Adl *et al.,* 2014), sequencing errors (Lee *et al.,* 2012; Kunin *et al*., 2013), as well as amplification of remnant DNA (e.g., e-DNA; Lorenz and Wackernagel 1987, Pawlowski *et al.*, 2011). The biases relevant to the patterns we observed in the data are discussed below.

All samples were extracted in the same way with MoBio's PowerSoil DNA Isolation kit, which has been shown to favor high quality DNA as well as similar yields of DNA from all different soil types (Santos *et al.,* 2013). While deep sequencing approaches have become a more effective way to survey microbial community diversity than traditional cloning, sequencing errors can result in false diversity, especially due to the presence of singletons (e.g., Kunin *et al.,* 2010,; Quince *et al.,* 2011). The discovery of this rare taxa has been termed the 'rare biosphere', but much of this low abundance diversity may in fact be spurious OTUs generated by sequencing noise. In an attempt to remove these spurious sequencing errors from our dataset, singleton OTUs as well as OTUs appearing with less than 0.01% relative abundance were removed during processing (see *Environmental Procedures* for more details).

Another shortcoming of DNA surveys is the detection of microorganisms that are not necessarily viable or metabolically active due to the amplification of remnant DNA

(Lorenz & Wackernagel, 1987, Pawlowski *et al.*, 2011). This can be particularly problematic when using deep-sequencing approaches on DNA isolated from the seafloor below a productive water column, such as the region above Hydrate Ridge, OR. Organisms from the water column (including dead and/or quiescent cells) can accumulate on the seafloor. Sequences from photosynthetic eukaryotes and pelagic metazoans have been frequently retrieved from deep-sea sediments, suggesting that these organisms (or sequences) are contaminants from the photic zone (e.g., Edgcomb *et al.*, 2002; Pawlowski *et al.*, 2011). The most abundant groups in our iTAG data (Bacillariophyta, Dinophyta, Filosa-Thecofilosa [cryomonads and ebriidas], Polycystinea and Apicomplexa [gregarines]) either contain some sort of shell (test or theca) and/or are known to be spore forming. In the case of Apicomplexa, most of the recovered sequences belonged to the gregarines, many of which are parasites commonly found in association with marine crustaceans in addition to forming cysts that are ubiquitously distributed throughout the marine environment (Leander, 2008; Rueckert *et al.*, 2011). It is therefore likely that these dominant groups do not represent active seep organisms, but occur as remnant DNA that has not degraded in the environment. The fact that these groups have high DNA abundance in all samples, regardless of seep heterogeneity (e.g., habitat, activity), further supports the notion that they are not representative of active seep protists.

The primers used for the clone libraries were different than those used for the iTAG data set, therefore, the difference in compositional patterns between the two datasets may be the result of primer biases (Stoeck *et al.,* 2010; 2014). Interestingly though, the community patterns we observed using the clone library primers were more

consistent with the RNA library created by Takishita *et al.* (2010) in a seep environment. This may suggest that the sequences in our clone libraries are more reflective of organisms that are active members of the community. However, different copies, pseudogenes (Santos et al., 2003) and other variants of the 18S rRNA of each organism, can lead to inflated diversity metrics by increasing the number of predicted OTUs (Decelle *et al.,* 2015). Ciliates in particular can often be overrepresented in rRNA surveys because of their shorter SSU rRNA sequences and high SSU rRNA gene copy number (Gong *et al.,* 2013). While this likely contributes to the large number of ciliates sequences in our DNA libraries, microscopy surveys of a methane seep in Monterey Bay also reveal a higher abundance of ciliates in active seep sediments relative to the surrounding deep sea sediments (Buck and Barry, 1998). Other methods such as qPCR (Santos *et al.,* 2015), fluorescent *in situ* hybridization (FISH; Hirst *et al.,* 2011), and metatranscriptomics (Geisen *et al.,* 2015) would provide valuable complementary information for interpreting 18S rRNA sequencing data in future studies.

References:

Adl, S.M., Habura, A., and Eglit, Y. (2014) Amplification primers of SSU rDNA for soil protists. Soil Biol Biochem. 69: 328-342.

Buck, K.R. and Barry, J.P. (1998) Monterey Bay cold seep infauna: quantitative comparison of bacterial mat meiofauna with non-seep control sites. Cah Biol Mar 39: 333-335.

Decelle, J., Romac, S., Sasaki, E., Not, F., and Mahe, F. (2014). Intracellular diversity of V4 and V9 regions of the 18S rNRA in marine protists (radiolarians) assessed by high-throughput sequencing. PLoS One 8: e104297.

Edgcomb, V.P., Kysela, D.T., Teske, A., *et al.* (2002) Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. PNAS 99: 7658-7662.

Geisen, S., Tveit, A.T., Clark, I.M., Richter, A., Svenning, M.M., Bonkowski, M., and Urich, T. (2015). Metatranscriptomic census of active protists in soils. ISME J 9: 2178-2190.

Gong, J., Dong, J., Liu, X., Massana, R. (2013) Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of Oligotrich and Peritrich ciliates. Protist 164: 369-379.

Hirst, M.B., Kita, K.N., Dawson, S.C. (2011). Uncultivated microbial eukaryote diversity: a method to link ssu rRNA gene sequences with morphology. PLoS One. 6: 28158.

Kunin,V., Engelbrektson, A., Ochman, H., Hugenholtz, P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ Microbiol 12: 118-123.

Leander, S.B. (2008) Marine gregarines - evolutionary prelude to the apicomplexan radiation? Trends Parasitol 24: 60-67.

Lee, C.K., Herbold, C.W., Polson, S.W., Wommack, K.E., Williamson, S.J., McDonald, I.R., and Cary S.C. (2012) Groundtruthing next-gen sequencing for microbial ecology-biases and errors in community structure estimates from PCR amplicon pyrosequencing. PLoS ONE 7: e44224.

Lorenz, M.G., and Wackernagel, G. (1987) Adsorption of DNA to sand and variable degradation rates of adsorbed DNA. Appl Environ Microbiol 67: 2354-2359.

Medinger, R., Nolte, V., Pandey, R.V., Jost, S., Ottenwälder, B., Schlötterer, C., et al. (2010). Diversity in a hidden world: potential and limitation of next generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. Mol Ecol 19: 32-40.

Pawlowski J, Christen R, Lecroq B, et al. Eukaryotic richness in the abyss: insights from Pawlowski, J., Christen, R., Lecroq, B., Bachar, D., Shahbazkia, H.R., Amaral-Zettler, L., and Guillou, L. (2011) Eukaryotic richness in the abyss: insights from pyrotag sequencing. PLOS One 6: e18169.

Ruecker, S., Simdyanov, T.G., Aleoshin, V.V., *et al.* (2011) Identification of a divergent environmental DNA sequence clade using the phylogeny of gregarine parasites (Apicomplexa) from crustacean hosts. PLOS One 6: e18163.

Santos, S.S., Nielsen, T.K., Hansen, L.H., and Winding, A. (2015) Comparison of three DNA extraction methods for recovery of soil protist DNA. J Microbiol Methods 115: 13-19.

Stoeck, T., Bass, D., Neble, M., Christen, R., Jobes, M.D.M, Breiner, H-W. and Richards, T.A. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Molec Ecol 19: 21-31.

Stoeck, T., Breiner, H-W., Filker, S., Ostermaier, V., Kammerlander, B., and Sonntag, B. (2014). A morphogenetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology. Environ Microbiol 16: 430-444.

Takishita, K., Kakizoe, N., Yoshida, T., and Maruyama, T. (2010) Molecular evidence that phylogenetically diverged ciliates are active in microbial mats of deep-sea cold-seep sediment. J Eukaryot Microbiol 57: 76-86

Quince, C., Lanzen, A., Davenport, R.J., and Turnbaugh, P.J. (2011) Removing noise from pyrosequenced amplicons. BMC Bioinformatics 12: 38.