Effect of Oxygen Minimum Zone Formation on Communities of Marine Protists

William Orsi[1, 2§], Young C. Song[3, 4§], Steven Hallam[3,4], Virginia Edgcomb[1*]

[1] Department of Geology and Geophysics, Woods Hole Oceanographic Institution

[2] Center for Deep Energy Biosphere Investigations

[3] Department of Microbiology & Immunology, University of British Columbia, Canada

[4] Graduate Program in Bioinformatics, University of British Columbia, Canada

Classification: *Biological Sciences, Environmental Science, Systems Biology*

Pages: 29, Figures: 6, Tables: 2, Supplementary: 5 figures and 2 tables

Abstract: 200 words, 1233 characters

Text: 4,970 words, 28,963 characters

[§]These authors contributed equally to this work

*To whom correspondence should be addressed:

Woods Hole Oceanographic Institution

266 Woods Hole Road MS# 52 Woods Hole, Ma. 02543 USA

Office: (508) 289 3734 e-mail: vedgcomb@whoi.edu

Keywords: protists, diversity, anoxic, oxygen minimum zone, 18S rRNA approach

Running title: *Diversity and ecology of microbial eukaryotes*

Abstract

Changes in ocean temperature and circulation patterns compounded by human activities are leading to oxygen minimum zone expansion with concomitant alteration in nutrient and climate active trace gas cycling. Here, we report the response of microbial eukaryote populations to seasonal changes in water column oxygen-deficiency using Saanich Inlet, a seasonally anoxic fjord on the coast of Vancouver Island British Columbia, as a model ecosystem. We combine small subunit ribosomal RNA gene sequencing approaches with multivariate statistical methods to reveal shifts in operational taxonomic units during successive stages of seasonal stratification and renewal. A meta-analysis is used to identify common and unique patterns of community composition between Saanich Inlet and the anoxic/sulfidic Cariaco Basin (Venezuela) and Framvaren Fjord (Norway) to show shared and unique responses of microbial eukaryotes to oxygen and sulfide in these three environments. Our analyses also reveal temporal fluctuations in rare populations of microbial eukaryotes, particularly anaerobic ciliates, that may be of significant importance to the biogeochemical cycling of methane in oxygen minimum zones.

Introduction

Dissolved oxygen (O_2) concentration is a primary driver of nutrient and energy flow patterns within marine ecosystems (Diaz and Rosenberg 2008; Diaz *et al.* 2009). Oxygen-deficiency selects for microbial groups capable of utilizing alternative respiratory substrates including nitrate (NO_3^-), nitrite (NO_2^-), manganese (Mn), iron (Fe), sulfate (SO_4^-) or carbon dioxide (CO_2) (Zehnder and Stumm 1988). Within oxygen-deficient waters the use of NO_2^- or NO_3^- as alternative electron acceptors results in the production of nitrous oxide (N_2O) and dinitrogen gas (N_2) (Lam and Kuypers, 2010). Similarly reduction of SO_4^- and CO_2 under anoxic conditions results in the production of toxic hydrogen sulfide (H_2S) (Teske 2010) and methane (CH_4), respectively (Naqvi *et al.* 2010). Recent studies of microbial community structure and systems metabolism within marine oxygen minimum zones (OMZs) indicate a versatile capacity to produce and consume climate active trace gases (defined as gases making up <1% of the atmosphere that absorb the electromagnetic energy resulting from reflection of solar radiation from the Earth's surface) or to limit accumulation of H_2S within the surrounding water column (Lavik *et al.* 2009; Walsh *et al.* 2009; Canfield *et al.* 2010; Zaikova *et al.* 2010). Although prokaryotic (bacteria and archaea) microorganisms are the primary drivers of these biogeochemical transformations (Arrigo, 2005), it is likely that microbial eukaryotes (protists) act as important biological controls through predation on (Taylor, 1982), parasitism of (Chambouvet *et al.*, 2008), and symbioses with (Edgcomb *et al.*, 2010), different microbes.

Information regarding the influence of OMZ formation on marine protists remains incremental, although recent studies using next generation sequencing approaches point to complex and diverse protistan communities in these habitats (Stoeck *et al.* 2009; Behnke *et al.* 2010). Restructuring of these protists' communities in response to changing levels of water column oxygen-deficiency likely influences prokaryotic population structure and activities with resulting feedback on nutrient and climate active trace gas cycling. Several recent studies

provided insight into the biogeography, species richness, endemicity, and habitat specialization of protists along oxygen gradients (Behnke *et al.* 2006; Zuendorf *et al*. 2006; Behnke *et al.* 2010; Edgcomb *et al.* 2011a; Edgcomb *et al.* 2011b; Orsi *et al.* 2011b). Work in the Cariaco Basin revealed a specialization of many protistan taxa to different biogeochemical niches and sites in the basin (Orsi *et al.* 2011b) and the estimated protistan species richness there was found to be exceptionally high (Edgcomb *et al.* 2011a). A recent study of Framvaren Fjord found evidence for seasonal fluctuations in protistan community structure (Behnke *et al.* 2010). Although major taxonomic lineages remained consistent throughout the time course of the study, subgroup diversity changed extensively from season to season among and between sampling depths consistent with dynamic recruitment from diverse low abundance populations. A similar finding was also reported by an additional study of microbial eukaryotes in the western North Atlantic and eastern North Pacific oceans (Caron *et al.*, 2009).

In the present study, we monitored changes in protistan community structure in Saanich Inlet, a seasonally anoxic fjord on the coast of Vancouver Island, Canada using small subunit ribosomal RNA gene (SSU rRNA gene, 18S rRNA gene) clone library sequencing. We charted the spatiotemporal variability of protists in relation to dissolved gases and nutrients, and employed multivariate statistical approaches to identify potential relationships between compositional profiles, taxonomic groups and environmental parameters at different stages of water column stratification and renewal. The resulting data sets are compared to related studies in anoxic Cariaco Basin and Framvaren Fjord and used to identify common and unique patterns of community composition.

Results

Community Diversity Measures

We sampled four depth intervals over three years (2006-2008) during the months of February, April, July, and November, representing different water column redox states resulting from seasonal stratification and renewal (Zaikova *et al.* 2010; Walsh *et al.* 2009). A total of 4,987 18S rRNA gene sequences recovered from 19 different clone libraries were analyzed (Table 1). Clustering of these sequences at the 99%, 98%, 95%, and 90% sequence identity thresholds resulted in a total of 1217, 993, 596, and 244 operational taxonomic units (OTUs), respectively. We estimated OTU richness using a statistical tool designed for estimating microbial species richness with a reliable standard error (Hong *et al.* 2006; Jeon *et al.* 2008; Edgcomb *et al.* 2011a; Orsi *et al.* 2011b). At the 99 and 98% sequence identity thresholds, we estimate 13,442 (-/+: 7,963-23,373 CI [95% confidence interval]) and 8,176 (-/+: 4,861-14,333 CI) taxa, respectively. At 95% and 90% sequence identity, we estimate 2,687 (-/+: 1,440-5,778 CI) and 510 (-/+: 376-781 CI) taxa, respectively (Table 2). Using the same statistical tool, we estimated taxonomic richness within the anoxic Framvaren Fjord using recently published 18S rRNA gene datasets from this environment (Behnke *et al.* 2006; Behnke *et al.* 2010). The number of OTUs estimated to exist in Framvaren Fjord at the 99%, 98%, 95%, and 90% sequence identity levels, amount to 28%, 13%, 17%, and 45% of the OTUs predicted for Saanich Inlet, respectively (Table 2). Non-parametric methods estimated similar richness and diversity for Saanich Inlet vs. Framvaren Fjord, although predicted richness was less than in Cariaco. Non-parametric estimates for all 3 sites were lower than parametric estimates (Table S2).

Protistan Community Structure in Saanich Inlet

Canonical correspondence analysis (CCA) showed a clear division of Saanich Inlet protistan communities into different clusters associated with oxic (>90 μM), dysoxic (20-90 μM), suboxic (1-20) μM or anoxic/sulfidic (<1 μM/±sulfide) water column conditions (Figure 1). The anoxic samples from 200 m clustered together, separate from all other samples on the biplot (Figure 1). However, the 200 m sample taken in November grouped with dysoxic samples. A Monte Carlo test of the null hypothesis of no relationship between the OTU distribution and the measured environmental variables oxygen, nitrate, methane, and sulfide had a p-value of 0.01. Multi-response permutation procedure (MRPP) tests of influence of season, depth, oxygen, nitrate, methane, and sulfide on the OTU distribution all yielded p-values less than or equal to 0.01.

Changes in taxonomic representation at 120 and 200 m coincided with the annual renewal cycle in fall and re-stratification of the water column in summer (Table 1) (Walsh *et al*., 2009). Only 10 and 3% of OTUs were detected in both July and November, respectively, at these depths. At 120 m in July, clone libraries were dominated by Stramenopile-affiliated sequences, while in November an increase in Cercozoan- and Dinophyceae-affiliated sequences was observed (Figure 2, 6). At 200 m in July, 66% and 11% of sequences were affiliated with the Ciliophora and Euglenozoa, respectively (Figure 2, 6), while in November, 70% of sequences were affiliated with the Dinophyceae, 80% of which were affiliated with the sub-group Syndiniales (Figure 2, 6). The majority of sequences recovered from anoxic waters exhibited low (< 92%) sequence identities with their closest described relatives (Table S1) and formed new lineages based on phylogenetic trees (Figure 3, 4). The uncultured environmental sequences with the highest identities to many of these sequences were recovered from oxygen-deficient marine waters from the Cariaco Basin (Stoeck *et al.* 2003), Framvaren Fjord (Behnke *et al.* 2006, 2010), and Mariager Fjord (Denmark) (Zuendorf *et al.* 2006).

151 At 10 m, Dinophyceae- and Stramenopile-affiliated sequences dominated during all
152 seasons (Figure 2, S4). The majority (90%) of Dinophyceae-affiliated sequences were related to
153 the parasitic Syndiniales. The same observation was made at 100 m in February, April, and
154 November while in July, radiolarian-affiliated sequences were most abundant (Figure 2, S4).
155 The number of cercozoan-affiliated sequences increased in April at 10 and 100 m, but were less
156 abundant during the rest of the year (Figure 2, 6). The occurrence of larger-sized protists (such
157 as the Radiolaria and Cercozoa) in pico-eukaryote size clone libraries has also been reported in
158 previous studies (Not *et al*., 2009)
159
160 *Taxonomic Relationships between Saanich Inlet, Framvaren Fjord, and Cariaco Basin*
161 To better constrain protistan community structure and dynamics to changing levels of
162 oxygen and sulfide on a global scale, we compared available 18S rRNA sequence datasets from
163 Saanich Inlet, Cariaco Basin and Framvaren Fjord. Overall, taxonomic assignments for the
164 Cariaco Basin and Framvaren Fjord (Figure S1-2) datasets were congruent with previous studies
165 (Edgcomb *et al*. 2011a; Orsi *et al*. 2011b, Behnke *et al*. 2010). Sequences affiliated with the
166 Ciliophora, Euglenozoa, Choanoflagellata, Stramenopiles, Fungi, and Dinophyceae were well
167 represented from anoxic samples in all three sites (Figure 6, S1-5). However, many taxa were
168 differentially represented. Examples include Stramenopile-affiliated sequences that were more
169 prevalent in Cariaco Basin and Framvaren Fjord than Saanich Inlet (S1-4), as well as
170 Polycystinea- and fungal-affiliated sequences that were more abundant in Cariaco Basin relative
171 to Saanich Inlet and Framvaren (Figure 2, 6, S1-5). Sequences affiliated with the Stramenopiles,
172 Cercozoa, Dinophyceae, Polycystinea, and Acantharea were represented in oxygenated samples
173 from both the Cariaco Basin and Saanich Inlet (Figure 2, 6, S1-5).
174 Combined principle component (PCA) and hierarchical cluster analyses of the Saanich
175 Inlet, Cariaco Basin (Edgcomb *et al.* 2011a; Orsi *et al.*, 2011b) and Framvaren Fjord (Behnke *et*

176 *al*. 2010) datasets revealed biogeographic and niche-specific clustering patterns. For the most
177 part, protistan communities clustered by location according to depth and oxygen concentration.
178 The majority of oxic, dysoxic and suboxic samples from Saanich Inlet formed a nested series in
179 one cluster, with anoxic Cariaco Basin and Framvaren Fjord samples, forming independent
180 clusters. Interestingly, anoxic/sulfidic samples from Saanich Inlet and one anoxic deep sample
181 from the Cariaco western sub-basin (site BC) clustered with the hyper-sulfidic Framvaren Fjord
182 samples (Figure 5).
183
184 *Fluctuations in rare microbial populations*
185 Our comparison of microbial populations during stratification and renewal suggests
186 selection for growth of many low abundance taxa once the preferred conditions arise. Examples
187 include an increase in the number of Ciliate- and Stramenopile-affiliated sequences recovered at
188 200 m during periods of anoxia (Figure 2, 6 and S4). One Stramenopile-affiliated OTU was
189 detected 228 times in July and 11 times in November (Figure S4). Furthermore, an increase in
190 Dinophyceae and Cercozoan-affiliated sequences at 120 m and 200 m was observed after the
191 oxygenated renewal event in autumn (Figure 2, 6 and S4). Several Dinophyceae-affiliated OTUs
192 detected in July and November at 200 m increased significantly in size in November after the
193 renewal event (Figure S4).
194
195 **Discussion**
196 Saanich Inlet provides a model ecosystem for studying microbial community responses to
197 changes in dissolved oxygen concentrations *i.e.* oxygen minimum zone formation, because the
198 bottom 80 m of the water column (120-200 m) annually fluctuates between oxygenated and
199 reduced states (Zaikova *et al.* 2010). In summer months, restriction of movement by a shallow
200 glacial sill at the mouth of the inlet results in water mass stability. High primary production at

the surface fuels aerobic respiration of microorganisms, which in turn leads to a progressive loss of dissolved oxygen as this organic matter sinks. The resulting chemical transformation of the water column produces a redoxcline between 100 and 120 m with anoxic waters stretching from 120 m depth to the seafloor. During autumn-winter, nutrient-rich oxygenated waters flow over the glacial sill shoaling anoxic waters upward.

We applied an 18S rRNA gene sequencing approach to assess the influence of oxygen minimum zone formation on community structure of marine protists in Saanich Inlet over a three-year period. Phylogenetic and multivariate statistical analyses of the resulting dataset revealed defined shifts in operational taxonomic units that correlate with changes in water column chemistry. Taxonomic and multivariate comparisons, as well as comparisons of statistically estimated richness, between the Saanich Inlet, the anoxic Cariaco Basin and Framvaren Fjord provide insights into how different communities of marine protists respond to anoxic and low oxygen conditions.

However, it is important to note that differences in sample processing across studies add potential bias to this meta-analysis. Water samples in the current survey of Saanich Inlet were filtered through a 2.7 μm prefilter, whereas no prefilter was applied in the studies of the Cariaco Basin and Framvaren Fjord. Furthermore, biomass from Saanich Inlet waters was collected on filters with a pore size of 0.22 μm, while in Cariaco and Framvaren filters with pore sizes of 0.65 and 0.45 μm were used. Additionally, different primer combinations were used in all three studies, increasing potential amplification biases. Given these biases, we argue that while the current meta-analysis reveals broad patterns of similarity in protistan community responses to low oxygen conditions, interpretation of site-specific differences must be made with extreme caution.

It is known that oxygen and sulfide concentrations have a strong influence on microbial distributions in anoxic marine environments such as the Cariaco Basin (Taylor *et al*. 2001; Li *et*

226 *al.* 2008; Lin *et al.* 2008; Edgcomb *et al*. 2011a; Orsi *et al.* 2011b) and the Framvaren Fjord
227 (Behnke *et al.* 2006; Stoeck *et al.* 2009; Stoeck *et al.* 2010). These findings are validated by our
228 CCA, MRPP, and Monte Carlo analyses that indicate oxygen and sulfide, as well as methane, to
229 be the primary drivers of protistan distribution in Saanich Inlet (see Results and Figure 1). Ours
230 is the first investigation into the influence of methane on the distribution of protistan
231 communities in low oxygen marine environments and our results suggest that methane may have
232 a stronger influence than sulfide, based on the differential length of the methane and sulfide
233 vectors in the CCA (Figure 1). Thus, the selective influence of methane on protistan
234 communities should be considered in future studies of OMZs. Interestingly, methanogenic
235 archaea were not detected in Saanich Inlet water samples that incorporated a 2.7 μm prefilter
236 (Zaikova *et al.* 2010). While diffusive flux from underlying sediments is one likely source of
237 methane in the water column, new methane production originating from methanogens associated
238 with the anaerobic ciliates detected in our study may also contribute to methane accumulation in
239 basin waters. Methanogenic symbionts of ciliates are well-known (e.g. Embley and Finlay, 1993;
240 Fenchel and Finlay, 1995; van Hoek *et al*. 2000; Edgcomb *et al*. 2011c). This finding suggests
241 that such symbioses may have the potential to contribute to climate active trace gas cycling in
242 low oxygen and anoxic marine environments.

243 The CCA analysis also suggests that unique assemblages of protists inhabit different
244 niches along the redoxcline in Saanich Inlet throughout the year depending on the intensity of
245 renewal events. The November 200 m sample collected during renewal does not group with the
246 other 200 m samples from April, February, and July on the biplot (Figure 1). Rather, this sample
247 groups with dysoxic samples on the biplot, most likely reflecting the physical movement of an
248 oxygenated water mass into basin waters (Table 1). This is supported by the minimal overlap in
249 OTUs observed at 200 m between July and November and the MRPP analysis of the influence of
250 season and depth. These findings confirm that OMZ formation in Saanich Inlet has a strong

influence on the protistan community, with different OTUs being selected for as a result of the annual stratification and renewal cycle.

At 200 m during November and February, the number of 18S rRNA gene sequences affiliated with the Syndiniales and Stramenopiles increased relative to July and April (Figure 2, S4). The appearance of these groups is not unexpected as the upwelling water originates from coastal marine sources, a habitat in which representatives of the Syndiniales and Stramenopiles have been detected previously (Guillou *et al.,* 2008, Lin *et al*. 2006; Massana *et al.* 2006). The majority of Stramenopile sequences were affiliated with the uncultured Marine Stramenopiles (MAST), which have been shown to exhibit a range of trophic modes and specificity for different prey species and sizes (Massana *et al.*, 2009). Thus, as MAST stramenopiles have been detected in the anoxic waters of Framvaren Fjord and Cariaco Basin (Behnke et al., 2010a; Orsi et al., 2011b), they likely play a role in regulating the abundances of microbial populations that mediate biogeochemical cycling in OMZs.

After re-stratification of the water column in July (Table 1), the majority of sequences in the anoxic portion of the water column were affiliated with the Ciliophora and Euglenozoa (Figure 2, 6). Similar observations have been made in the Cariaco Basin where waters below the oxic/anoxic interface contain over twice the number of ciliate- and euglenozoan-affiliated operational taxonomic units (OTUs) relative to oxygentated waters (Orsi *et al*. 2011b). However, as no other studies of eukaryotic communities in seasonal OMZs have been conducted, we can only speculate at this point that such shifts may represent seasonally-related succession within the protist community. Furthermore, we can only speculate that the dominant ciliates and euglenozoans found at 200 m in July represent species that survive periodic exposure to oxygen during renewal events by becoming less active (and less numerous) until favorable conditions are restored. Overall, these survey results indicate that Ciliophora and Euglenozoa are selected for in Saanich Inlet during periods of water column stratification. Both contain many species of

anaerobes and microaerophiles, and indeed most of the closest described relatives of the sequences affiliated with these groups (i.e. *Calkinsia*, *Cyclidium, Strombidium,* and *Nyctotherus*) fall into this category (Table S1).

Our phylogenetic analyses of ciliate and euglenozoan-affiliated OTUs recovered from anoxic waters (Figures 3 and 4), as well as the relatively low (< 92%) identities of most of these sequences to their closest described species in public databases, suggests that OMZ formation in Saanich Inlet selects for novel lineages within these phyla. The new Symbiontida-affiliated lineages (Figure 4) with low (<90%) identities to the euglenozoan *Calkinsia aureus* (Yubuki *et al*. 2009) (Table S1) may correspond to protists exhibiting symbioses with bacteria. *C. aureus* is a euglenozoan flagellate recently recovered from the Santa Barbara Basin (California) (Bernhard *et al.* 2000; Yubuki *et al*. 2009; Edgcomb *et al*. 2010) with a cortex that is completely covered by epibiotic bacteria belonging to the *Arcobacter*, a group that includes chemoautotrophs and chemoorganotrophs capable of nitrate reduction and sulfide oxidization (Edgcomb *et al.* 2010). Phylogenetic analyses of ciliate-affiliated sequences reveal two clades branching basal to the novel ciliate class Cariacotrichea (Orsi *et al.* 2011a) recovered from the Cariaco Basin, suggesting this newly discovered taxon to be highly diverse.

Despite the use of a 2.7 µm prefilter, sequences from ciliates larger than 2.7 µm were recovered in Saanich Inlet samples, indicating the presence of DNA from lysed cells. Aside from their high copy number of ribosomal RNA genes (e.g. Prescott 1994), the increase in abundance of ciliate-affiliated sequences may be induced by oxygen depletion and accumulation of sulfide that selects for ciliates adapted to such conditions. Also, ciliates, being significant grazers of bacteria, may be responding to spikes in prey species that occur after OMZ formation. Ciliates can act as primary bacterial grazers (Sherr and Sherr 2002) and may regulate abundances of denitrifying and anammox bacteria responsible for the production of nitrous oxide that are known to exist in the Inlet at this depth (Zaikova *et al*. 2010). Because nitrous oxide is a

greenhouse gas and causes ozone depletion, a potentially important relationship may exist between the abundances of ciliate grazers, denitrifying and anammox bacteria, and the release of nitrous oxide from the surrounding water column.

Our dataset reveals a possible linkage between environmental perturbations and a response from microbial populations present at relatively low abundances, also termed "the rare biosphere" (Pedros-Alio 2007). While sequence abundance in clone libraries is by no means an exact indicator of cell numbers at the time of sampling, gross differences can be used as a proxy for cellular abundance (e.g. Not *et al*., 2009). Fluctuations in sequence representation within OTUs affiliated with the Stramenopiles, Dinophyceae, Ciliophora, and Euglenozoa (see results and Figures 2, 6, S4) all suggest that some temporally-rare microbial populations become abundant in Saanich Inlet after preferred conditions arise. The potential impact of seasonally-abundant ciliate grazers and heterotrophic flagellates on the release of biologically-produced nitrous oxide during periods of anoxia would serve as a prime example of the potential ecological importance of such a 'rare biosphere'.

Comparisons of parametric and nonparametric richness estimates for the Saanich Inlet, Cariaco Basin and Framvaren Fjord datasets need to be interpreted with caution because of methodological differences in sample collection. Although potentially biased, this comparison suggests that Cariaco Basin contains roughly twice the number of species- (defined as OTUs sharing 99-98% sequence identity, see Caron *et al.* 2009 for discussion) and genus-level (defined as OTUs sharing 95-90% sequence identity) taxa (Edgcomb *et al.* 2011) than are estimated for the Saanich Inlet and roughly ten times the number of such taxa from the Framvaren Fjord (Table 2). Non-parametric estimations, which typically underestimate microbial diversity (Chao and Bunge, 2002), expectedly resulted in smaller predictions (Table S2), with differences between the three locations being proportional to the parametric richness estimates. Non-parametric richness estimations (Chao, Simpsons, and Shannon indices) indicate a general trend

of decreasing microbial richness with increasing depth (and anoxia) in the Saanich Inlet water column (Table S2). The same trend is apparent in the empirically registered OTU richness (Table S2). Together with phylogenetic and multivariate analyses (Figures 1, 3 and 4), these data suggest that the anoxic waters of Saanich Inlet house a genetically distinct, albeit less rich, protistan community relative to that present in the shallower, oxygenated waters. Differences in estimated richness between the Framvaren, Saanich Inlet, and Cariaco (Table 2) are supported by a multivariate comparative analysis (Figure 5), indicating that the two fjord communities are more similar to one another under anoxic/sulfidic conditions than either is to Cariaco communities. However, differences in library size may influence diversity estimates (Chao and Bunge 2002). Consequently, the relatively low taxon richness detected in Saanich and Framvaren may be impacted by undersampling.

The differences between the communities at the three locations are likely due to several parameters that differ between oceanographic provinces. One likely reason for the variation in richness is the large difference in size between the Cariaco Basin, Framvaren Fjord and Saanich Inlet. Protistan communities in the eastern and western sub-basins of the Cariaco contain widely divergent assemblages, and this phenomenon is likely driven in part by differences in primary production, riverine inputs and trophic responses to differential prey items (Orsi *et al.* 2011b). The presence of additional niches that are attributed to the larger size of the Cariaco Basin may permit higher protist diversity relative to Saanich Inlet and Framvaren Fjord. Second, the difference in climate and seawater temperatures between Saanich Inlet, Framvaren Fjord, and Cariaco may contribute to the differences in richness, as temperature has been shown to be a significant driver of the diversification of marine microbial populations (Rutherford *et al.* 1999; Fuhrman *et al.* 2008). Also, unlike Saanich Inlet and Cariaco, the oxic/anoxic interface of Framvaren Fjord lies in the photic zone and contains significantly higher sulfide levels (Table 1). The Cariaco Basin has remained anoxic for millions of years (Schubert, 1982) and experiences

only limited oxygen intrusion events (Lin *et al.* 2008). This timeframe has likely allowed for a higher amount of speciation and diversification and could explain, in part, the higher richness of this environment. This may also contribute to the differential representation of taxonomic groups recovered in Cariaco, such as Fungi and Polycystinea (Supplementary Figure 2), as well as the separation of the fjord communities from Cariaco on the PCA biplot (Figure 5).

Our analysis revealed common and unique responses of protists to water column oxygen-deficiency in space and time. Similar to studies of Framvaren Fjord and Cariaco Basin, we observed that protistan taxon representation in Saanich Inlet changed in response to environmental perturbations associated with altered redox status. However, differences in taxonomic representation and diversity estimations between the three locations indicated patterns of endemism not fully explained by sampling and detection biases alone. At the same time, we obtained evidence for temporal fluctuations in rare protistan populations, particularly anaerobic ciliates, that may be of significant importance to biogeochemical cycling, *e.g.* methane, within OMZs.

Experimental Procedures

Sample Collection and Processing

Samples from Saanich Inlet were collected and processed as described previously (Walsh and Hallam, 2011; Zaikova *et al*., 2010) as part of a monthly monitoring program in Saanich Inlet aboard the *MSV John Strickland* (JS) or *CCGS John P. Tulley* (JPT). Briefly, water samples and environmental parameter data were collected from station S3 (48°35.30N, 123°30.22W). Approximately 20 L from 10, 100, 120 and 200 m depth intervals representing oxic, dysoxic, suboxic and anoxic regions of the water column was prefiltered through 2.7 μm GF/D prefilters onto 0.22 μm Sterivex filters for downstream molecular analyses. Biomass samples were accompanied by higher resolution physical and chemical data spanning sixteen depth intervals including cell counts, temperature, salinity, oxygen (O_2), nitrate (NO_3^-), phosphate (PO_4^{3-}), silicate (SiO_4^-), nitrite (NO_2^-), nitrous oxide (N_2O), ammonia (NH_4^+), carbon dioxide (CO_2), methane (CH_4), and hydrogen sulfide (H_2S) concentration measurements. Aspects of this workflow are presented as a series of on-line video protocols including: (1) seawater collection and environmental sampling (URL: http://www.jove.com/index/Details.stp?ID=1159) (Zaikova *et al*., 2009), (2) small volume filtration (URL: http://www.jove.com/index/Details.stp?ID=1163) and large volume filtration (URL: http://www.jove.com/index/Details.stp?ID=1161) (Walsh *et al*., 2009a; Walsh *et al*., 2009b), (3) genomic DNA extraction and purification (URL: http://www.jove.com/index/Details.stp?ID=1352) (Wright *et al.,* 2009). Additional information relating to hydrology and water column chemistry in Saanich Inlet is available through the Saanich undersea array, a streaming cabled observatory node situated on the seafloor near the mouth of the Inlet (URL: http://www.venus.uvic.ca/locations/saanich_inlet.php).

392 *PCR Amplification of 18S rRNA genes*

393 DNA extracts from 10, 100, 120 and 200 m depth intervals were amplified using small subunit
394 ribosomal DNA primers targeting the eukaryotic domain: Euk515F 5' -
395 GTGCCAAGCAGCCGCGGTAA) and Euk1209R 5' GACGGGCRGTGWGTRCA) under the
396 following PCR conditions. PCR conditions: 2 min at 95°C followed by 20 cycles of 95°C for 40
397 sec, 55°C for 30 sec, 72°C for 90 sec and final extension of 7 min at 72°C. Each 50 µl reaction
398 contained 1 µl of template DNA and 1 µl each 0.4 uM forward and reverse primer added to a
399 PCR Master Mix (Stratagene, Cat #600640). Reactions were aliquoted into 3 x 15 ul reactions
400 prior to PCR (to minimize bias) and re-pooled after PCR. For specific details see URL:
401 http://my.jgi.doe.gov/general/protocols/SOP_16S18S_rRNA_PCR_Library_Creation.pdf
402
403 *Clone Library Construction and Sequencing*
404 Resulting amplicons were gel purified using the MiniElute gel extraction Kit (Qiagen) according
405 to the manufacturer's instructions. Ligation, transformation and sequencing steps were
406 performed as described in Zaikova and colleagues (2009). The number of resulting transformants
407 per ligation ranged between ~100,000 to 800,000 colony forming units (CFUs). One 384-well
408 plate per sample ligation was sequenced with variable success using M13F and M13R primers as
409 described at http://jgi.doe.gov/sequencing/protocols/prots_production.html.
410
411 *Quality control of 18S SSU rRNA dataset*
412 The sequences were checked for chimeras using the Bellerophon Chimera Check and the
413 Check_Chimera utilities (Ribosomal Database Project) (Cole *et al*. 2003). After removal of
414 putative chimeras, bacterial, archaeal and metazoan sequences, the remaining sequences were
415 grouped into Operational Taxonomic Units (OTUs) based on 98% rRNA gene sequence
416 similarity levels. This was achieved by first making all possible pair-wise sequence alignments

by using ClustalW (Thompson *et al*. 1994), calculating % sequence identities, followed by clustering the sequences by using the unweighted pair group method with arithmetic mean (UPGMA) as implemented in the OC clustering program (http:// www.compbio.dundee.ac.uk/Software/OC/oc.html). OTUs clustered at the 98% identity threshold were subjected to ordination and multivariate statistical analyses.

Community composition analysis

18S rRNA gene sequences from Saanich Inlet (19 samples), Cariaco Basin (16 samples) (Edgcomb *et al.* 2011a; Orsi *et al.* 2011b), and Framvaren Fjord (9 samples) (Behnke *et al.* 2010) were obtained from the GenBank nt database and were aligned using the Needleman-Wunsch algorithm implemented in MOTHUR version 1.18 (Schloss *et al*. 2009), with gap penalty of -1 and k-mer size of 9. The resulting alignment file was used as an input to generate 6,768 operational taxonomic units (OTUs) based on 98% similarity threshold and a representative sequence for each OTU was selected using the recommended distance-based method, Get.oturep, (http://www.mothur.org/wiki/Get.oturep). The resulting cluster file and list of representative sequences for each OTU were combined to generate an OTU table containing the number of sequences in each OTU across the 44 samples. Representative sequences were then used in a BLASTn search against Silva SSU reference database version 10.6 with sequences affiliated with the Bacteria and Archaea removed. The results of clustering were also used to calculate non-parametric Chao, Shannon and Simpson indices of alpha diversity (Table S2).

Circos Visualization

The BLASTn output and the OTU table were combined together to generate a community composition table containing the number of sequences across each samples belonging to a specific taxonomic group. From this table, the relative abundance of each

442 protistan group in a given sample was calculated as a percentage value by dividing the raw
443 number of sequences associated with the specific taxa by the total number of sequences in the
444 sample. Circos was used to generate circular link diagrams illustrating community composition
445 differences among and between samples (Krzywinski *et al.* 2009).
446
447 *Histoheatmap Generation*
448 The same table used in Circos visualization was also used to generate combined histograms and
449 heatmaps (*i.e.* histoheatmaps illustrating the presence of different OTUs across each sample)
450 using the R statistical software package (http://www.r-project.org). This script is available
451 upon request from the authors.
452
453 *Phylogenetic Analysis*
454 For our phylogenetic analyses we focused on ciliate and euglenozoan-affiliated sequences in the
455 July 200 m sample. We focused on this sample because it contained the most novel sequences of
456 all clone libraries/samples in our study. For these analyses, we used representative sequences
457 from each ciliate and euglenozoan-affiliated OTU clustered at the 98% sequence identity level.
458 Representative sequences were compared against the Genbank-nt database using BLASTn in
459 search of their closest relatives and the highest scoring cultured and uncultured sequence
460 relatives were retrieved. Sequences were aligned using the ARB automated aligner (Ludwig *et*
461 *al*. 2004), the alignment was manually refined using secondary structure information, and only
462 unambiguous positions were used to construct phylogenetic trees. Phylogenetic trees were
463 constructed using Bayesian inference (Ronquist and Huelsenbeck 2003) and PhyML (Guindon *et*
464 *al*. 2005).
465
466

Statistical Analyses

After clustering of our Sanger sequence dataset, we obtained "frequency count" data at the 90, 95, 98 and 99% sequence identity levels. These are the numbers of OTUs registered once (the "singletons"), or twice (the "doubletons"), etc. Using these data we estimated the total number of OTUs at each level of sequence identity, representing the sum of seen (empirically registered) and unseen OTUs (present but undetected due to limited sequencing effort). This was performed using the program CatchAll (Bunge 2010) to compute eight parametric (Poisson; negative binomial; inverse Gaussian, Pareto and lognormal-mixed Poisson; and mixtures of one, two, or three geometrics) estimators as described previously (Hong *et al*. 2006).

Canonical correspondence analysis

Canonical correspondence analysis (CCA) was used to elucidate relationships between protistan community structure and concentrations of dissolved oxygen (O_2), nitrate (NO_3^-), methane (CH_4), and hydrogen sulfide (H_2S). MRPP was used to test for a statistically significant influence of season, depth, nitrate, sulfide, and oxygen on the observed OTU distribution. A Monte Carlo test was also used to assess the null hypothesis of no relationship between OTU distributions and environmental variables. All ordination and multivariate statistical analyses were performed on our dataset clustered at the 98% sequence identity threshold. Monte Carlo tests, MRPP, and CCA were implemented using the PC-ORD software package (MjM Software Design).

Principle component and hierarchical clustering analysis

To determine the correlation between protistan community structure and the oxygen concentration in each sample, a table containing the raw number of sequences associated each major taxon was used as an input for principal component analysis (PCA) using the

FactorMineR module (http://factominer.free.fr.). Based on the first two principal components calculated from the PCA analysis, samples were hierarchically clustered using the Manhattan distance method with complete linkage implemented in the same software module. The results of the analysis were visualized as dendrograms with dot plots using the custom perl script, bubble.pl, (http://www.cmde.science.ubc.ca/hallam/bubble.php).

Acknowledgements

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory under Contract No., and Los Alamos National Laboratory (Contract No. DE-AC02-05CH11231, DE-AC52-07NA27344, DE-AC02-06NA25396), the Natural Sciences and Engineering Research Council (NSERC) of Canada 328256-07 and STPSC 356988, Canada Foundation for Innovation (CFI) 17444; Canadian Institute for Advanced Research (CIFAR), NSF MCB-0348407 to VE, NSF Center for Deep Energy Biosphere Investigations, and the Center for Bioinorganic Chemistry (CEBIC). We thank M. Robert at the Institute for Ocean Science (IOS), C. Payne, L. Pakhomova, and Philippe Tortell at the University of British Columbia (UBC) for help in sampling and chemical analyses and the captains and crews of the *CCGS John P. Tulley* and *HMS John Strickland* for logistical support. We thank the entire technical staff at the Joint Genome Institute and S. Tringe and Tijana Glavina del Rio for project management support, Richard Christen at Universite´ de Nice, France) for assistance in the preliminary taxonomic parsing of our dataset and Nathan Cahoon at Northeastern University for his helpful assistance with sequence clustering. This is contribution number 190 from the Center for Dark Energy Biosphere Investigations (C-DEBI).

517

518

519 **Accession Numbers**

520 Eukaryotic 18S rRNA gene sequences recovered from the Saanich Inlet water column on

521 were deposited in Genbank under accession numbers HQ864863 - HQ871151.

522

523 **Figure and table legends**

524

525 **Figure 1.** Canonical correspondence analysis of the Saanich Inlet 18S rRNA gene sequence
526 dataset clustered at the 98% identity threshold. Samples are represented in the biplot by dots, the
527 size and color of which indicates the presence and concentration of dissolved oxygen (O_2). Axis
528 1 and 2 explained 8% and 7.7% of the variance in OTU distribution, respectively. A Monte Carlo
529 test for significance of the Eigenvalues yielded a p-value equal to 0.03.

530

531 **Figure 2.** Phylum-level taxonomic affiliations of Saanich Inlet 18S rRNA gene sequences (See
532 Table 1 for sample_id information). The color-coded outer histogram represents the abundance
533 of seventeen major taxonomic groups identified. The relative abundance (% of total) of
534 sequences affiliated with each taxonomic group within a sample is indicated by the thickness of
535 the colored area at the perimeter of the circle. Black circles represent individual samples. The
536 concentration of oxygen within each sample is represented by a black bar, and gray bars indicate
537 samples with no detectable oxygen. The height of each bar is scaled according to the value of
538 the oxygen concentration (in μM) normalized using natural logarithm (ln).

539

540 **Figure 3.** Phylogenetic relationships of ciliate-affiliated 18S rRNA gene sequences. The tree
541 was constructed under Maximum Likelihood using an alignment of 757 unambiguous positions
542 under the GTR+I+Gamma model of sequence evolution. Bootstrap (PhyML, 1000 iterations) and
543 posterior probability (5,000,000 generations with 25% of trees discarded as burnin) values
544 greater than 50% are shown at the nodes in the order PP/ML (posterior probability/maximum
545 likelihood bootstrap). Black circles at nodes represent full posterior probability and bootstrap
546 support. OTUs from our study appear in bold font. The number of sequences per OTU

recovered from oxic, suboxic, dysoxic, and oxic samples are represented by circles. The size and color of the circles denotes the number of sequences and oxygen concentration, respectively.

Figure 4. Phylogenetic relationships of Euglenozoa-affiliated 18S rRNA gene sequences. The tree was constructed under Maximum Likelihood using an alignment of 809 unambiguous positions under the GTR+I+Gamma model of sequence evolution. Bootstrap (PhyML, 1000 iterations) and posterior probability (5,000,000 generations with 25% of trees discarded as burning) values greater than 50% are shown at the nodes in the order PP/ML. Black circles at nodes represent full posterior probability and bootstrap support. OTUs from our study appear in bold font. The number of sequences per OTU recovered from oxic, suboxic, dysoxic, and oxic samples are represented by circles. The size and color of the circles denotes the number of sequences and oxygen concentration, respectively.

Figure 5. Principal component analysis and hierarchical clustering of the 38 samples from Saanich Inlet, Cariaco Basin and Framvaren Fjord (See Table 1 for sample_id information). The x and y axes of the grid represent the first and second principal components, respectively. Each dot represents one of the 38 samples used in the analysis. The visual properties of each dot can be divided into three categories. The shape of each dot represents sample location, Saanich Inlet (circle), Cariaco Basin (hexagon) and Framvaren Fjord (square). Each dot is color-coded based on dissolved oxygen concentration, oxic (red), dysoxic (green), suboxic (light blue) and anoxic (purple). The size of each dot is scaled according to the value of the oxygen concentration (in μM) normalized using natural logarithm (ln). The clustering pattern is further linked to the dendrogram generated from hierarchical clustering.

571 **Figure 6**. Combined histogram and heatmap describing the diversity of Ciliophora, Cercozoa,
572 Fungi and Euglenozoa OTUs among and between sampling depths and locations (See Table 1 for
573 sample_id information). The heatmap shows the number of sequences in each OTU that are
574 affiliated with each color-coded protistan group. Color intensity of each cell is proportional to
575 the log-corrected number of sequences in the OTU. The histogram shows total number of
576 uncorrected sequences in corresponding OTU (*i.e.* sum of sequences in a given OTU across all
577 the samples). Only those OTUs are shown for which the total number of sequences is greater
578 than or equal to 1% of the total sequences affiliated with a given taxa. The count legends indicate
579 the number of cells in each heatmap that contain the designated log-corrected value.
580
581 **Table 1.** Summary of eukaryotic SSU ribosomal RNA gene (18S rRNA) sequence data, as well
582 as geochemical data, generated from the Saanich Inlet water column spanning the 2006-2008
583 time-series.
584
585 **Table 2.** Parametric predicted richness of protistan assemblages in Saanich Inlet, Framvaren
586 Fjord (data taken from Behnke *et al*., 2010), and the Cariaco Basin (data taken from Edgcomb *et*
587 *al*., 2011a) with associated statistics (SE: standard error; GOF: goodness-of-fit (p-value for the
588 corrected Pearson chi-square goodness-of-fit test); CI: 95% Confidence Interval; NA: Not
589 applicable).
590
591 **Supplementary Figure 1.** Phylum-level taxonomic affiliations of successfully sequenced
592 Cariaco Basin 18S rRNA gene sequences (See Table 1 for sample_id information). The color-
593 coded histogram represents the abundance of seventeen major taxonomic groups identified. The
594 relative abundance (% of total) of sequences affiliated with each taxonomic group within a
595 sample is indicated by the thickness of the colored area at the perimeter of the circle. Black

596 hexagons represent individual samples. The concentration of oxygen within each sample is
597 represented by a black bar, gray bars indicate samples with no detectable oxygen. The height of
598 each bar is scaled according to the value of the oxygen concentration (in μM) normalized using
599 natural logarithm (ln).
600
601 **Supplementary Figure 2.** Phylum-level taxonomic affiliations of successfully sequenced
602 Framvaren Fjord 18S rRNA gene sequences (See Table 1 for sample_id information). The color-
603 coded histogram represents the abundance of seventeen major taxonomic groups identified. The
604 relative abundance (% of total) of sequences affiliated with each taxonomic group within a
605 sample is indicated by the thickness of the colored area at the perimeter of the circle. Black
606 squares represent individual samples. The concentration of oxygen within each sample is
607 represented by a black bar, gray bars indicate samples with no detectable oxygen. The height of
608 each bar is scaled according to the value of the oxygen concentration (in μM) normalized using
609 natural logarithm (ln).
610
611 **Supplementary Figure 3.** Combined phylum-level taxonomic affiliations of successfully
612 sequenced 18S rRNA gene sequences from Saanich Inlet, Cariaco Basin and Framvaren Fjord
613 (See Table 1 for sample_id information). The color-coded histograms represents the abundance
614 of seventeen major taxonomic groups identified. The relative abundance (% of total) of
615 sequences affiliated with each taxonomic group within a sample is indicated by the thickness of
616 the colored area at the perimeter of the circle. The shape of each dot represents sample location,
617 Saanich Inlet (circle), Cariaco Basin (hexagon) and Framvaren Fjord (square). The
618 concentration of oxygen within each sample is represented by a black bar, gray bars indicate
619 samples with no detectable oxygen. The height of each bar is scaled according to the value of
620 the oxygen concentration (in μM) normalized using natural logarithm (ln).

Supplementary Figure 4. Combined histogram and heatmap describing the diversity of Dinophyceae, Stramenopiles, and Polycystinea OTUs among and between sampling depths and locations (See Table 1 for sample_id information). The heatmap shows the number of sequences in each OTU that are affiliated with each color-coded protistan group. Color intensity of each cell is proportional to the log-corrected number of sequences in the OTU. The histogram shows total number of uncorrected sequences in corresponding OTU (i.e. sum of sequences in a given OTU across all the samples). Only those OTUs are shown for which the total number of sequences is greater than or equal to 1% of the total sequences affiliated with a given taxa. The count legends indicate the number of cells in each heatmap that contain the designated log-corrected value.

Supplementary Figure 5. Combined histogram and heatmap describing the diversity of Apusozoa, Cryptophyta, Choanoflagellida, and Acantharea OTUs among and between sampling depths and locations (See Table 1 for sample_id information). The heatmap shows the number of sequences in each OTU that are affiliated with each color-coded protistan group. Color intensity of each cell is proportional to the log-corrected number of sequences in the OTU. The histogram shows total number of uncorrected sequences in corresponding OTU (i.e. sum of sequences in a given OTU across all the samples). Only those OTUs are shown for which the total number of sequences is greater than or equal to 1% of the total sequences affiliated with a given taxa. The count legends indicate the number of cells in each heatmap that contain the designated log-corrected value.

Supplementary Table 1. Summary of 18S rRNA gene sequences retrieved from the July 200 m library sample affiliated with the phyla Ciliophora and Euglenozoa.

646

647 **Supplementary Table 2.** Non-parametric diversity estimates for the protistan communities of

648 the Saanich Inlet, Framvaren Fjord, and Cariaco Basin.

649 **References**

650 Arrigo KR (2005). Marine microorganisms and global nutrient cycles. *Nature* **437:** 349-55.
651
652 Behnke A, Barger KJ, Bunge J, Stoeck T (2010a). Spatio-temporal variations in protistan
653 communities along an O/HS gradient in the anoxic Framvaren Fjord (Norway). *FEMS Microbiol*
654 *Ecol* **72:** 89-102.
655
656 Behnke A, Bunge J, Barger K, Breiner HW, Alla V, Stoeck T (2006). Microeukaryote
657 community patterns along an O2/H2S gradient in a supersulfidic anoxic fjord (Framvaren,
658 Norway). *Appl Environ Microbiol* **72:** 3626-36.
659
660 Behnke A, Engel M, Christen R, Nebel M, Klein RR, Stoeck T (2010b). Depicting more accurate
661 pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA
662 gene regions. *Environ Microbiol* **13:** 340-9.
663
664 Bernhard JM, Buck KR, Farmer MA, Bowser SS (2000). The Santa Barbara Basin is a symbiosis
665 oasis. *Nature* **403:** 77-80.
666
667 Bunge J. (2010). *Forthcoming in Proceedings of the 2011 Pacific Symposium on Biocomputing*:
668 Hawaii.
669
670 Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, Delong EF *et al* (2010).
671 A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. *Science* **330:** 1375-
672 8.
673
674 Caron DA, Countway PD, Savai P, Gast RJ, Schnetzer A, Moorthi SD *et al* (2009). Defining
675 DNA-based operational taxonomic units for microbial-eukaryote ecology. *Appl Environ*
676 *Microbiol* **75:** 5797-808.
677
678 Chambouvet A, Morin P, Marie D, Guillou L (2008). Control of toxic marine dinoflagellate
679 blooms by serial parasitic killers. *Science* **322:** 1254-7.
680
681 Chao A, Bunge J (2002). Estimating the number of species in a stochastic abundance model.
682 *Biometrics* **58:** 531-539.
683
684 Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA *et al* (2003). The Ribosomal
685 Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the
686 new prokaryotic taxonomy. *Nucleic Acids Res* **31:** 442-3.
687
688 Diaz RJ, Rosenberg R (2008). Spreading dead zones and consequences for marine ecosystems.
689 *Science* **321:** 926-9.
690
691 Diaz RJ, Rosenberg R, Rabalais NN, Levin LA (2009). Dead zone dilemma. *Mar Pollut Bull* **58:**
692 1767-8.
693
694 Edgcomb V, Orsi W, Bunge J, Jeon S, Christen R, Leslin C *et al* (2011a). Protistan microbial
695 observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species
696 richness. *ISME J* **5**: 1344-56.

Edgcomb V, Orsi W, Taylor GT, Vdacny P, Taylor C, Suarez P *et al* (2011b). Accessing marine protists from the anoxic Cariaco Basin. *ISME J* **5**: 1237-41.

Edgcomb V, Beaudoin D, Gast R, Biddle JF, Teske A (2010b). Marine subsurface eukaryotes: the fungal majority. *Environ Microbiol*. doi:10.1111/j.1462-2920.2010.02318.x

Edgcomb V, Leadbetter ER, Bourland W, Beaudoin D, Bernhard JM (2011c). Structured multiple endosymbiosis of bacteria and archaea in a ciliate from marine sulfidic sediments: a survival mechanism in low oxygen, sulfidic sediments? *Front Microbiol* **2:** 55.

Embley TM, Finlay BJ (1993). Systematic and morphological diversity of endosymbiotic methanogens in anaerobic ciliates. *Antonie Van Leeuwenhoek* **64:** 261-71.

Fenchel T, Finlay BJ (1995). *Ecology and evolution in anoxic worlds.* Oxford University Press: Oxford, New York, and Tokyo.

Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL *et al* (2008). A latitudinal diversity gradient in planktonic marine bacteria. *Proc Natl Acad Sci U S A* **105:** 7774-8.

Guillou L, Viprey M, Chambouvet A, Welsh RM, Kirkham AR, Massana R *et al* (2008). Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environ Microbiol* **10:** 3349-65.

Guindon S, Lethiec F, Duroux P, Gascuel O (2005). PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* **33:** W557-9.

Hong SH, Bunge J, Jeon SO, Epstein SS (2006). Predicting microbial species richness. *Proc Natl Acad Sci U S A* **103:** 117-22.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D *et al* (2009). Circos: an information aesthetic for comparative genomics. *Genome Res* **19:** 1639-45.

Lam P, Kuypers MM (2010). Microbial nitrogen cycling processes in oxygen minimum zones. *Ann Rev Mar Sci* **3:** 317-45.

Lavik G, Stuhrmann T, Bruchert V, Van der Plas A, Mohrholz V, Lam P *et al* (2009). Detoxification of sulphidic African shelf waters by blooming chemolithotrophs. *Nature* **457:** 581-4.

Li XN, Taylor GT, Astor Y, Scranton MI (2008). Sulfur speciation in the Cariaco Basin with reference to chemoautotrophic production. *Mar. Chem.* **112:** 53-64.

Lin S, Zhang H, Hou Y, Miranda L, Bhattacharya D (2006). Development of a dinoflagellate-oriented PCR primer set leads to detection of picoplanktonic dinoflagellates from Long Island Sound. *Appl Environ Microbiol* **72:** 5626-30.

Lin XJ, Scranton MI, Chistoserdov A, Varela R, Taylor GT (2008). Spatiotemporal dynamics of bacterial populations in the anoxic Cariaco Basin. *Limnology and Oceanography* **53:** 37-51.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32:** 1363-71.

Massana R, Terrado R, Forn I, Lovejoy C, Pedros-Alio C (2006). Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. *Environ Microbiol* **8:** 1515-22.

Massana R, Unrein F, Rodriguez-Martinez R, Forn I, Lefort T, Pinhassi J *et al* (2009). Grazing rates and functional diversity of uncultured heterotrophic flagellates. *ISME J* **3:** 588-96.

McCune B, Grace JB (2002). *Analysis of ecological communities*. MJM Software Design: Gleneden, Oregon.

Naqvi SWA, Bange HW, Farıas L, Monteiro PMS, Scranton MI, Zhang J (2010). Marine hypoxia/anoxia as a source of CH4 and N2O. *Biogeosciences* **7:** 2159-2190.

Not F, del Campo J, Balague V, de Vargas C, Massana R (2009). New insights into the diversity of marine picoeukaryotes. *PLoS One* **4:** e7143.

Orsi W, Edgcomb V, Faria J, Foissner W, Fowle WH, Hohmann T *et al* (2011a). Class Cariacotrichea, a novel ciliate taxon from the anoxic Cariaco Basin, Venezuela. *Int J Syst Evol Microbiol* doi:10.1099/ijs.0.034710-0.

Orsi W, Edgcomb V, Jeon S, Leslin C, Bunge J, Taylor GT *et al* (2011b). Protistan microbial observatory in the Cariaco Basin, Caribbean. II. Habitat specialization. *ISME J* **5**: 1357-73.

Pedros-Alio C (2007). Ecology. Dipping into the rare biosphere. *Science* **315:** 192-3.

Prescott DM (1994). The DNA of ciliated protozoa. *Microbiol. Rev.* **58:** 233-267.

Ronquist F, Huelsenbeck JP (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572-4.

Schloss PD (2009). A high-throughput DNA sequence aligner for microbial ecology studies. *PLoS One* **4:** e8230.

Schubert C (1982). Origin of Cariaco Basin, southern Caribbean Sea. *Marine Geology* **47:** 345-360.

Sherr EB, Sherr BF (2002). Significance of predation by protists in aquatic microbial food webs. *Antonie Van Leeuwenhoek* **81:** 293-308.

Stoeck T, Bass D, Nebel M, Christen R, Jones MD, Breiner HW *et al* (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol* **19 Suppl 1:** 21-31.

Stoeck T, Behnke A, Christen R, Amaral-Zettler L, Rodriguez-Mora MJ, Chistoserdov A *et al* (2009). Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. *BMC Biol* **7:** 72.

Stoeck T, Taylor GT, Epstein SS (2003). Novel eukaryotes from the permanently anoxic Cariaco Basin (Caribbean Sea). *Appl Environ Microbiol* **69:** 5656-5663.

Taylor GT (1982). The role of pelagic heterotrophic protozoa in nutrient cycling: A review. *Ann. Inst. Oceanogr., (Suppl.), Paris* **58(S):** 227-241.

Taylor GT, Scranton MI, Iabichella M, Y. HT, Thunell RC, Muller-Karger F *et al* (2001). Chemoautotrophy in the redox transition zone of the Cariaco Basin: A significant midwater source of organic carbon production. *Limnol. Oceanogr.* **46:** 148-163.

Teske A (2010). Oceans. Cryptic links in the ocean. *Science* **330:** 1326-7.

Thompson JD, Higgens DG, Gibson TJ (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **33:** 4673-4680.

van Hoek AH, van Alen TA, Sprakel VS, Leunissen JA, Brigge T, Vogels GD *et al* (2000). Multiple acquisition of methanogenic archaeal symbionts by anaerobic ciliates. *Mol Biol Evol* **17:** 251-8.

Walsh DA, Hallam SJ (2011). Bacterial community diversity and population dynamics in a seasonally anoxic fjord: Saanich Inlet. In: Bruijn Fd (ed). *Handbook of Molecular Microbial Ecology II: Metagenomics and Complementary Approaches*. Wiley-Blackwell. pp 253-267.

Walsh DA, Zaikova E, Hallam SJ (2009a). Large volume filtration of coastal seawater samples. *J. Vis. Exp.* **28**. doi: 10.3791/1161

Walsh DA, Zaikova E, Hallam SJ (2009b). Small volume filtration of coastal seawater samples. *J. Vis. Exp.* **28**. doi: 10.3791/1163

Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG *et al* (2009). Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* **326:** 578-82.

Wright JJ, Lee S, Zaikova E, Walsh DA, Hallam SJ (2009). DNA extraction from 0.22 micron Sterivex filters and cesium chloride density gradient centrifugation. *J. Vis. Exp.* **31**. doi:10.3791/1352

Yubuki N, Edgcomb VP, Bernhard JM, Leander BS (2009). Ultrastructure and molecular phylogeny of Calkinsia aureus: cellular identity of a novel clade of deep-sea euglenozoans with epibiotic bacteria. *BMC Microbiol* **9:** 16.

Zaikova E, Hawley A, Walsh DA, Hallam SJ (2009). Seawater sampling and collection (2009). *J. Vis. Exp*. doi: 10.3791/1159

Zaikova E, Walsh DA, Stilwell CP, Mohn WW, Tortell PD, Hallam SJ (2010). Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environ Microbiol* **12:** 172-91.

Zehnder AJ, Stumm W (1988). Geochemistry and biogeochemistry of anaerobic habitats. In: Zehnder AJ (ed). *Biology of anaerobic microorganisms*. John Wiley & Sons: New York.

Zuendorf A, Bunge J, Behnke A, Barger KJ, Stoeck T (2006). Diversity estimates of microeukaryotes below the chemocline of the anoxic Mariager Fjord, Denmark. *FEMS Microbiol Ecol* **58:** 476-91.

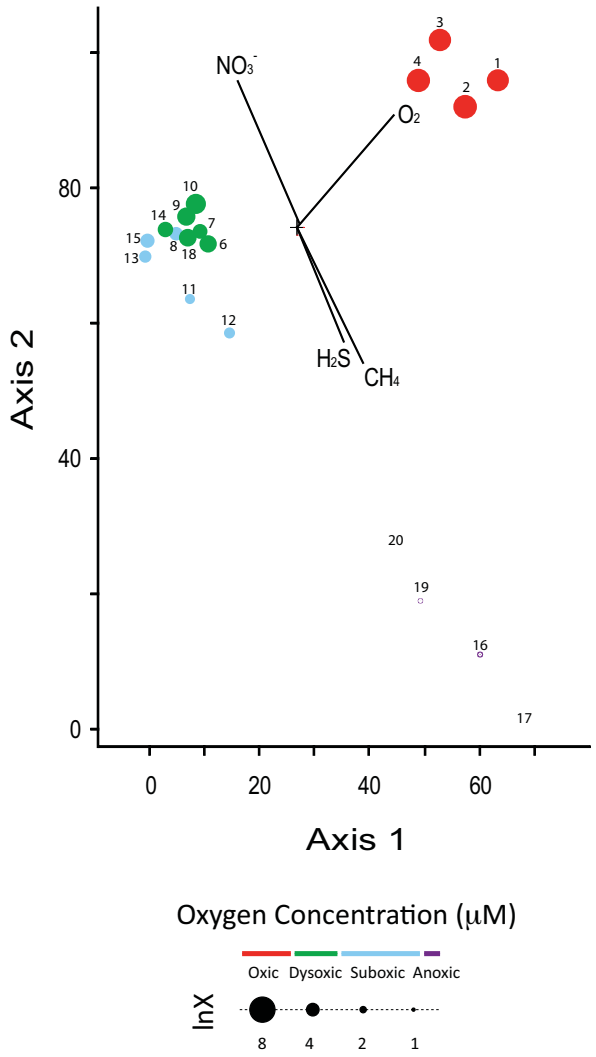

Axis 2
Axis 1
0
40
80
0
20
40
60
NO3-
O2
H2S
CH4
1
2
3
4
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Oxygen Concentration (μM)
Oxic
Dysoxic
Suboxic
Anoxic
lnX
8
4
2
1

Figure 1

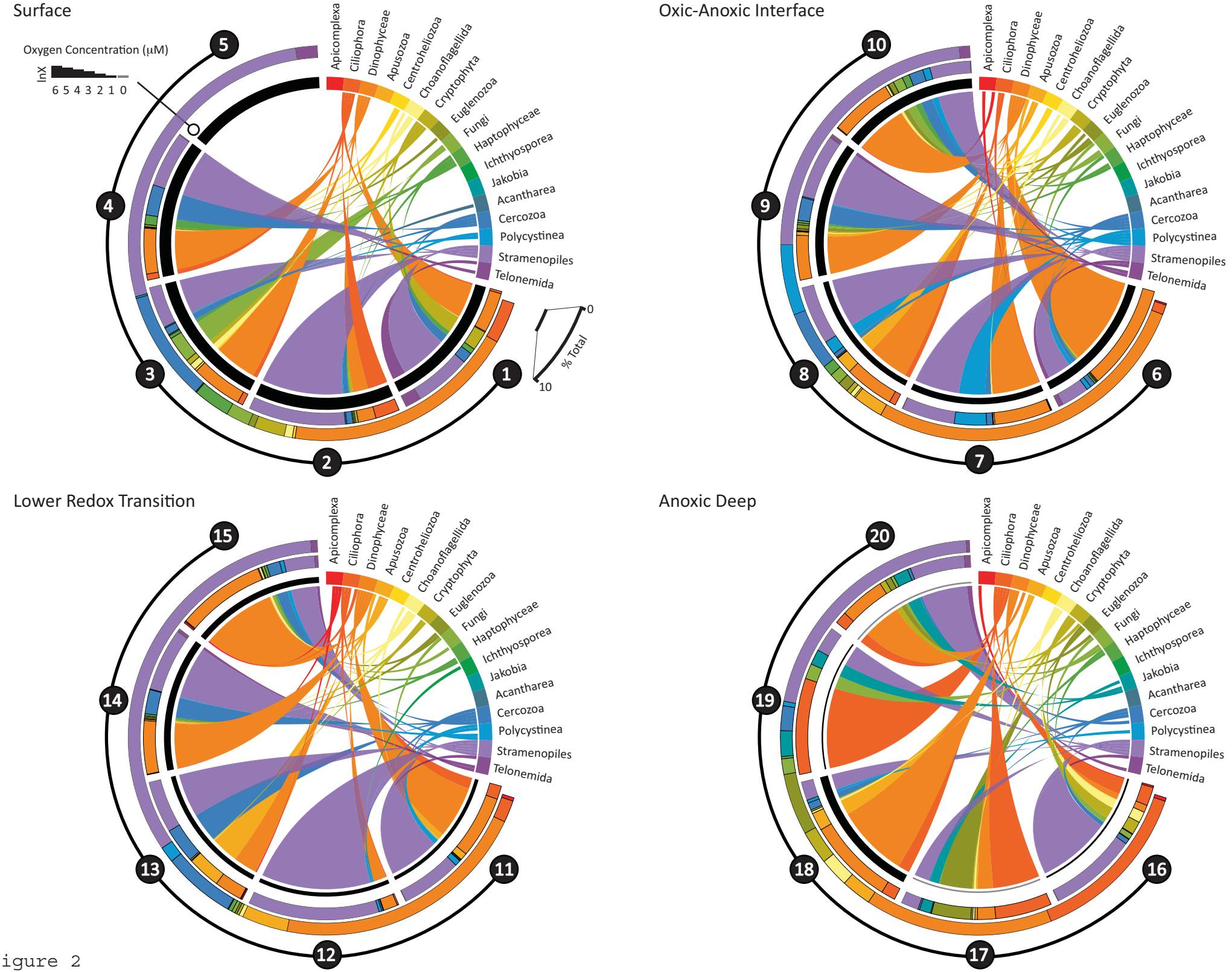
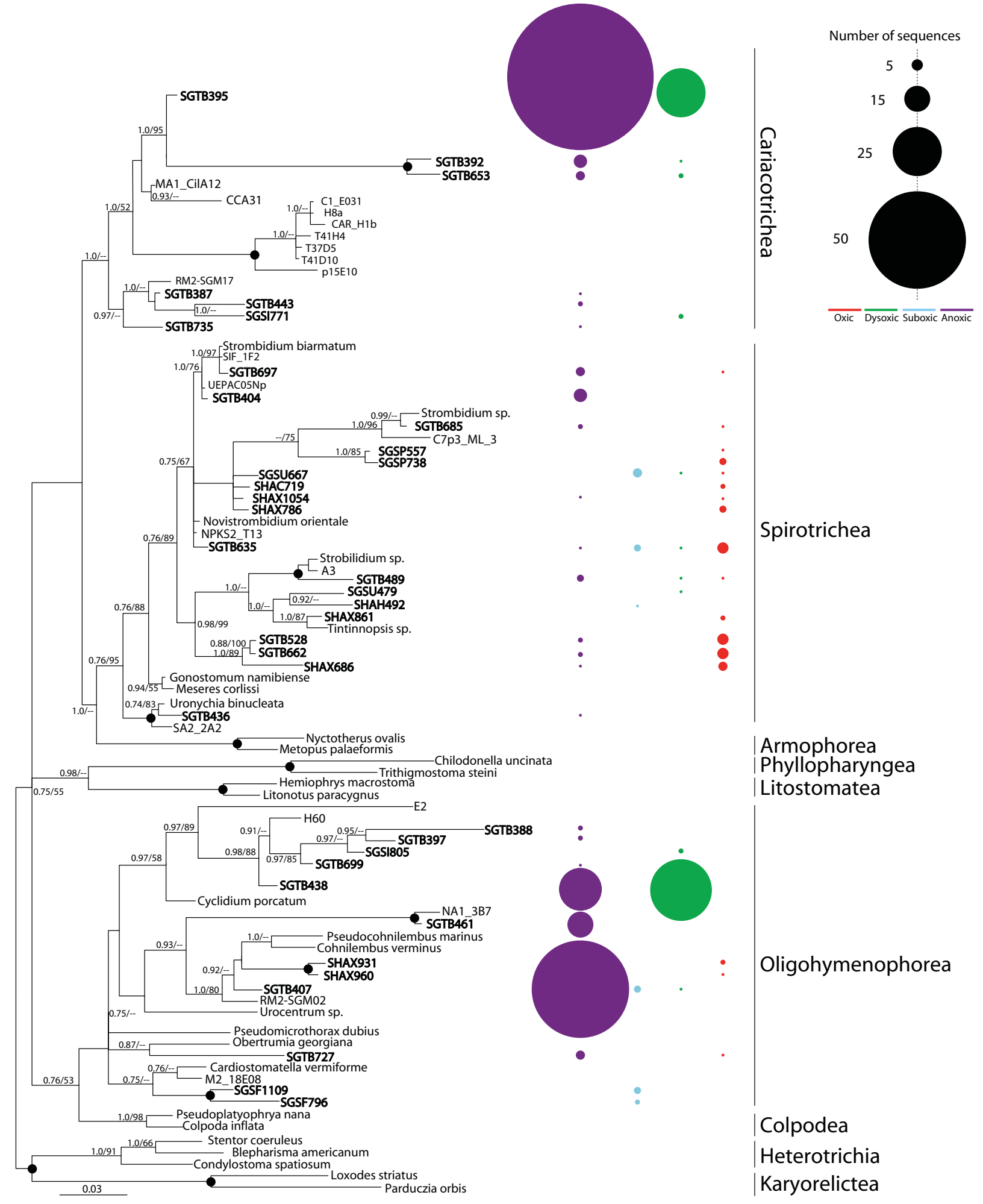
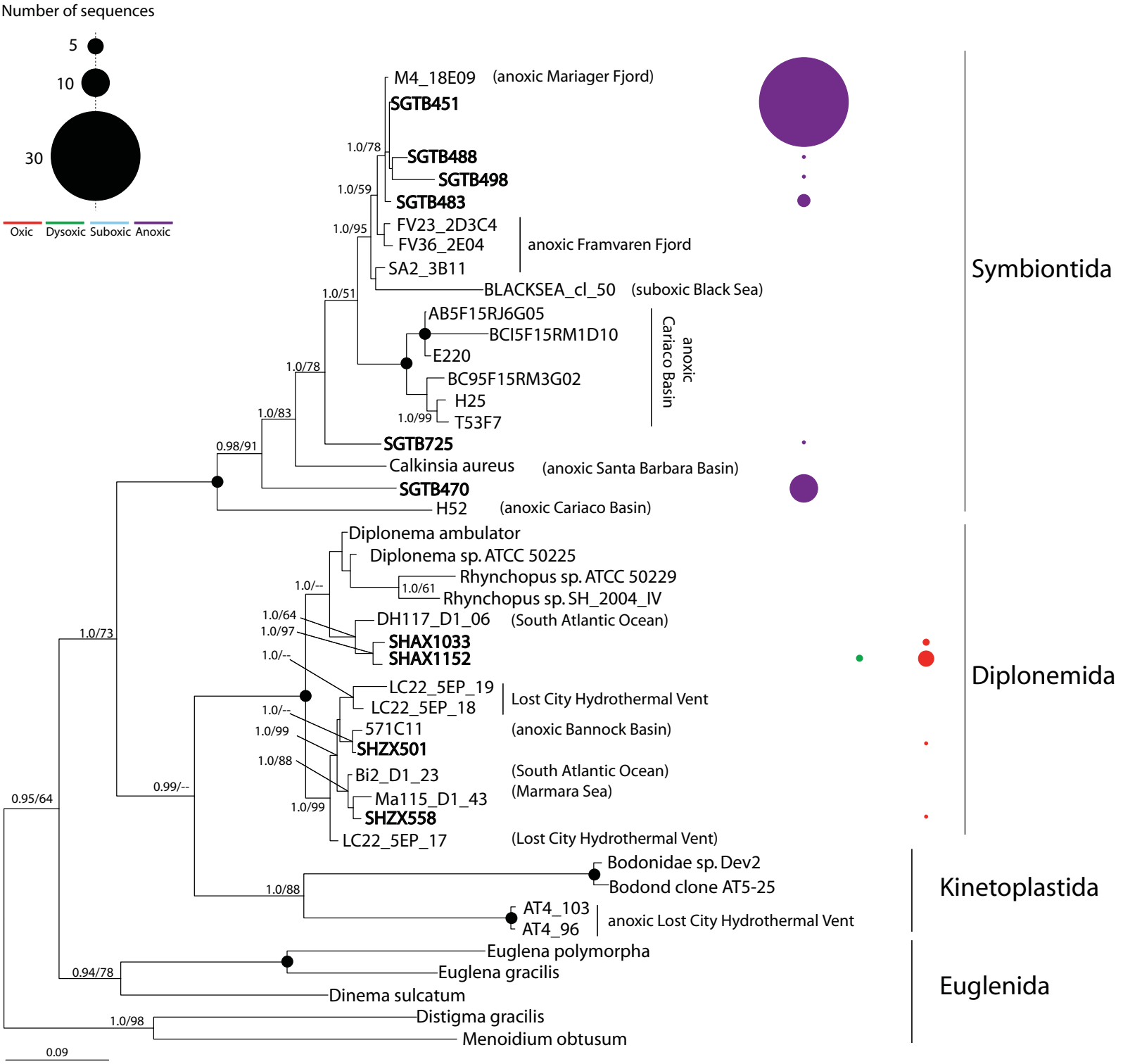

Surface
Oxygen Concentration (μM)
lnX
6 5 4 3 2 1 0
Apicomplexa
Ciliophora
Dinophyceae
Apusozoa
Centroheliozoa
Choanoflagellida
Cryptophyta
Euglenozoa
Fungi
Haptophyceae
Ichthyosporea
Jakobia
Acantharea
Cercozoa
Polycystinea
Stramenopiles
Telonemida
0
% Total
10
1
2
3
4
5
Oxic-Anoxic Interface
6
7
8
9
10
Lower Redox Transition
11
12
13
14
15
Anoxic Deep
16
17
18
19
20

Figure 2


Number of sequences
5
15
25
50
Oxic
Dysoxic
Suboxic
Anoxic
Cariacotrichea
Spirotrichea
Armophorea
Phyllopharyngea
Litostomatea
Oligohymenophorea
Colpodea
Heterotrichia
Karyorelictea
SGTB395
SGTB392
SGTB653
MA1_CilA12
CCA31
C1_E031
H8a
CAR_H1b
T41H4
T37D5
T41D10
p15E10
RM2-SGM17
SGTB387
SGTB443
SGSI771
SGTB735
Strombidium biarmatum
SIF_1F2
SGTB697
UEPAC05Np
SGTB404
Strombidium sp.
SGTB685
C7p3_ML_3
SGSP557
SGSP738
SGSU667
SHAC719
SHAX1054
SHAX786
Novistrombidium orientale
NPKS2_T13
SGTB635
Strobilidium sp.
A3
SGTB489
SGSU479
SHAH492
SHAX861
Tintinnopsis sp.
SGTB528
SGTB662
SHAX686
Gonostomum namibiense
Meseres corlissi
Uronychia binucleata
SGTB436
SA2_2A2
Nyctotherus ovalis
Metopus palaeformis
Chilodonella uncinata
Trithigmostoma steini
Hemiophrys macrostoma
Litonotus paracygnus
E2
H60
SGTB388
SGTB397
SGSI805
SGTB699
SGTB438
Cyclidium porcatum
NA1_3B7
SGTB461
Pseudocohnilembus marinus
Cohnilembus verminus
SHAX931
SHAX960
SGTB407
RM2-SGM02
Urocentrum sp.
Pseudomicrothorax dubius
Obertrumia georgiana
SGTB727
Cardiostomatella vermiforme
M2_18E08
SGSF1109
SGSF796
Pseudoplatyophrya nana
Colpoda inflata
Stentor coeruleus
Blepharisma americanum
Condylostoma spatiosum
Loxodes striatus
Parduczia orbis
0.03


Number of sequences
5
10
30
Oxic
Dysoxic
Suboxic
Anoxic
M4_18E09 (anoxic Mariager Fjord)
SGTB451
1.0/78
SGTB488
SGTB498
1.0/59
SGTB483
1.0/95
FV23_2D3C4
FV36_2E04
anoxic Framvaren Fjord
SA2_3B11
BLACKSEA_cl_50 (suboxic Black Sea)
1.0/51
AB5F15RJ6G05
BCI5F15RM1D10
E220
anoxic Cariaco Basin
1.0/78
BC95F15RM3G02
H25
1.0/99
T53F7
1.0/83
SGTB725
0.98/91
Calkinsia aureus (anoxic Santa Barbara Basin)
SGTB470
H52 (anoxic Cariaco Basin)
Symbiontida
Diplonema ambulator
Diplonema sp. ATCC 50225
1.0/--
Rhynchopus sp. ATCC 50229
1.0/61
Rhynchopus sp. SH_2004_IV
1.0/64
DH117_D1_06 (South Atlantic Ocean)
1.0/97
SHAX1033
1.0/--
SHAX1152
1.0/73
LC22_5EP_19
LC22_5EP_18
Lost City Hydrothermal Vent
1.0/--
571C11 (anoxic Bannock Basin)
1.0/99
SHZX501
1.0/88
Bi2_D1_23 (South Atlantic Ocean)
0.99/--
Ma115_D1_43 (Marmara Sea)
0.95/64
1.0/99
SHZX558
LC22_5EP_17 (Lost City Hydrothermal Vent)
Diplonemida
Bodonidae sp. Dev2
Bodond clone AT5-25
1.0/88
AT4_103
AT4_96
anoxic Lost City Hydrothermal Vent
Kinetoplastida
Euglena polymorpha
0.94/78
Euglena gracilis
Dinema sulcatum
1.0/98
Distigma gracilis
Menoidium obtusum
Euglenida
0.09

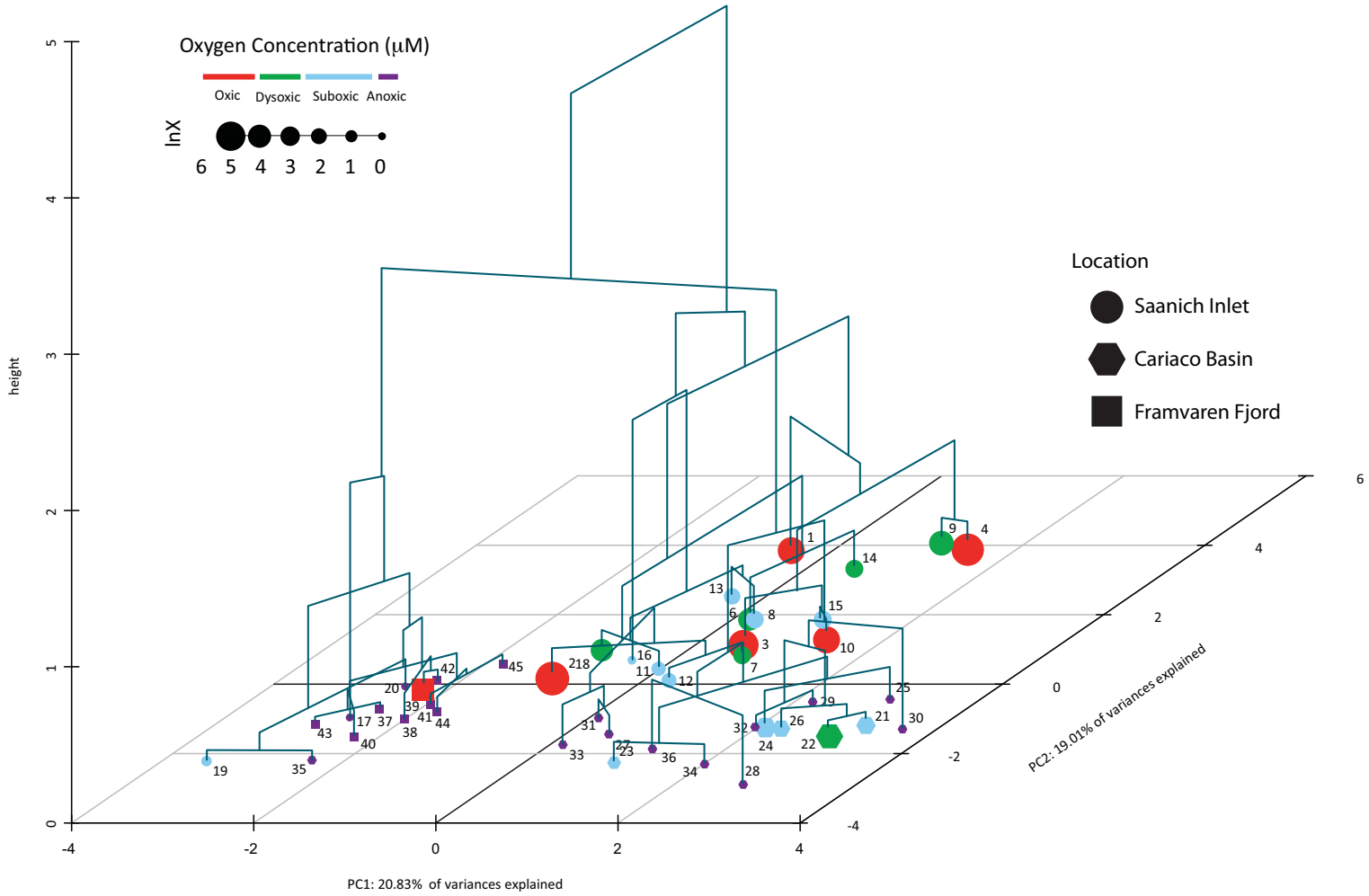

Oxygen Concentration (μM)
Oxic
Dysoxic
Suboxic
Anoxic
lnX
6 5 4 3 2 1 0
Location
Saanich Inlet
Cariaco Basin
Framvaren Fjord
height
PC1: 20.83% of variances explained
PC2: 19.01% of variances explained

Figure 5

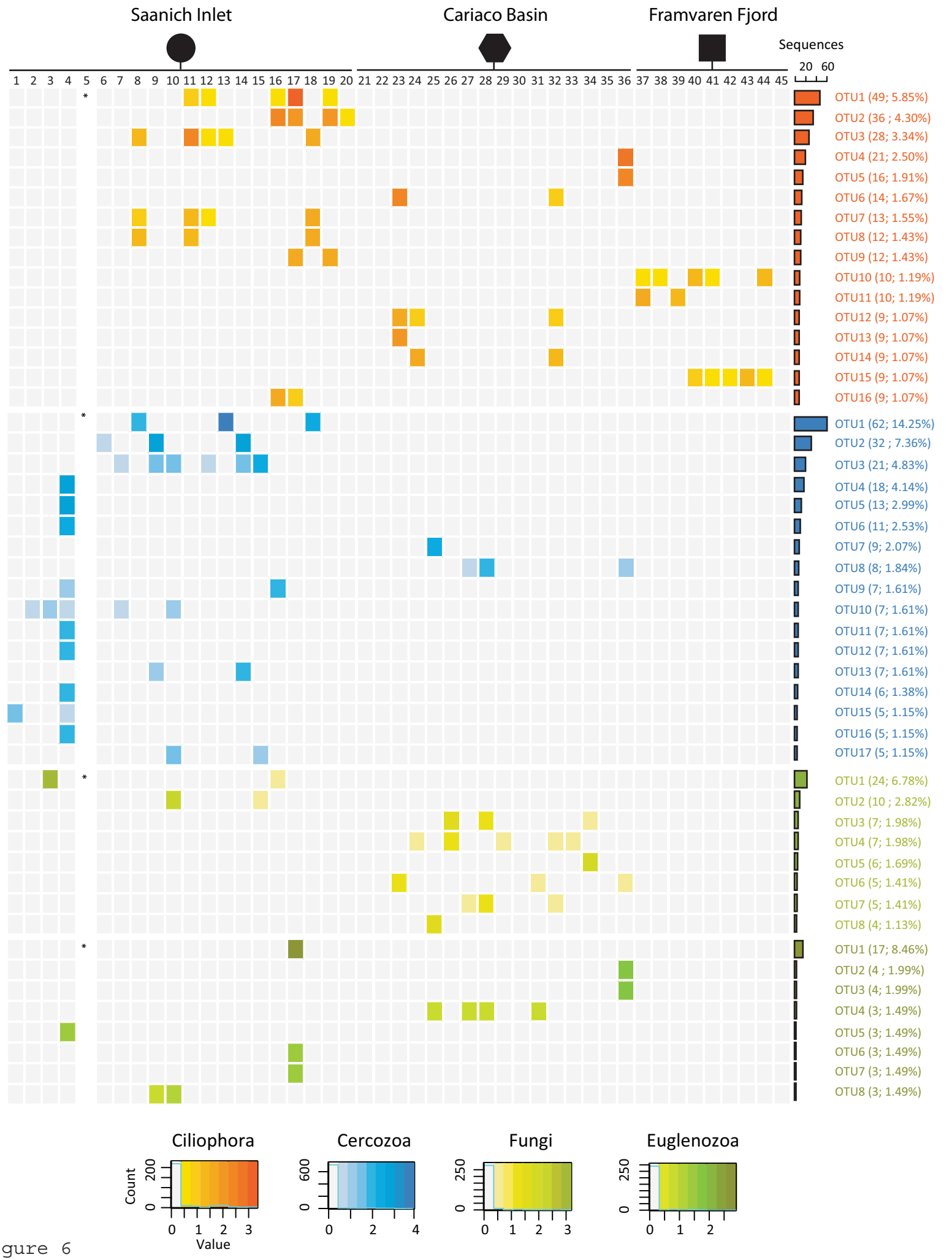

Saanich Inlet
Cariaco Basin
Framvaren Fjord
Sequences
20 60
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
OTU1 (49; 5.85%)
OTU2 (36 ; 4.30%)
OTU3 (28; 3.34%)
OTU4 (21; 2.50%)
OTU5 (16; 1.91%)
OTU6 (14; 1.67%)
OTU7 (13; 1.55%)
OTU8 (12; 1.43%)
OTU9 (12; 1.43%)
OTU10 (10; 1.19%)
OTU11 (10; 1.19%)
OTU12 (9; 1.07%)
OTU13 (9; 1.07%)
OTU14 (9; 1.07%)
OTU15 (9; 1.07%)
OTU16 (9; 1.07%)
OTU1 (62; 14.25%)
OTU2 (32 ; 7.36%)
OTU3 (21; 4.83%)
OTU4 (18; 4.14%)
OTU5 (13; 2.99%)
OTU6 (11; 2.53%)
OTU7 (9; 2.07%)
OTU8 (8; 1.84%)
OTU9 (7; 1.61%)
OTU10 (7; 1.61%)
OTU11 (7; 1.61%)
OTU12 (7; 1.61%)
OTU13 (7; 1.61%)
OTU14 (6; 1.38%)
OTU15 (5; 1.15%)
OTU16 (5; 1.15%)
OTU17 (5; 1.15%)
OTU1 (24; 6.78%)
OTU2 (10 ; 2.82%)
OTU3 (7; 1.98%)
OTU4 (7; 1.98%)
OTU5 (6; 1.69%)
OTU6 (5; 1.41%)
OTU7 (5; 1.41%)
OTU8 (4; 1.13%)
OTU1 (17; 8.46%)
OTU2 (4 ; 1.99%)
OTU3 (4; 1.99%)
OTU4 (3; 1.49%)
OTU5 (3; 1.49%)
OTU6 (3; 1.49%)
OTU7 (3; 1.49%)
OTU8 (3; 1.49%)
Ciliophora
Count
0 200
0 1 2 3
Value
Cercozoa
0 600
0 2 4
Fungi
0 250
0 1 2 3
Euglenozoa
0 250
0 1 2

Figure 6

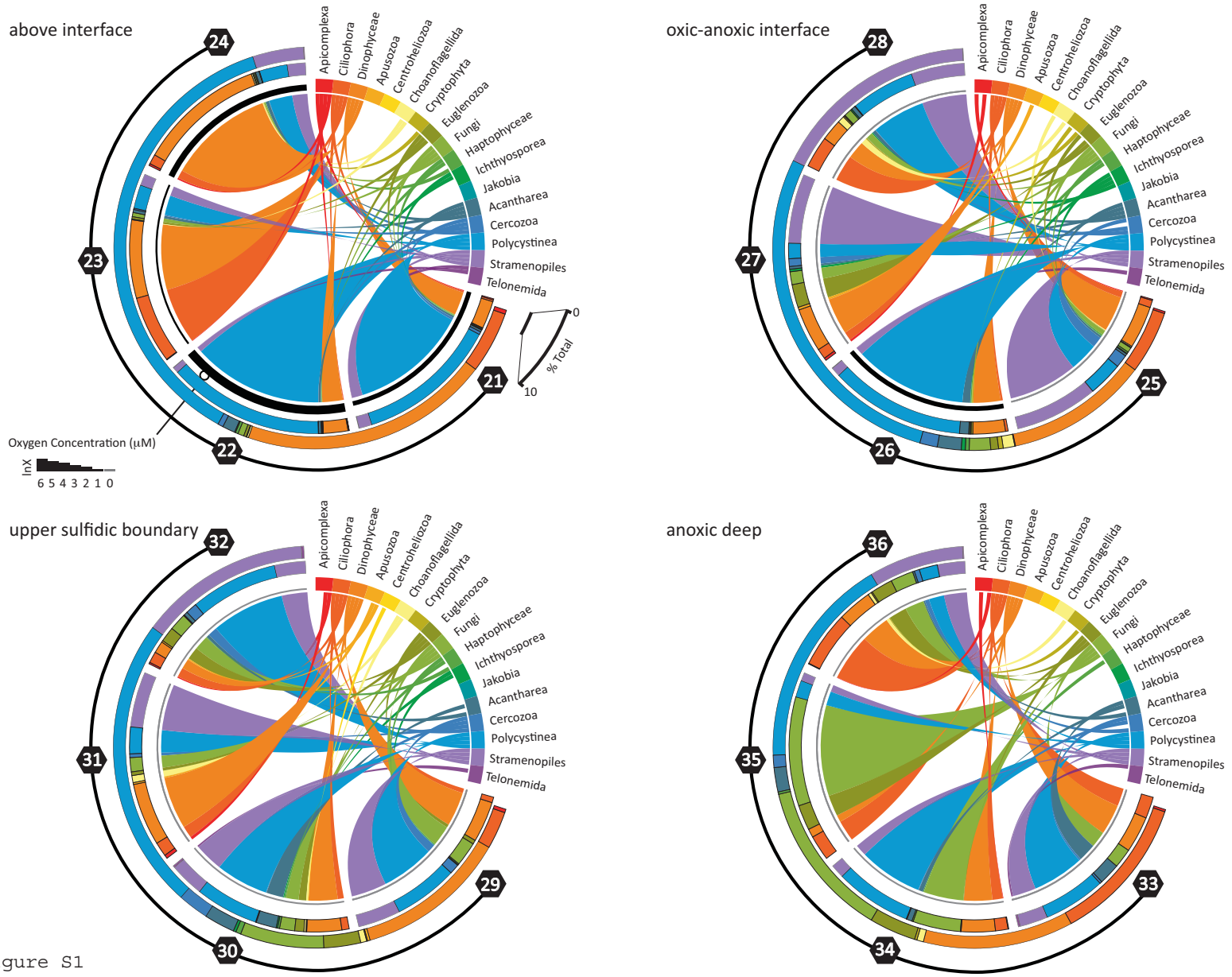

above interface
oxic-anoxic interface
upper sulfidic boundary
anoxic deep
Apicomplexa
Ciliophora
Dinophyceae
Apusozoa
Centroheliozoa
Choanoflagellida
Cryptophyta
Euglenozoa
Fungi
Haptophyceae
Ichthyosporea
Jakobia
Acantharea
Cercozoa
Polycystinea
Stramenopiles
Telonemida
% Total
0
10
Oxygen Concentration (µM)
lnX
6 5 4 3 2 1 0
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

Figure S1

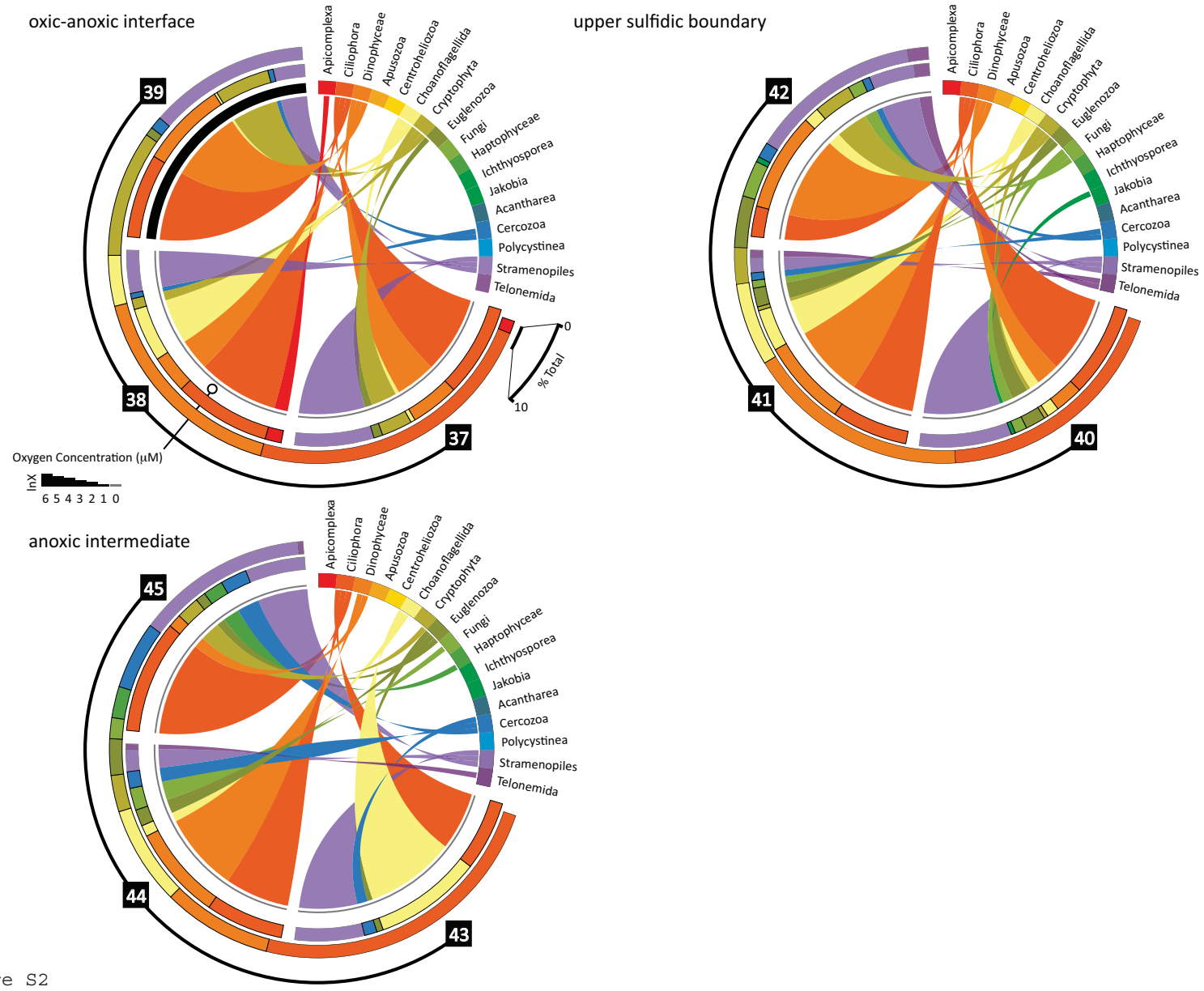

oxic-anoxic interface
upper sulfidic boundary
anoxic intermediate
Apicomplexa
Ciliophora
Dinophyceae
Apusozoa
Centrohelozoa
Choanoflagellida
Cryptophyta
Euglenozoa
Fungi
Haptophyceae
Ichthyosporea
Jakobia
Acantharea
Cercozoa
Polycystinea
Stramenopiles
Telonemida
39
38
37
42
41
40
45
44
43
0
% Total
10
Oxygen Concentration (μM)
lnX
6 5 4 3 2 1 0

Figure S2

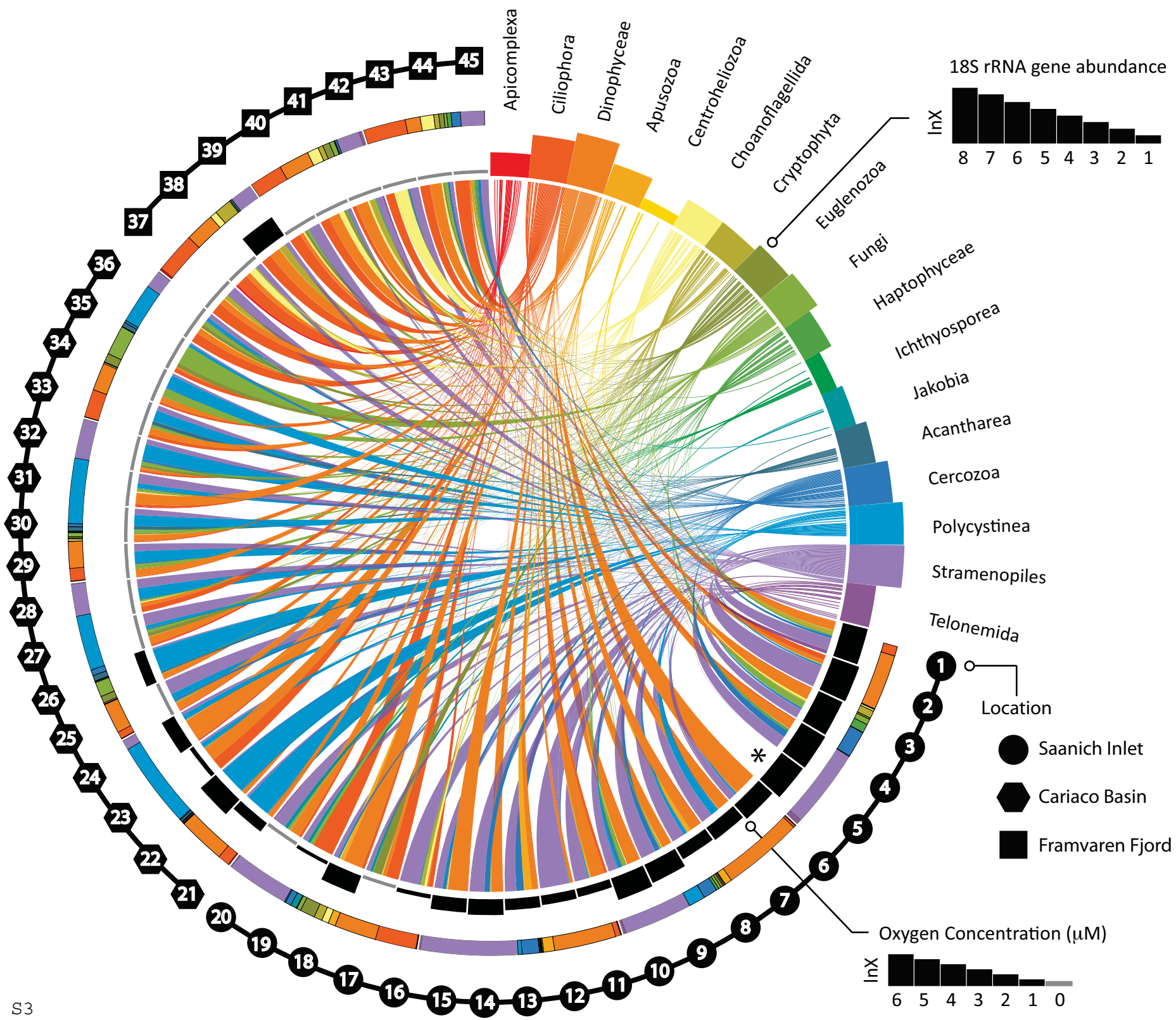

Apicomplexa
Ciliophora
Dinophyceae
Apusozoa
Centroheliozoa
Choanoflagellida
Cryptophyta
Euglenozoa
Fungi
Haptophyceae
Ichthyosporea
Jakobia
Acantharea
Cercozoa
Polycystinea
Stramenopiles
Telonemida
18S rRNA gene abundance
lnX
8 7 6 5 4 3 2 1
Location
Saanich Inlet
Cariaco Basin
Framvaren Fjord
Oxygen Concentration (µM)
lnX
6 5 4 3 2 1 0
*

Figure S3

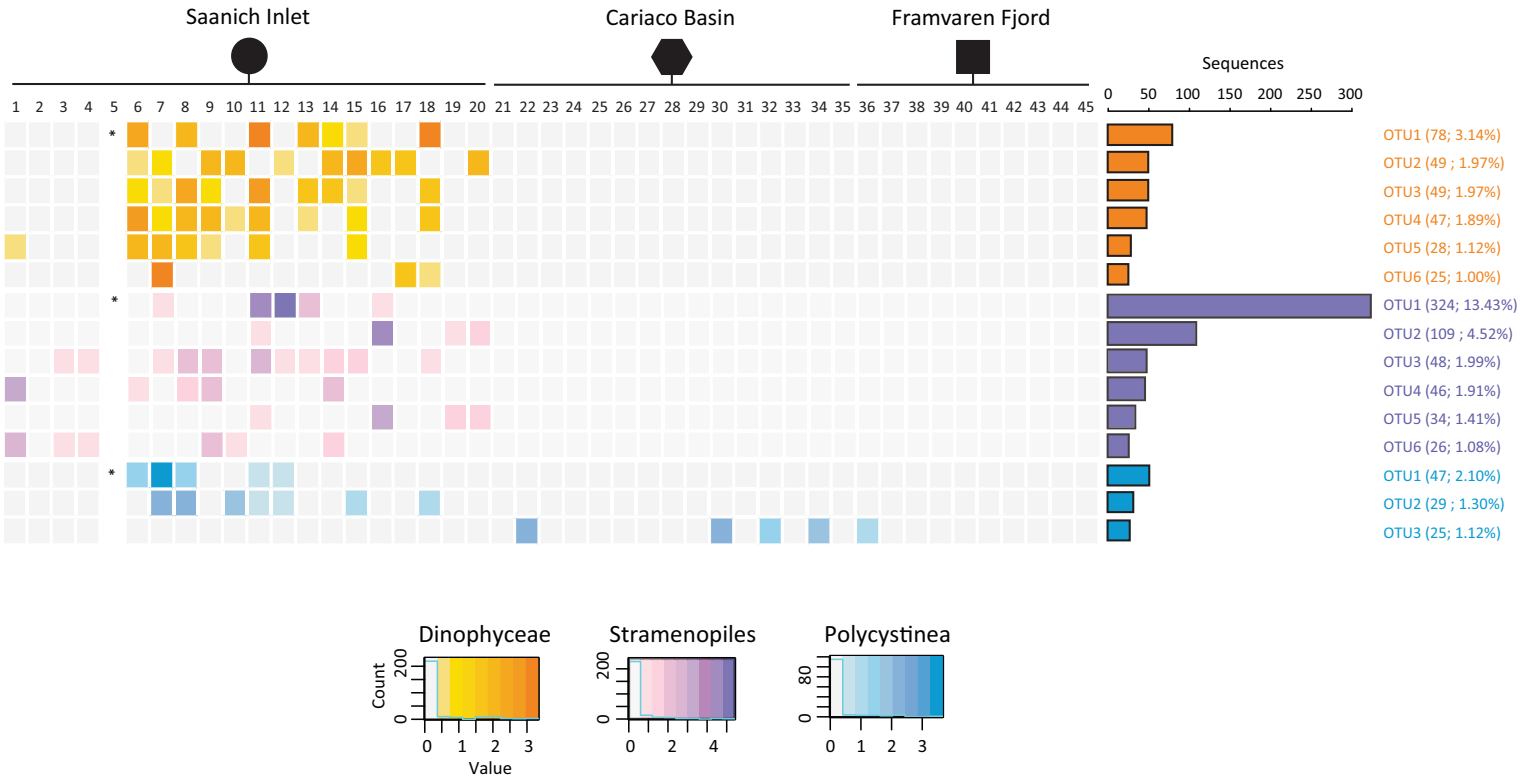

Saanich Inlet
Cariaco Basin
Framvaren Fjord
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
Sequences
0 50 100 150 200 250 300
OTU1 (78; 3.14%)
OTU2 (49 ; 1.97%)
OTU3 (49; 1.97%)
OTU4 (47; 1.89%)
OTU5 (28; 1.12%)
OTU6 (25; 1.00%)
OTU1 (324; 13.43%)
OTU2 (109 ; 4.52%)
OTU3 (48; 1.99%)
OTU4 (46; 1.91%)
OTU5 (34; 1.41%)
OTU6 (26; 1.08%)
OTU1 (47; 2.10%)
OTU2 (29 ; 1.30%)
OTU3 (25; 1.12%)
Dinophyceae
Count
0 200
0 1 2 3
Value
Stramenopiles
0 200
0 2 4
Polycystinea
0 80
0 1 2 3

Figure S4

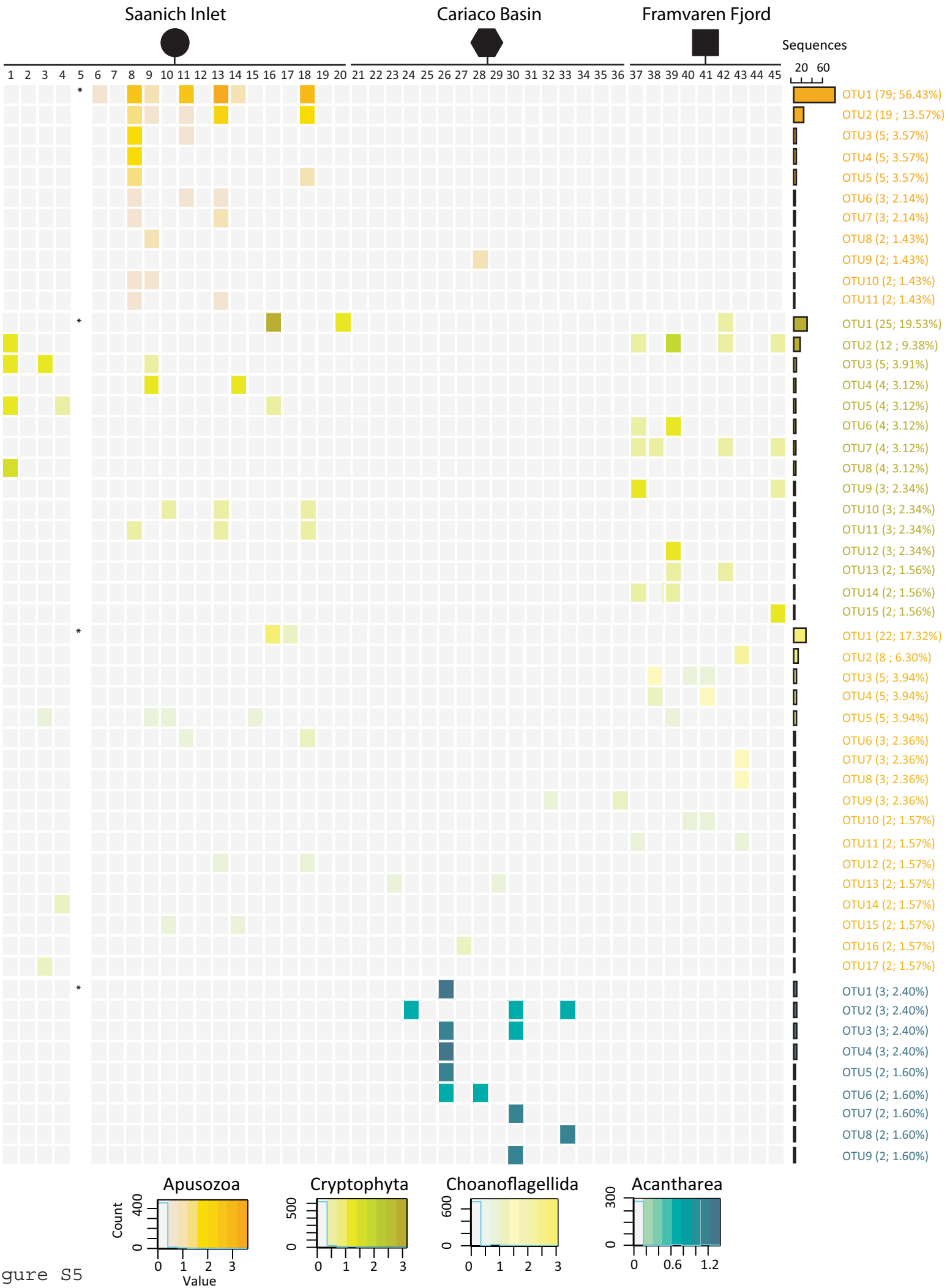

Saanich Inlet
Cariaco Basin
Framvaren Fjord
Sequences
20 60
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
OTU1 (79; 56.43%)
OTU2 (19 ; 13.57%)
OTU3 (5; 3.57%)
OTU4 (5; 3.57%)
OTU5 (5; 3.57%)
OTU6 (3; 2.14%)
OTU7 (3; 2.14%)
OTU8 (2; 1.43%)
OTU9 (2; 1.43%)
OTU10 (2; 1.43%)
OTU11 (2; 1.43%)
OTU1 (25; 19.53%)
OTU2 (12 ; 9.38%)
OTU3 (5; 3.91%)
OTU4 (4; 3.12%)
OTU5 (4; 3.12%)
OTU6 (4; 3.12%)
OTU7 (4; 3.12%)
OTU8 (4; 3.12%)
OTU9 (3; 2.34%)
OTU10 (3; 2.34%)
OTU11 (3; 2.34%)
OTU12 (3; 2.34%)
OTU13 (2; 1.56%)
OTU14 (2; 1.56%)
OTU15 (2; 1.56%)
OTU1 (22; 17.32%)
OTU2 (8 ; 6.30%)
OTU3 (5; 3.94%)
OTU4 (5; 3.94%)
OTU5 (5; 3.94%)
OTU6 (3; 2.36%)
OTU7 (3; 2.36%)
OTU8 (3; 2.36%)
OTU9 (3; 2.36%)
OTU10 (2; 1.57%)
OTU11 (2; 1.57%)
OTU12 (2; 1.57%)
OTU13 (2; 1.57%)
OTU14 (2; 1.57%)
OTU15 (2; 1.57%)
OTU16 (2; 1.57%)
OTU17 (2; 1.57%)
OTU1 (3; 2.40%)
OTU2 (3; 2.40%)
OTU3 (3; 2.40%)
OTU4 (3; 2.40%)
OTU5 (2; 1.60%)
OTU6 (2; 1.60%)
OTU7 (2; 1.60%)
OTU8 (2; 1.60%)
OTU9 (2; 1.60%)
Apusozoa
Count
0
400
0 1 2 3
Value
Cryptophyta
0
500
0 1 2 3
Choanoflagellida
0
600
0 1 2 3
Acantharea
0
300
0 0.6 1.2

Figure S5

Table 1. Sample index for Saanich Inlet, Cariaco Basin and Framvaren Fjord microbial eukaryotes

Sample ID	Location	Station[a,b]	Date (mm/yy)	Latitude, Longitude	Description	Oxytype	Depth (m)	O_2 (nM)[c]	H_2S (μM)	NO_3^- (μM)	CH_4 (nM)	Size Fraction (μ)[A,B,C]	Sequences
1	Saanich Inlet	SI03	02/06	48.35°N, 123.30°W	surface	oxic	10	212.0	0.0	26.7	123.0	2.7-0.2	324
2	Saanich Inlet	SI03	07/06	48.35°N, 123.30°W	surface	oxic	10	381.6	0.0	4.6	94.0	2.7-0.2	147
3	Saanich Inlet	SI03	11/06	48.35°N, 123.30°W	surface	oxic	10	249.0	0.0	25.5	47.7	2.7-0.2	283
4	Saanich Inlet	SI03	04/07	48.35°N, 123.30°W	surface	oxic	10	316.1	0.0	18.2	45.0	2.7-0.2	695
5	Saanich Inlet	SI03	04/08	48.35°N, 123.30°W	surface	oxic	10	263	0.0	19.2	28.8	2.7-0.2	ND
6	Saanich Inlet	SI03	02/06	48.35°N, 123.30°W	oxic-anoxic interface	dysoxic	100	51.1	0.0	21.0	93.8	2.7-0.2	243
7	Saanich Inlet	SI03	07/06	48.35°N, 123.30°W	oxic-anoxic interface	dysoxic	100	22.9	0.0	16.1	59.6	2.7-0.2	270
8	Saanich Inlet	SI03	11/06	48.35°N, 123.30°W	oxic-anoxic interface	suboxic	100	15.4	0.0	13.0	38.2	2.7-0.2	360
9	Saanich Inlet	SI03	04/07	48.35°N, 123.30°W	oxic-anoxic interface	dysoxic	100	67.3	0.0	26.5	14.8	2.7-0.2	355
10	Saanich Inlet	SI03	04/08	48.35°N, 123.30°W	oxic-anoxic interface	dysoxic	100	120.0	0.0	20.5	22.4	2.7-0.2	326
11	Saanich Inlet	SI03	02/06	48.35°N, 123.30°W	lower redox transition	suboxic	125	5.0	0.0	9.7	145.0	2.7-0.2	344
12	Saanich Inlet	SI03	07/06	48.35°N, 123.30°W	lower redox transition	suboxic	120	6.5	0.0	4.8	167.0	2.7-0.2	328
13	Saanich Inlet	SI03	11/06	48.35°N, 123.30°W	lower redox transition	suboxic	120	9.8	0.0	8.9	22.1	2.7-0.2	278
14	Saanich Inlet	SI03	04/07	48.35°N, 123.30°W	lower redox transition	dysoxic	120	26.9	0.0	20.6	9.0	2.7-0.2	267
15	Saanich Inlet	SI03	04/08	48.35°N, 123.30°W	lower redox transition	suboxic	120	18.1	0.0	13.2	7.1	2.7-0.2	336
16	Saanich Inlet	SI03	02/06	48.35°N, 123.30°W	anoxic deep	anoxic	215	1.4	ND	1.8	734.0	2.7-0.2	345
17	Saanich Inlet	SI03	07/06	48.35°N, 123.30°W	anoxic deep	anoxic	200	0.0	ND	0.5	1273.8	2.7-0.2	343
18	Saanich Inlet	SI03	11/06	48.35°N, 123.30°W	anoxic deep	dysoxic	200	54	ND	19.8	19.9	2.7-0.2	339
19	Saanich Inlet	SI03	04/07	48.35°N, 123.30°W	anoxic deep	anoxic	200	1.1	5.6	0.0	857.8	2.7-0.2	334
20	Saanich Inlet	SI03	04/08	48.35°N, 123.30°W	anoxic deep	anoxic	200	0.0	2.1	0.1	708.3	2.7-0.2	375
21	Cariaco Basin	A	01/05	10.50°N, 64.66°W	above interface	suboxic	240	8.4	0.0			≥4.5	600
22	Cariaco Basin	A	05/05	10.50°N, 64.66°W	above interface	dysoxic	185	66.3	0.0			≥4.5	744
23	Cariaco Basin	BC	01/05	B;10.40°N, 64.46°W/C;10.40°N, 65.35°W	above interface	anoxic	260/240	0.0/3.1	4.0/0.0			≥4.5	356
24	Cariaco Basin	BC	05/05	B;10.40°N, 64.46°W/C;10.40°N, 65.35°W	above interface	suboxic	216/230	12.5/23.5	0.0/0.0			≥4.5	519
25	Cariaco Basin	A	01/05	10.50°N, 64.66°W	oxic-anoxic interface	anoxic	280	0.0	1.1			≥4.5	495
26	Cariaco Basin	A	05/05	10.50°N, 64.66°W	oxic-anoxic interface	suboxic	225	8.7	0.0			≥4.5	553
27	Cariaco Basin	BC	01/05	B;10.40°N, 64.46°W/C;10.40°N, 65.35°W	oxic-anoxic interface	anoxic	300/280	0.0/0.0	4.7/0.9			≥4.5	153
28	Cariaco Basin	BC	05/05	B;10.40°N, 64.46°W/C;10.40°N, 65.35°W	oxic-anoxic interface	anoxic	256/270	0.0/0.0	0.3/0.1			≥4.5	536
29	Cariaco Basin	A	01/05	10.50°N, 64.66°W	upper sulfidic boundary	anoxic	320	0.0	4.0			≥4.5	565
30	Cariaco Basin	A	05/05	10.50°N, 64.66°W	upper sulfidic boundary	anoxic	265	0.0	0.6			≥4.5	394
31	Cariaco Basin	BC	01/05	B;10.40°N, 64.46°W/C;10.40°N, 65.35°W	upper sulfidic boundary	anoxic	340/320	0.0/0.0	4.7/3.7			≥4.5	162
32	Cariaco Basin	BC	05/05	B;10.40°N, 64.46°W/C;10.40°N, 65.35°W	upper sulfidic boundary	anoxic	296/310	0.0/0.0	4.7/1.4			≥4.5	433
33	Cariaco Basin	A	01/05	10.50°N, 64.66°W	anoxic deep	anoxic	900	0.0	53.1			≥4.5	140
34	Cariaco Basin	A	05/05	10.50°N, 64.66°W	anoxic deep	anoxic	900	0.0	51.5			≥4.5	400
35	Cariaco Basin	BC	01/05	B;10.40°N, 64.46°W/C;10.40°N, 65.35°W	anoxic deep	anoxic	400/900	0.0/0.0	10.9/45.3			≥4.5	39
36	Cariaco Basin	BC	05/05	B;10.40°N, 64.46°W/C;10.40°N, 65.35°W	anoxic deep	anoxic	670/900	0.0/0.0	24.5/52.4			≥4.5	400
37	Framvaren Fjord		05/04	58.09°N, 06.45°E	oxic-anoxic interface	anoxic	18	0.0	0.0			≥6.5	76
38	Framvaren Fjord		11/04	58.09°N, 06.45°E	oxic-anoxic interface	anoxic	18	0.0	0.0			≥6.5	55
39	Framvaren Fjord		09/05	58.09°N, 06.45°E	oxic-anoxic interface	oxic	18	100.0	0.0			≥6.5	127
40	Framvaren Fjord		05/04	58.09°N, 06.45°E	upper sulfidic boundary	anoxic	20	0.0	180.0			≥6.5	74
41	Framvaren Fjord		11/04	58.09°N, 06.45°E	upper sulfidic boundary	anoxic	20	0.0	470.0			≥6.5	79
42	Framvaren Fjord		09/05	58.09°N, 06.45°E	upper sulfidic boundary	anoxic	20	0.0	20.0			≥6.5	42
43	Framvaren Fjord		05/04	58.09°N, 06.45°E	anoxic intermediate	anoxic	36	0.0	600.0			≥6.5	52
44	Framvaren Fjord		11/04	58.09°N, 06.45°E	anoxic intermediate	anoxic	36	0.0	470.0			≥6.5	54
45	Framvaren Fjord		09/05	58.09°N, 06.45°E	anoxic intermediate	anoxic	36	0.0	670.0			≥6.5	108

[a] Cariaco Basin samples from sites B and C were composited prior to sequencing based on related depth intervals with shared physical and chemical properties [ref].

[b] Only full-length amplicon sequences from Framvaren Fjord sequences were deposited in Genbank [ref].

[c] Limit of detection for oxygen sensor in Saanich Inlet is 3μM

ND; not done

A; GF/D prefilters (2.7 mm pore size) in-line with 0.22mm Sterivex-GV filters (Millipore)

B; 47mm Durapore membranes (0.45 mm pore size)

C; 47mm Durapore membranes (0.65 mm pore size)

oxytype ln [μM]O2; oxic >4.5, dysoxic 3.0-4.5, suboxic 0.5-3.0, anoxic <0.5

Table 2. Predicted richness of protistan assemblages in Saanich Inlet, Cariaco Basin and Framvaren Fjord

		Saanich Inlet	Framvaren Fjord[a]	Cariaco Basin[b]
% Identity of sequences within OTU	Statistic	Best Parametric Model	Best Parametric Model	Best Parametric Model
≥99		2-mixed exponential	2-mixed exponential	3-mixed exponential
	Estimated total number of OTUs	13442	3703	35968
	Standard Error	3,795	2,995	13,123
	Confidence Interval	7,963-23,373	1,090-15,370	18,373-72,596
	Goodness-of-fit	0.12	0.26	0.14
≥98		3-mixed exponential	3-mixed exponential	3-mixed exponential
	Estimated total number of OTUs	8176	1055	12480
	Standard Error	2,326	290	3489
	Confidence Interval	4,861-14,333	659-1,854	7,480-21,666
	Goodness-of-fit	0.03	0.02	0.82
≥95		3-mixed exponential	2-mixed exponential	2-mixed exponential
	Estimated total number of OTUs	2687	457	4039
	Standard Error	1,023	58	798
	Confidence Interval	1,440-5,778	366-597	2,815-6,015
	Goodness-of-fit	0.03	0.02	0.06
≥90		3-mixed exponential	2-mixed exponential	3-mixed exponential
	Estimated total number of OTUs	510	232	1143
	Standard Error	99	28	162
	Confidence Interval	376-781	189-300	887-1,532
	Goodness-of-fit	0.015	0.1	0.68

SE; standard error

GOF; godness-of-fit (p-value for the corrected Pearson chi-square goodness-of-fit test

CB; 95% confidence interval

NA; not applicable

[a] Sequence data from Behnke et al., 2010

[b] Sequence data from Edgcomb et al., 2011

Table S1. Ciliate and Euglenozoa 18S rRNA gene sequences recovered from the July 2006 200 meter Saanich Inlet clone library

	Closest Described Species			Closest Uncultured Clone		
Sequences	Species Name	Genbank Accession Numbe	% Identity	Clone Name	Genbank Accession Numbe	% Identity
1	*Gonostomum strenuum*	AJ310493	92	RM2-SGM17	AB505525	98
2	*ardiostomatella vermiforme*	AY881632	90	M2_18E08	DQ103844	91
17	*Nyctotherus cordiformis*	AJ006712	89	NA1_4G10	EF526790	90
73	*Strombidium conicum*	FJ422992	89-91	MA1_CilA12	EF526985	95-96
5	*Strombidium stylifer*	DQ631805	97	UEPAC05Np2	AY129035	99
4	*Strombidium biarmatum*	AY541684	99	SIF_1F2	EF527106	99
3	*Strombidium sp. SBB99-1*	AY143565	99	C7p3_ML_383	FJ353184	99
2	*Novistrombidium orientale*	FJ422988	93	NPKS2_T13	EU371394	93
1	*Uronychia binucleata*	EF198667	98	SA2_2A2	EF527163	98
1	*Condylostoma spatiosum*	DQ822483	89	NA1_4G10	EF526790	90
1	*Protogastrostyla pulchra*	EF194082	93	MA1_CilA12	EF526985	96
5	*Cyclidium glaucoma*	DQ442840	92	NA1_3B7	EF526828	97
38	*Homalogastra setosa*	EF158848	96	RM2-SGM02	AB505510	96
1	*Strobilidium sp.*	AF399123	99	A3	FN263264	99
44	*Calkinsia aureus*	EU753419	87-94	H52	AY256215	89-98
13	*Cyclidium porcatum*	Z29517	90-92	H60	AY256216	94-96
2	*Meseres corlissi*	EU399529	90	MA1_CilA12	EF526985	96

Table S2. Non-parametric diversity estimates for Saanich Inlet, Cariaco Basin and Framvaren Fjord eukaryotic microbial communities

									chao (S)			simpson (D)			shannon (H)		
Sample ID	Location	Site	Date (mm/yy)	Description	Oxytype	Depth (m)	Sequences	OTUs	average	lower	upper	average	lower	upper	average	lower	upper
1	Saanich Inlet	SI03	02/06	surface	oxic	10	324	182	967	624	1577	0.029	0.020	0.039	4.51	4.34	4.68
2	Saanich Inlet	SI03	07/06	surface	oxic	10	147	138	964	562	1745	0.001	0.000	0.002	4.91	4.79	5.02
3	Saanich Inlet	SI03	11/06	surface	oxic	10	283	182	1114	702	1849	0.034	0.021	0.047	4.54	4.35	4.73
4	Saanich Inlet	SI03	04/07	surface	oxic	10	695	254	484	399	616	0.010	0.008	0.012	5.02	4.93	5.11
5	Saanich Inlet	SI03	04/08	surface	oxic	10											
6	Saanich Inlet	SI03	02/06	oxic-anoxic interface	dysoxic	100	243	119	335	232	533	0.015	0.010	0.019	4.39	4.26	4.53
7	Saanich Inlet	SI03	07/06	oxic-anoxic interface	dysoxic	100	270	175	873	572	1405	0.025	0.014	0.035	4.63	4.46	4.80
8	Saanich Inlet	SI03	11/06	oxic-anoxic interface	suboxic	100	360	172	413	311	589	0.009	0.007	0.011	4.82	4.71	4.92
9	Saanich Inlet	SI03	04/07	oxic-anoxic interface	dysoxic	100	355	183	354	287	466	0.008	0.006	0.009	4.91	4.81	5.01
10	Saanich Inlet	SI03	04/08	oxic-anoxic interface	dysoxic	100	326	159	442	321	650	0.027	0.018	0.036	4.46	4.30	4.61
11	Saanich Inlet	SI03	02/06	lower redox transition	suboxic	125	344	106	185	147	258	0.078	0.056	0.100	3.67	3.49	3.84
12	Saanich Inlet	SI03	07/06	lower redox transition	suboxic	120	328	50	102	70	181	0.461	0.394	0.528	1.76	1.54	1.99
13	Saanich Inlet	SI03	11/06	lower redox transition	suboxic	120	278	82	172	123	280	0.049	0.035	0.063	3.67	3.52	3.83
14	Saanich Inlet	SI03	04/07	lower redox transition	dysoxic	120	267	138	343	251	510	0.012	0.008	0.016	4.59	4.46	4.71
15	Saanich Inlet	SI03	04/08	lower redox transition	suboxic	120	336	139	303	231	429	0.055	0.036	0.073	4.11	3.93	4.29
16	Saanich Inlet	SI03	02/06	anoxic deep	anoxic	215	345	83	155	118	233	0.109	0.083	0.134	3.19	3.01	3.37
17	Saanich Inlet	SI03	07/06	anoxic deep	anoxic	200	343	108	238	176	359	0.071	0.053	0.089	3.64	3.46	3.82
18	Saanich Inlet	SI03	11/06	anoxic deep	dysoxic	200	339	99	213	156	328	0.068	0.050	0.086	3.62	3.46	3.79
19	Saanich Inlet	SI03	04/07	anoxic deep	anoxic	200	334	20	37	24	84	0.781	0.721	0.841	0.67	0.50	0.83
20	Saanich Inlet	SI03	04/08	anoxic deep	anoxic	200	375	35	52	40	91	0.628	0.564	0.692	1.20	1.00	1.39
21	Cariaco Basin	A	01/05	above interface	suboxic	240	600	491	3113	2299	4293	0.001	0.001	0.002	6.06	5.99	6.13
22	Cariaco Basin	A	05/05	above interface	dysoxic	185	744	619	4407	3295	5981	0.001	0.001	0.001	6.32	6.26	6.38
23	Cariaco Basin	BC	01/05	above interface	anoxic	260	356	273	1349	957	1965	0.004	0.002	0.005	5.41	5.30	5.51
24	Cariaco Basin	BC	05/05	above interface	suboxic	216	519	437	2487	1848	3417	0.001	0.001	0.002	5.97	5.90	6.04
25	Cariaco Basin	A	01/05	oxic-anoxic interface	anoxic	280	495	441	3363	2380	4843	0.001	0.000	0.001	6.02	5.96	6.09
26	Cariaco Basin	A	05/05	oxic-anoxic interface	suboxic	225	553	457	2348	1779	3160	0.001	0.001	0.002	6.01	5.94	6.08
27	Cariaco Basin	BC	01/05	oxic-anoxic interface	anoxic	300	153	113	732	409	1405	0.012	0.005	0.020	4.46	4.30	4.63
28	Cariaco Basin	BC	05/05	oxic-anoxic interface	anoxic	256	536	504	9532	5756	16023	0.000	0.000	0.001	6.18	6.12	6.25
29	Cariaco Basin	A	01/05	upper sulfidic boundary	anoxic	320	565	389	1378	1083	1800	0.006	0.004	0.009	5.64	5.54	5.74
30	Cariaco Basin	A	05/05	upper sulfidic boundary	anoxic	265	394	331	2790	1827	4371	0.002	0.001	0.002	5.70	5.61	5.78
31	Cariaco Basin	BC	01/05	upper sulfidic boundary	anoxic	340	162	154	1330	751	2471	0.001	0.000	0.001	5.02	4.91	5.13
32	Cariaco Basin	BC	05/05	upper sulfidic boundary	anoxic	296	433	331	1680	1217	2385	0.004	0.002	0.006	5.58	5.49	5.68
33	Cariaco Basin	A	01/05	anoxic deep	anoxic	900	140	123	543	349	906	0.002	0.000	0.004	4.76	4.63	4.88
34	Cariaco Basin	A	05/05	anoxic deep	anoxic	900	400	332	2102	1460	3110	0.001	0.001	0.002	5.71	5.63	5.79
35	Cariaco Basin	BC	01/05	anoxic deep	anoxic	400	39	35	200	91	523	0.007	-0.003	0.016	3.51	3.27	3.74
36	Cariaco Basin	BC	05/05	anoxic deep	anoxic	670	400	310	1481	1080	2093	0.005	0.003	0.008	5.50	5.39	5.60
37	Framvaren Fjord		05/04	oxic-anoxic interface	anoxic	18	76	56	217	124	438	0.022	0.005	0.039	3.78	3.55	4.01
38	Framvaren Fjord		11/04	oxic-anoxic interface	anoxic	18	55	47	262	127	625	0.008	0.000	0.017	3.77	3.56	3.98
39	Framvaren Fjord		09/05	oxic-anoxic interface	oxic	18	127	69	168	113	294	0.018	0.010	0.026	3.97	3.80	4.13
40	Framvaren Fjord		05/04	upper sulfidic boundary	anoxic	20	74	63	448	212	1055	0.006	0.001	0.010	4.07	3.89	4.25
41	Framvaren Fjord		11/04	upper sulfidic boundary	anoxic	20	79	68	282	164	546	0.004	0.001	0.008	4.16	4.00	4.33
42	Framvaren Fjord		09/05	upper sulfidic boundary	anoxic	20	42	37	169	84	408	0.007	-0.001	0.015	3.56	3.34	3.78
43	Framvaren Fjord		05/04	anoxic intermediate	anoxic	36	52	33	384	200	770	0.037	0.011	0.063	3.23	2.96	3.49
44	Framvaren Fjord		11/04	anoxic intermediate	anoxic	36	54	46	477	193	1306	0.009	0.000	0.018	3.74	3.52	3.95
45	Framvaren Fjord		09/05	anoxic intermediate	anoxic	36	108	74	264	161	490	0.013	0.006	0.019	4.10	3.93	4.27

chao (S); Sobs + (a2/2b) where Sobs is the number of species observed, is the number of species observed just once and b is the number of species observed just twice.

simpson (D); $S\ (n+N)^2$ where n = the total number of organisms of a particular species and N = the total number of organisms of all species

shannon (H); S - (Pi * ln Pi) - [S-1/2N)] where Pi = fraction of the entire population made up of species I, S = numbers of species encountered

* Based on the average value for composited samples.

oxytype ln $[\mu M]_{O2}$; oxic >4.5, dysoxic 3.0-4.5, suboxic 0.5-3.0, anoxic <0.5