ORIGINAL ARTICLE

Effect of oxygen minimum zone formation on communities of marine protists

William Orsi[1,2,5], Young C Song[3,4,5], Steven Hallam[3,4] and Virginia Edgcomb[1]

[1]Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, MA, USA; [2]Center for Dark Energy Biosphere Investigations, Los Angeles, CA, USA; [3]Department of Microbiology and Immunology, University of British Columbia, Vancouver, Canada and [4]Graduate Program in Bioinformatics, University of British Columbia, Vancouver, Canada

Changes in ocean temperature and circulation patterns compounded by human activities are leading to oxygen minimum zone (OMZ) 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 OMZs.

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). O_2-deficiency selects for microbial groups capable of utilizing alternative respiratory substrates including nitrate (NO_3^-), NO_2^- (nitrite), Mn (manganese), Fe (iron), SO_4^- (sulfate) or carbon dioxide (CO_2) (Zehnder and Stumm, 1988). Within O_2-deficient waters, the use of NO_2^- or NO_3^- as alternative electron acceptors results in the production of N_2O (nitrous oxide) 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 O_2 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., 2011c), 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 O_2-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 O_2 gradients (Behnke et al.,

2006; Zuendorf et al., 2006; Behnke et al., 2010; Edgcomb et al., 2011a, b; 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 with related studies in anoxic Cariaco Basin and Framvaren Fjord and used to identify common and unique patterns of community composition.

Materials and methods

Sample collection and processing
Samples from Saanich Inlet were collected and processed as described previously (Zaikova et al., 2010; Walsh and Hallam, 2011) 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 16 depth intervals including cell counts, temperature, salinity, O_2, NO_3^-, PO_4^{3-} (phosphate), SiO_4^- (silicate), NO_2^-, N_2O, NH_4^+ (ammonia), CO_2, CH_4 and 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, b), and (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://venus.uvic.ca).

PCR amplification of 18S rRNA genes
DNA extracts from 10, 100, 120 and 200 m depth intervals were amplified using small subunit ribosomal DNA primers targeting the eukaryotic domain: Euk515F (5'-GTGCCAAGCAGCCGCGG TAA) and Euk1209R (5'-GACGGGCRGTGWGTRCA) under the following PCR conditions. PCR conditions: 2 min at 95 °C followed by 20 cycles of 95 °C for 40 s, 55 °C for 30 s, 72 °C for 90 s and final extension of 7 min at 72 °C. Each 50 μl reaction contained 1 μl of template DNA and 1 μl each 0.4 μM forward and reverse primer added to a PCR Master Mix (Stratagene, Santa Clara, CA, USA, Cat #600640). Reactions were aliquoted into 3×15 μl reactions before PCR (to minimize bias) and re-pooled after PCR. For specific details see URL: http://my.jgi.doe.gov/general/protocols/SOP_16S18S_rRNA_PCR_Library_Creation.pdf

Clone library construction and sequencing
Resulting amplicons were gel purified using the MiniElute gel extraction Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Ligation, transformation and sequencing steps were performed as described in Zaikova et al. (2009). The number of resulting transformants per ligation ranged between ~100 000 and 800 000 colony forming units. One 384-well plate per sample ligation was sequenced with variable success using M13F and M13R primers as described at http://jgi.doe.gov/sequencing/protocols/prots_production.html.

Quality control of 18S SSU rRNA data set
The sequences were checked for chimeras using the Bellerophon Chimera Check and the Check_Chimera utilities (Ribosomal Database Project) (Cole et al., 2003). After removal of putative chimeras, bacterial, archaeal and metazoan sequences, the remaining sequences were grouped into operational taxonomic units (OTUs) based on 98% rRNA gene sequence similarity levels. This was achieved by first making all possible pairwise sequence alignments by using ClustalW (Thompson et al., 1994), calculating percentage sequence identities, followed by clustering the sequences by using the unweighted pair group method with arithmetic mean 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 (nine 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 (Patrick Schloss, Detroit, MI, USA) (Schloss, 2009), with gap penalty of −1 and k-mer size of 9. The resulting alignment file was used as an input to generate 6768 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 (Bremen, Germany) 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 α diversity (Supplementary 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 sample belonging to a specific taxonomic group. From this table, the relative abundance of each protistan group in a given sample was calculated as a percentage value by dividing the raw number of sequences associated with the specific taxa by the total number of sequences in the sample. Circos was used to generate circular link diagrams illustrating community composition differences among and between samples (Krzywinski *et al.*, 2009).

Histoheatmap generation
The same table used in Circos visualization was also used to generate combined histograms and heatmaps (that is, histoheatmaps illustrating the presence of different OTUs across each sample) using the R statistical software package (http://www.r-project.org). This script is available upon request from the authors.

Phylogenetic analysis
For our phylogenetic analyses, we focused on ciliate and euglenozoan-affiliated sequences in the July 200 m sample. We focused on this sample because it contained the most novel sequences of all clone libraries/samples in our study. For these analyses, we used representative sequences from each ciliate and euglenozoan-affiliated OTU clustered at the 98% sequence identity level. Representative sequences were compared against the Genbank-nt database using BLASTn in search of their closest relatives and the highest scoring cultured and uncultured sequence relatives were retrieved. Sequences were aligned using the ARB automated aligner (Bremen, Germany) (Ludwig *et al.*, 2004), the alignment was manually refined using secondary structure information, and only unambiguous positions were used to construct phylogenetic trees. Phylogenetic trees were constructed using Bayesian inference (Ronquist and Huelsenbeck, 2003) and PhyML (Montpellier, France) (Guindon *et al.*, 2005).

Statistical analyses
After clustering of our Sanger sequence data set, 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 (Ithaca, NY, USA) (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 O_2, NO_3^-, CH_4 and H_2S. Multiresponse permutation procedure (MRPP) was used to test for a statistically significant influence of season, depth, NO_3^-, sulfide and O_2 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 data set clustered at the 98% sequence identity threshold. Monte Carlo tests, MRPP and CCA were implemented using the PC-ORD software package (MjM Software Design, Gleneden Beach, OR, USA).

Principal component and hierarchical clustering analysis
To determine the correlation between protistan community structure and the O_2 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).

Results

Community diversity measures
We sampled four depth intervals over 3 years (2006–2008) during the months of February, April, July and November, representing different water column redox states resulting from seasonal stratification and renewal (Walsh *et al.*, 2009; Zaikova *et al.*, 2010). A total of 4987 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 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; Edgcomb *et al.*, 2011a; Orsi *et al.*, 2011b). At the 99% and 98% sequence identity thresholds, we estimate 13 442 ($-/+$: 7963–23 373 CI (95% confidence interval)) and 8176 ($-/+$: 4861–14 333 CI) taxa, respectively. At 95% and 90% sequence identity, we estimate 2687 ($-/+$: 1440–5778 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 data sets 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 three sites were lower than parametric estimates (Supplementary Table S2).

Protistan community structure in Saanich Inlet
CCA showed a clear division of Saanich Inlet protistan communities into different clusters associated with oxic ($>90\,\mu$M), dysoxic (20–90 μM), suboxic (1–20) μM or anoxic/sulfidic ($<1\,\mu$M/ ± 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 O_2, NO_3^-, CH_4 and sulfide had a *P*-value of 0.01. MRPP tests of influence of season, depth, O_2, NO_3^-, CH_4 and sulfide on the OTU distribution all yielded *P*-values $\leqslant 0.01$.

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 (Figures 2 and 3). At 200 m in July, 66% and 11% of sequences were affiliated with the Ciliophora and Euglenozoa, respectively (Figures 2 and 3), while in November, 70% of sequences were affiliated with the Dinophyceae, 80% of which were affiliated with the subgroup Syndiniales (Figures 2 and 3). The majority of sequences recovered from anoxic waters exhibited low ($<92\%$) sequence identities with their closest described relatives (Supplementary Table S1) and formed new lineages based on phylogenetic trees (Figures 4 and 5). The uncultured environmental sequences with the highest identities to many of these sequences were recovered from O_2-deficient marine waters from the Cariaco Basin (Stoeck *et al.*, 2003), Framvaren Fjord (Behnke *et al.*, 2006; Behnke *et al.*, 2010) and Mariager Fjord (Denmark) (Zuendorf *et al.*, 2006).

At 10 m, Dinophyceae- and Stramenopile-affiliated sequences dominated during all seasons (Figure 2; Supplementary Figure S4). The majority (90%) of Dinophyceae-affiliated sequences were related to the parasitic Syndiniales. The same observation was made at 100 m in February, April and November while in July, radiolarian-affiliated sequences were most abundant (Figure 2; Supplementary Figure S4). The number of cercozoan-affiliated sequences increased in April at 10 and 100 m, but were less abundant during the rest of the year (Figures 2 and 3). The occurrence of larger-sized protists (such as the Radiolaria and Cercozoa) in pico-eukaryote size clone libraries has also been reported in previous studies (Not *et al.*, 2009)

Taxonomic relationships between Saanich Inlet, Framvaren Fjord and Cariaco Basin
To better constrain protistan community structure and dynamics to changing levels of O_2 and sulfide on a global scale, we compared available 18S rRNA sequence data sets from Saanich Inlet, Cariaco Basin and Framvaren Fjord. Overall, taxonomic assignments for the Cariaco Basin and Framvaren Fjord (Supplementary Figures S1 and S2) data sets were congruent with previous studies (Behnke *et al.*,

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)	O₂ (mM)[c]	H₂S (µM)	NO₃ (µM)	CH₄ (nM)	Size fraction (µ)[A,B,C]	Sequences	Library
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	SGPW
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	SGSO
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	SHAB
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	SHAW
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	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	SGPZ
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	SGST
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	SHAG
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	SHAZ
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	SHZW
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	SGSC
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	SGSX
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	SHAN
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	SHBC
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	SHZZ
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	SGSH
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	SGTA
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	SHAS
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	SHBH
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	SIAC
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	

Table 1 (Continued)

Sample ID	Location	Station[a,b]	Date (mm/yy)	Latitude, longitude	Description	Oxytype	Depth (m)	O_2 (mM)[c]	H_2S (µM)	NO_3^- (µM)	CH_4 (nM)	Size fraction (µ)[A,B,C]	Sequences	Library
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	M-IF
38	Framvaren Fjord		11/04	58.09°N, 06.45°E	Oxic–anoxic interface	Anoxic	18	0.0	0.0			≥6.5	55	N-IF
39	Framvaren Fjord		09/05	58.09°N, 06.45°E	Oxic–anoxic interface	Oxic	18	100.0	0.0			≥6.5	127	S-IF
40	Framvaren Fjord		05/04	58.09°N, 06.45°E	Upper sulfidic boundary	Anoxic	20	0.0	180.0			≥6.5	74	M-A1
41	Framvaren Fjord		11/04	58.09°N, 06.45°E	Upper sulfidic boundary	Anoxic	20	0.0	470.0			≥6.5	79	N-A1
42	Framvaren Fjord		09/05	58.09°N, 06.45°E	Upper sulfidic boundary	Anoxic	20	0.0	20.0			≥6.5	42	S-A1
43	Framvaren Fjord		05/04	58.09°N, 06.45°E	Anoxic intermediate	Anoxic	36	0.0	600.0			≥6.5	52	M-A2
44	Framvaren Fjord		11/04	58.09°N, 06.45°E	Anoxic intermediate	Anoxic	36	0.0	470.0			≥6.5	54	N-A2
45	Framvaren Fjord		09/05	58.09°N, 06.45°E	Anoxic intermediate	Anoxic	36	0.0	670.0			≥6.5	108	S-A2

Abbreviations: CH_4, methane; H_2S, hydrogen sulfide; ND, not done; NO_3^-, nitrate; O_2, oxygen.

[a]Cariaco Basin samples from sites B and C were composited before 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 3 µM.

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

B denotes 47 mm Durapore membranes (0.45 mm pore size).

C denotes 47 mm Durapore membranes (0.65 mm pore size).

Oxytype ln [µM]O_2: oxic >4.5, dysoxic 3.0–4.5, suboxic 0.5–3.0, anoxic <0.5.

Summary of eukaryotic SSU ribosomal RNA gene (18S rRNA) sequence data, as well as geochemical data, generated from the Saanich Inlet water column spanning the 2006–2008 time series.

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

% Identity of sequences within OTU	Statistic	Saanich Inlet	Framvaren Fjord[a]	Cariaco Basin[b]
		Best parametric model	Best parametric model	Best parametric model
⩾99		Two-mixed exponential	Two-mixed exponential	Three-mixed exponential
	Estimated total number of OTUs	13442	3703	35968
	Standard error	3795	2995	13 123
	Confidence interval	7963–23 373	1090–15 370	18 373–72 596
	Goodness-of-fit	0.12	0.26	0.14
⩾98		Three-mixed exponential	Three-mixed exponential	Three-mixed exponential
	Estimated total number of OTUs	8176	1055	12480
	Standard error	2326	290	3489
	Confidence interval	4861–14 333	659–1854	7480–21 666
	Goodness-of-fit	0.03	0.02	0.82
⩾95		Three-mixed exponential	Two-mixed exponential	Two-mixed exponential
	Estimated total number of OTUs	2687	457	4039
	Standard error	1023	58	798
	Confidence interval	1440–5778	366–597	2815–6015
	Goodness-of-fit	0.03	0.02	0.06
⩾90		Three-mixed exponential	Two-mixed exponential	Three-mixed exponential
	Estimated total number of OTUs	510	232	1143
	Standard error	99	28	162
	Confidence interval	376–781	189–300	887–1532
	Goodness-of-fit	0.015	0.1	0.68

Abbreviations: GOF, goodness-of-fit (*P*-value for the corrected Pearson χ^2 goodness-of-fit test; OTU, operational taxonomic unit.
[a]Sequence data are from Behnke *et al.* (2010).
[b]Sequence data are from Edgcomb *et al.* (2011a).
Parametric predicted richness of protistan assemblages in Saanich Inlet, Framvaren Fjord (data are taken from Behnke *et al.*, 2010) and the Cariaco Basin (data are taken from Edgcomb *et al.*, 2011a) with associated statistics.

2010; Edgcomb *et al.*, 2011a; Orsi *et al.*, 2011b). Sequences affiliated with the Ciliophora, Euglenozoa, Choanoflagellata, Stramenopiles, Fungi and Dinophyceae were well represented from anoxic samples in all three sites (Figure 3; Supplementary Figures S1–S5). However, many taxa were differentially represented. Examples include Stramenopile-affiliated sequences that were more prevalent in Cariaco Basin and Framvaren Fjord than Saanich Inlet (Supplementary Figures S1–S4), as well as Polycystinea- and fungal-affiliated sequences that were more abundant in Cariaco Basin relative to Saanich Inlet and Framvaren (Figures 2 and 3; Supplementary Figures S1–S5). Sequences affiliated with the Stramenopiles, Cercozoa, Dinophyceae, Polycystinea and Acantharea were represented in oxygenated samples from both the Cariaco Basin and Saanich Inlet (Figures 2 and 3; Supplementary Figures S1–S5).

Combined principal component (PCA) and hierarchical cluster analyses of the Saanich Inlet, Cariaco Basin (Edgcomb *et al.*, 2011a; Orsi *et al.*, 2011b) and Framvaren Fjord (Behnke *et al.*, 2010) data sets revealed biogeographic and niche-specific clustering patterns. For the most part, protistan communities clustered by location according to depth and O_2 concentration. The majority of oxic, dysoxic and suboxic samples from Saanich Inlet formed a nested series in one cluster, with anoxic Cariaco Basin and Framvaren Fjord samples, forming independent clusters. Interestingly, anoxic/sulfidic samples from Saanich Inlet and one anoxic deep sample from the Cariaco western subbasin (site BC) clustered with the hyper-sulfidic Framvaren Fjord samples (Figure 6).

Fluctuations in rare microbial populations
Our comparison of microbial populations during stratification and renewal suggests selection for growth of many low abundance taxa once the preferred conditions arise. Examples include an increase in the number of Ciliate- and Stramenopile-affiliated sequences recovered at 200 m during periods of anoxia (Figures 2 and 3; Supplementary Figure S4). One Stramenopile-affiliated OTU was detected 228 times in July and 11 times in November (Supplementary Figure S4). Furthermore, an increase in Dinophyceae and Cercozoan-affiliated sequences at 120 and 200 m was observed after the oxygenated renewal event in autumn (Figures 2 and 3; Supplementary Figure S4). Several Dinophyceae-affiliated OTUs detected in July and November at 200 m increased significantly in size in November after the renewal event (Supplementary Figure S4).

Discussion

Saanich Inlet provides a model ecosystem for studying microbial community responses to changes in dissolved O_2 concentrations, that is, OMZ formation, because the bottom 80 m of the water column (120–200 m) annually fluctuates between oxyge-
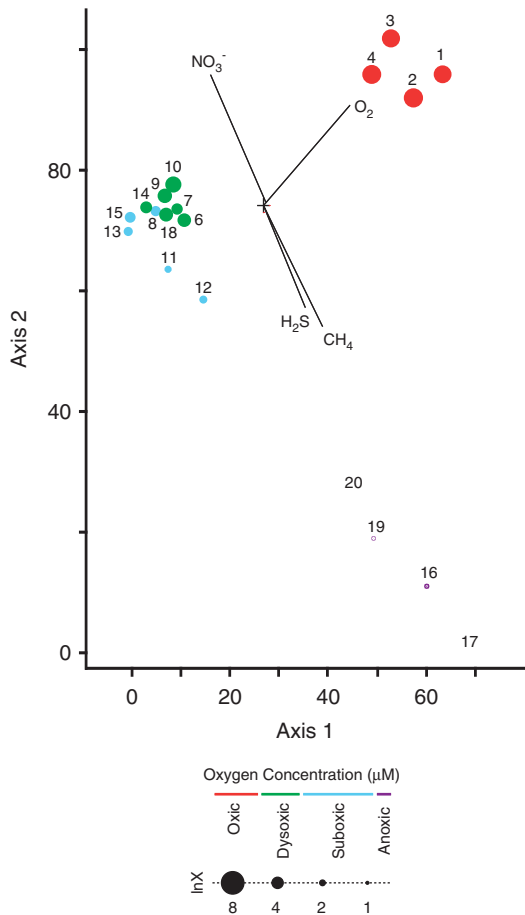
Figure 1 CCA of the Saanich Inlet 18S rRNA gene sequence data set clustered at the 98% identity threshold. Samples are represented in the biplot by dots, the size and color of which indicates the presence and concentration of dissolved O_2. Axis 1 and 2 explained 8% and 7.7% of the variance in OTU distribution, respectively. A Monte Carlo test for significance of the Eigenvalues yielded a *P*-value of 0.03.

nated and reduced states (Zaikova *et al.*, 2010). In summer months, restriction of movement by a shallow 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 O_2 as this organic matter sinks and is remineralized. 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 OMZ formation on community structure of marine protists in Saanich Inlet over a 3-year period. Phylogenetic and multivariate statistical analyses of the resulting data set revealed defined shifts in OTUs that correlated 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 changing levels of water column oxygen-deficiency.

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 O_2 conditions, interpretation of site-specific differences must be made with extreme caution.

It is known that O_2 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 al.*, 2008; Lin *et al.*, 2008; Edgcomb *et al.*, 2011a; Orsi *et al.*, 2011b) and the Framvaren Fjord (Behnke *et al.*, 2006; Stoeck *et al.*, 2009, 2010). These findings are validated by our CCA, MRPP and Monte Carlo analyses that indicate O_2 and sulfide, as well as CH_4, to be the primary drivers of protistan distribution in Saanich Inlet (see Results and Figure 1). Ours is the first investigation into the influence of CH_4 on the distribution of protistan communities in low O_2 marine environments and our results suggest that CH_4 may have a stronger influence than sulfide, based on the differential length of the CH_4 and sulfide vectors in the CCA (Figure 1). Thus, the selective influence of CH_4 on protistan communities should be considered in future studies of OMZs. Interestingly, methanogenic archaea were not detected in Saanich Inlet water samples that incorporated a 2.7-µm prefilter (Zaikova *et al.*, 2010). While diffusive flux from underlying sediments is one likely source of CH_4 in the water column, new CH_4 production originating from methanogens associated with the anaerobic ciliates detected in our study may also contribute to CH_4 accumulation in basin waters. Methanogenic symbionts of ciliates are well known (Embley and Finlay, 1993; Fenchel and Finlay, 1995; van Hoek *et al.*, 2000; Edgcomb *et al.*, 2011c). This finding suggests that such symbioses may have the potential to contribute to climate active trace gas cycling in low O_2 and anoxic marine environments.

The CCA analysis also suggests that unique assemblages of protists inhabit different niches along the redoxcline in Saanich Inlet throughout the year depending on the intensity of renewal events. The November 200 m sample collected during renewal does not group with the other 200 m samples from April, February and July on the biplot (Figure 1).

Rather, this sample groups with dysoxic samples on the biplot, most likely reflecting the physical movement of an oxygenated water mass into basin waters (Table 1). This is supported by the minimal overlap in OTUs observed at 200 m between July and November and the MRPP analysis of the influence of 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; Supplementary Figure 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 (Lin *et al.*, 2006; Massana *et al.*, 2006; Guillou *et al.*, 2008). The majority of Stramenopile sequences were affiliated with the
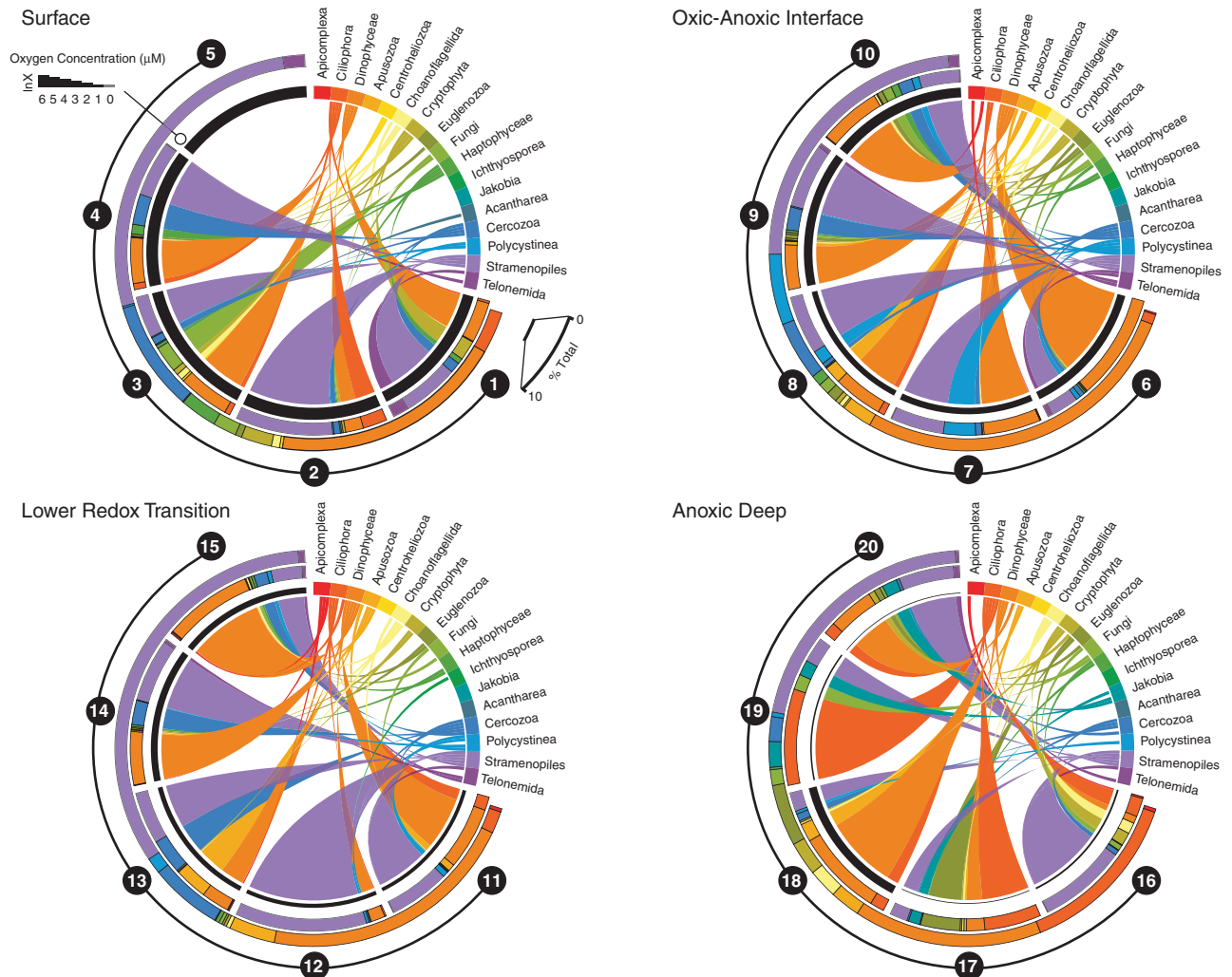
Figure 2 Phylum-level taxonomic affiliations of Saanich Inlet 18S rRNA gene sequences (see Table 1 for sample_id information). The color-coded outer histogram represents the abundance of 17 major taxonomic groups identified. The relative abundance (percentage of total) of sequences affiliated with each taxonomic group within a sample is indicated by the thickness of the colored area at the perimeter of the circle. Black circles represent individual samples. The concentration of O_2 within each sample is represented by a black bar, and gray bars indicate samples with no detectable O_2. The height of each bar is scaled according to the value of the O_2 concentration (in μM) normalized using natural logarithm (ln).

Figure 3 Combined histogram and heatmap describing the diversity of Ciliophora, Cercozoa, Fungi and Euglenozoa 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 ≥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.
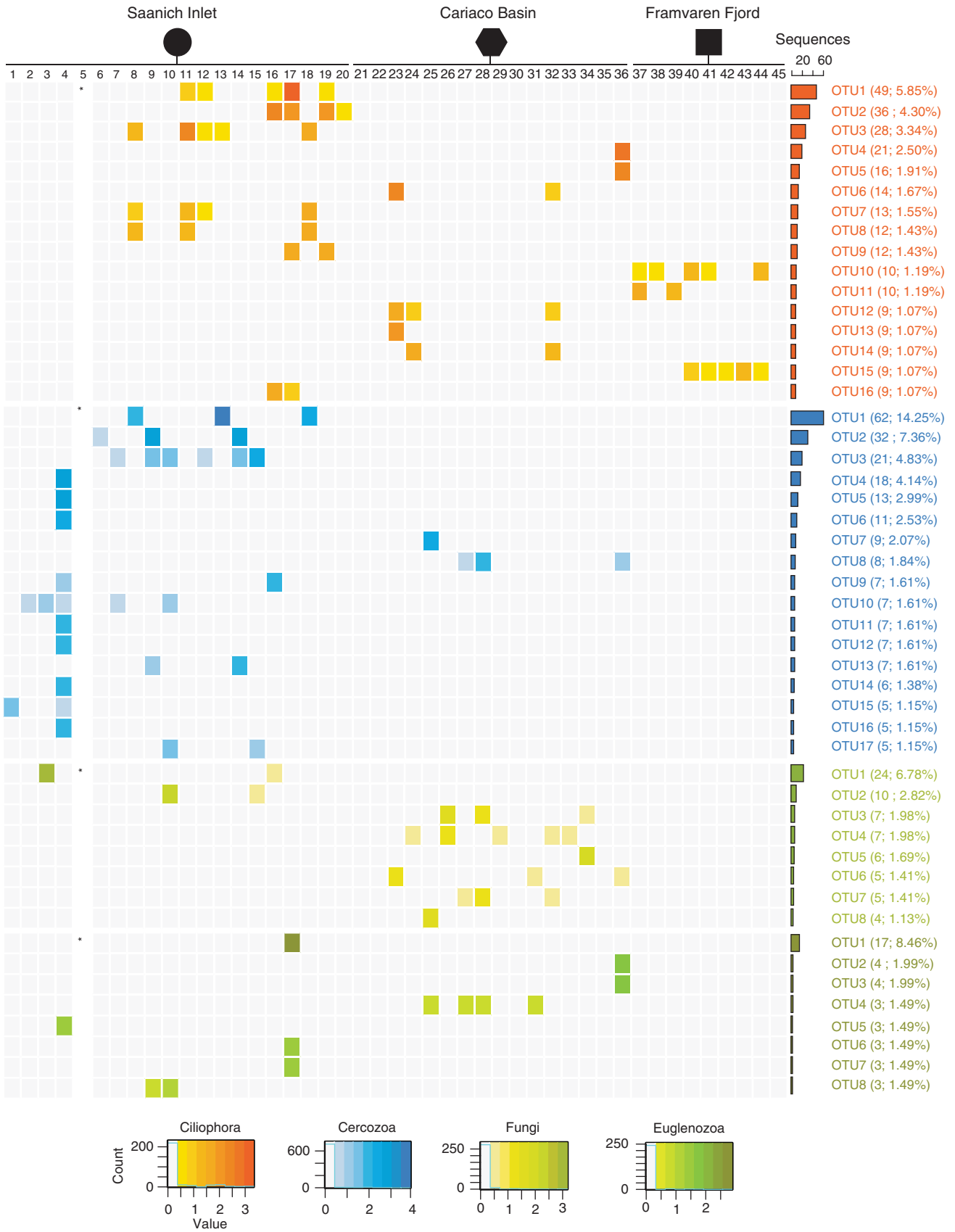
uncultured MAST (Marine Stramenopiles), which have been shown to exhibit a range of trophic modes and specificity for different prey species and sizes (Massana et al., 2009). Thus, as MAST stramenopiles have been detected in the anoxic waters of Framvaren Fjord and Cariaco Basin (Behnke et al., 2010; 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 (Figures 2 and 3). 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 OTUs relative to oxygenated 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 O_2 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 (that is, *Calkinsia*, *Cyclidium, Strombidium* and *Nyctotherus*) fall into this category (Supplementary Table S1).

Our phylogenetic analyses of ciliate and euglenozoan-affiliated OTUs recovered from anoxic waters (Figures 4 and 5), 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 5) with low (<90%) identities to the euglenozoan *Calkinsia aureus* (Yubuki et al., 2009) (Supplementary 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., 2011c) with a cortex that is completely covered by epibiotic bacteria belonging to the *Arcobacter*, a group that includes chemoautotrophs and chemoorganotrophs capable of NO_3^- reduction and sulfide oxidization (Edgcomb et al., 2011c). 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 >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 (Prescott, 1994), the increase in abundance of ciliate-affiliated sequences may be induced by O_2 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 N_2O that are known to exist in the Inlet at this depth (Zaikova et al., 2010). Because N_2O 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 N_2O from the surrounding water column.

Our data set 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 (Not et al., 2009). Fluctuations in sequence representation within OTUs affiliated with the Stramenopiles, Dinophyceae, Ciliophora and Euglenozoa (see Results and Figures 2 and 3; Supplementary Figure 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 N_2O during periods of anoxia would serve as a prime example of the potential ecological importance of such a 'rare biosphere.'

Comparisons of parametric and non-parametric richness estimates for the Saanich Inlet, Cariaco Basin and Framvaren Fjord data sets need to be interpreted with caution because of methodological differences in sample collection. Although potentially biased, this comparison suggests that the 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 than are estimated for the Saanich Inlet and roughly 10 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 (Supplementary 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
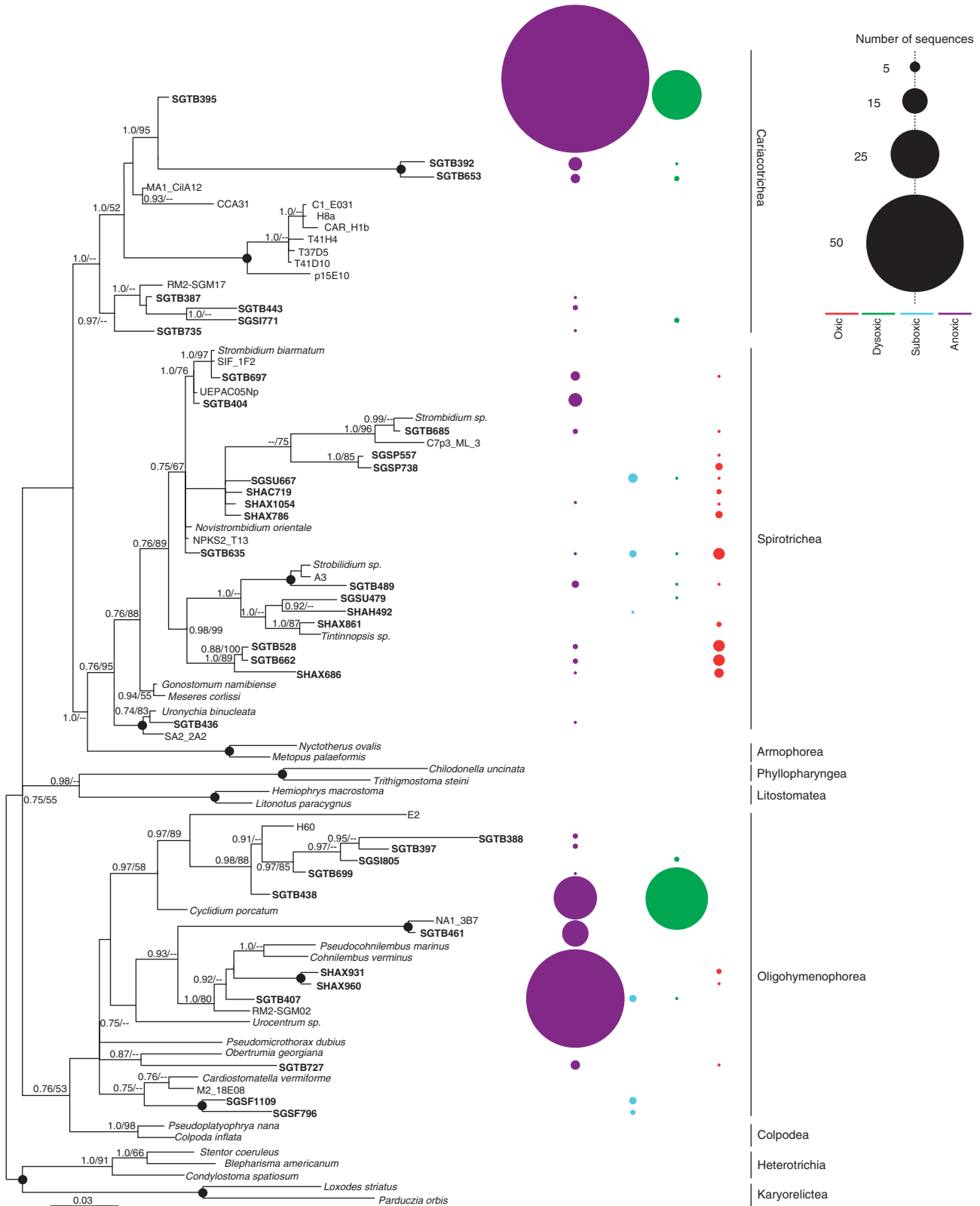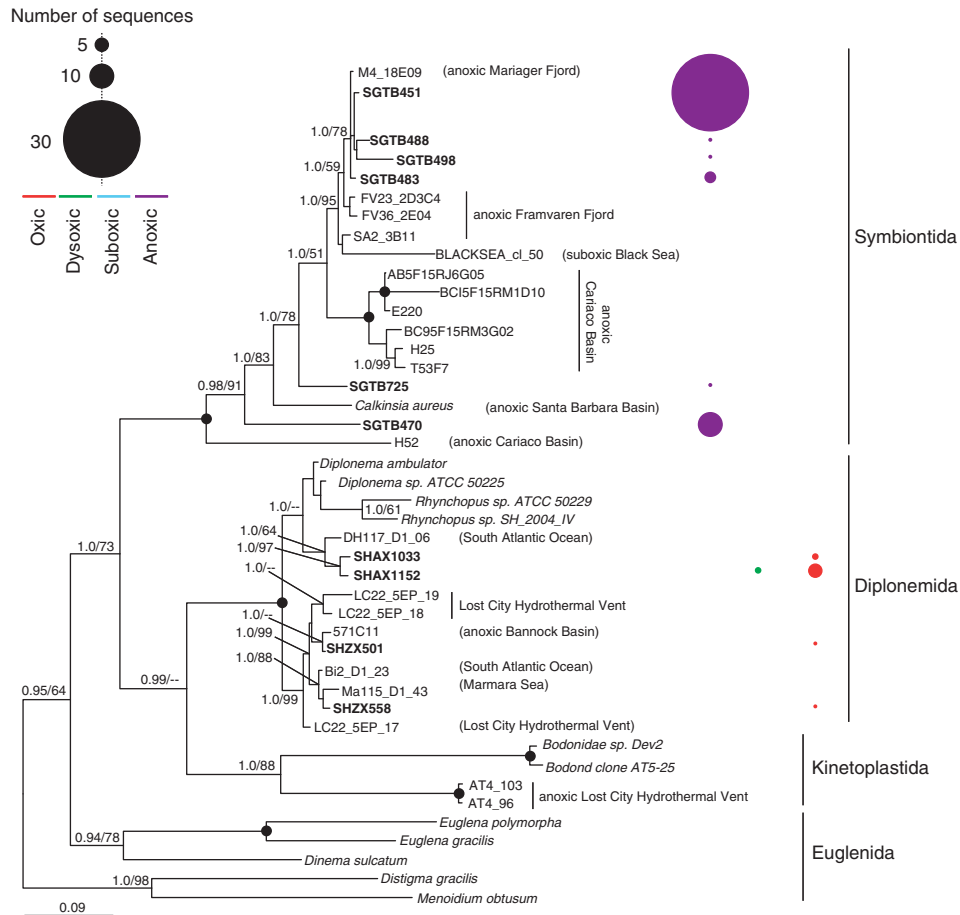
Figure 4 Phylogenetic relationships of ciliate-affiliated 18S rRNA gene sequences. The tree was constructed under maximum likelihood using an alignment of 757 unambiguous positions under the GTR+I+Gamma model of sequence evolution. Bootstrap (PhyML, 1000 iterations) and posterior probability (5 000 000 generations with 25% of trees discarded as burnin) values >50% are shown at the nodes in the order PP/ML (posterior probability/maximum likelihood bootstrap). 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 O_2 concentration, respectively.

Figure 5 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 >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 O_2 concentration, respectively.

Saanich Inlet water column (Supplementary Table S2). The same trend is apparent in the empirically registered OTU richness (Supplementary Table S2). Together with phylogenetic and multivariate analyses (Figures 1, 4 and 5), 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 6), 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 subbasins 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 (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
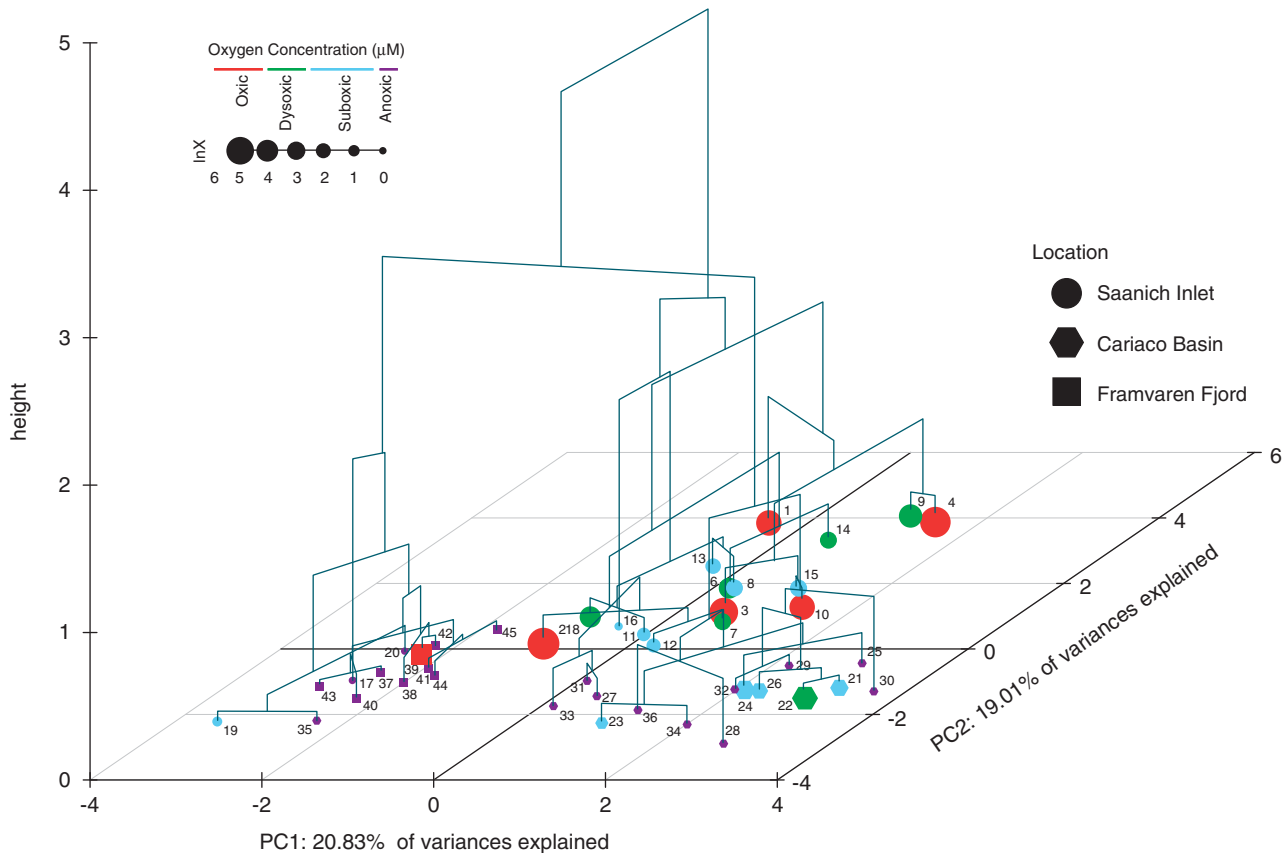
Figure 6 PCA and hierarchical clustering of the 38 samples from Saanich Inlet, Cariaco Basin and Framvaren Fjord (see Table 1 for sample_id information). The x and y axes of the grid represent the first and second principal components, respectively. Each dot represents one of the 38 samples used in the 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 O_2 concentration, oxic (red), dysoxic (green), suboxic (light blue) and anoxic (purple). The size of each dot is scaled according to the value of the O_2 concentration (in µM) normalized using natural logarithm (ln). The clustering pattern is further linked to the dendrogram generated from hierarchical clustering.

for millions of years (Schubert, 1982) and experiences only limited O_2 intrusion events (Lin *et al.*, 2008). This timeframe has likely allowed for expanded 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 S2), as well as the separation of the fjord communities from Cariaco on the PCA biplot (Figure 6).

Our analysis revealed common and unique responses of protists to water column O_2-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 within OMZs.

Acknowledgements

1600

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 data set and Nathan Cahoon at Northeastern University for his helpful assistance with sequence clustering.

Accession numbers

Eukaryotic 18S rRNA gene sequences recovered from the Saanich Inlet water column on were deposited in Genbank under accession numbers HQ864863–HQ871151.

References

Arrigo KR. (2005). Marine microorganisms and global nutrient cycles. *Nature* 437: 349–355.

Behnke A, Barger KJ, Bunge J, Stoeck T. (2010). Spatio-temporal variations in protistan communities along an O/HS gradient in the anoxic Framvaren Fjord (Norway). *FEMS Microbiol Ecol* 72: 89–102.

Behnke A, Bunge J, Barger K, Breiner HW, Alla V, Stoeck T. (2006). Microeukaryote community patterns along an O2/H2S gradient in a supersulfidic anoxic fjord (Framvaren, Norway). *Appl Environ Microbiol* 72: 3626–3636.

Bernhard JM, Buck KR, Farmer MA, Bowser SS. (2000). The Santa Barbara Basin is a symbiosis oasis. *Nature* 403: 77–80.

Bunge J. (2010). *Forthcoming in Proceedings of the 2011 Pacific Symposium on Biocomputing*. Pacific Symposia on Biocomputing: Hawaii.

Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, Delong EF et al. (2010). A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. *Science* 330: 1375–1378.

Caron DA, Countway PD, Savai P, Gast RJ, Schnetzer A, Moorthi SD et al. (2009). Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. *Appl Environ Microbiol* 75: 5797–5808.

Chambouvet A, Morin P, Marie D, Guillou L. (2008). Control of toxic marine dinoflagellate blooms by serial parasitic killers. *Science* 322: 1254–1257.

Chao A, Bunge J. (2002). Estimating the number of species in a stochastic abundance model. *Biometrics* 58: 531–539.

Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA et al. (2003). The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res* 31: 442–443.

Diaz RJ, Rosenberg R. (2008). Spreading dead zones and consequences for marine ecosystems. *Science* 321: 926–929.

Diaz RJ, Rosenberg R, Rabalais NN, Levin LA. (2009). Dead zone dilemma. *Mar Pollut Bull* 58: 1767–1768.

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.

Edgcomb V, Orsi W, Bunge J, Jeon S, Christen R, Leslin C et al. (2011a). Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME J* 5: 1344–1356.

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: 1237–1241.

Embley TM, Finlay BJ. (1993). Systematic and morphological diversity of endosymbiotic methanogens in anaerobic ciliates. *Antonie Van Leeuwenhoek* 64: 261–271.

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 USA* 105: 7774–7778.

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: 3349–3365.

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–W559.

Hong SH, Bunge J, Jeon SO, Epstein SS. (2006). Predicting microbial species richness. *Proc Natl Acad Sci USA* 103: 117–122.

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: 1639–1645.

Lam P, Kuypers MM. (2010). Microbial nitrogen cycling processes in oxygen minimum zones. *Ann Rev Mar Sci* 3: 317–345.

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: 581–584.

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: 5626–5630.

Lin XJ, Scranton MI, Chistoserdov A, Varela R, Taylor GT. (2008). Spatiotemporal dynamics of bacterial populations in the anoxic Cariaco Basin. *Limnol Oceanogr* 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: 1363–1371.

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: 1515–1522.

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: 588–596.

Naqvi SWA, Bange HW, Farỳas 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**: 1357–1373.

Pedros-Alio C. (2007). Ecology. Dipping into the rare biosphere. *Science* **315**: 192–193.

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**: 1572–1574.

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. *Mar Geol* **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, Ho T-Y, 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**: 1326–1327.

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**: 251–258.

Walsh DA, Hallam SJ. (2011). Bacterial community diversity and population dynamics in a seasonally anoxic fjord: Saanich Inlet. In Bruijn Fd (eds). *Handbook of Molecular Microbial Ecology II: Metagenomics and Complementary Approaches*. Wiley: Hoboken, NJ, 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**: 578–582.

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. *J Vis Exp*; e-pub ahead of print 17 June 2009; 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**: 172–191.

Zehnder AJ, Stumm W. (1988). Geochemistry and biogeochemistry of anaerobic habitats. In Zehnder AJ (eds). *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**: 476–491.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)