

1

2 **Metabarcoding analysis on European coastal samples**
3 **reveals new molecular metazoan diversity**

4

5 David López-Escardó¹, Jordi Paps², Colomán de Vargas^{3,4}, Ramon Massana⁵, Iñaki
6 Ruiz-Trillo^{1,6,7*}, Javier del Campo^{1,8*}

7 **Supplementary Materials and Methods**

8 *Sampling, 454 sequencing and curation of the sequences*

9 During the BioMarKs project (biomarks.eu), samples were collected in six European
10 coastal sites: the North Sea (Oslo, Norway), the English Channel (Roscoff, France),
11 the Bay of Biscay (Gijón, Spain), the Mediterranean Sea (Blanes, Spain, and Naples,
12 Italy) and the Black Sea (Varna, Bulgaria). Water column samples were taken with
13 Niskin bottles attached to a CTD rosette at surface and deep chlorophyll maximum
14 depths. Twenty liters of water per sample were pre-filtered through 20 µm filters and
15 then sequentially filtered through 3µm and 0.8µm polycarbonate filters (diameter:
16 142mm) using a peristaltic pump. Filtration time did not surpass 30min to avoid RNA
17 degradation. For dissolved DNA, 20 liters of 0.2 µm-filtered seawater was mixed with
18 400 ml of 0.5% CTAB (cetyltrimethylammonium bromide) (pH = 8) for 5 h and
19 filtered through 0.2 µm polycarbonate membranes (142 mm). To collect the micro-
20 (20– 200 µm) and meso- (200–2000 µm) planktonic fractions (micro/mesoplankton),
21 a plankton net of 20 µm mesh size was towed for 5–15 min, and the large protists

22 collected were rinsed with 0.2 µm-filtered seawater, passed through a 2000 µm
23 metallic sieve and filtered with 12µm PC membranes (47 mm). Filters were flash
24 frozen and stored at -80°C. Sediment samples were taken with sediment cores and
25 small aliquots were frozen at -80°C (Table S1). The total number of samples
26 considered in this study was 137 (Table S2). Total DNA and RNA were extracted at
27 the same time from the same filter using the NucleoSpin RNA L kit (Macherey-
28 Nagel, Düren, Germany). After quantification with a Nanodrop ND-1000
29 spectrophotometer (NanoDrop Technologies Inc, Wilmington, DE, USA), the quality
30 was further checked on a 1.5% agarose gel. Contaminating DNA was removed from
31 RNA samples using the TurboDNA kit (Ambion, Carlsbad, CA, USA). Extracted
32 RNA was immediately reverse transcribed using the RT Superscript III random
33 primers kit (Invitrogen, Carlsbad, CA, USA). The universal primers
34 TAReuk454FWD1 (50-CCAGCASCYGCGGTAATTCC-30) and TAReukREV3
35 (50-ACTTTCGTTCTTGATYRA-30) were used to amplify the V4 region (~380 bp)
36 of the eukaryotic 18S rDNA (1). Primers were adapted for 454 following the
37 manufacturer's specifications. They had the configuration A-adapter-tag (7 or 8bp)-
38 forward primer and B-adapter-reverse primer. PCRs were performed as explained in
39 Logares et al. 2012 (2), where amplifications were done in a volume of solution of
40 25ml and consisted on a 1x MasterMix Phusion High-Fidelity DNA Polymerase
41 (Finnzymes, Espoo, Finland), 0.35 mM of each primer and 3% DMSO. 5ng of
42 template DNA/cDNA was added to each PCR sample. PCRs cycles started with a
43 denaturation step at 98°C for 30s, followed by 10 cycles of 10 s at 98 °C, 30s at 53 °C
44 and 30s at 72°C, and afterwards by 15 cycles of 10 s at 98 °C, 30s at 48°C and 30s at
45 72 °C. Amplicons were evaluated in a 1.5% agarose gel to check for successful
46 amplifications. Triplicate amplicons were pooled and purified using the NucleoSpin

47 Extract II (Macherey-Nagel). Purified amplicons were eluted in 30 ml of elution
48 buffer and quantified again using a Nanodrop ND-1000 spectrophotometer. The final
49 number of pooled amplicons for 454 tag- sequencing was approximately 5mg.
50 Amplicon sequencing was carried out on a 454 GS FLX Titanium system (454 Life
51 Sciences, Branford, CT, USA) installed at Genoscope
52 (<http://ig.cea.fr/drf/ig/Pages/Genoscope.aspx>, France). Pyroreads were inspected to
53 remove short reads, reads with low quality and chimeras, as described in Massana et
54 al. 2015 (3) (Table S2).

55 *Taxonomic affiliation of the OTUs*

56 OTUs were taxonomically filtered using several eukaryotic reference datasets in order
57 to discard non-metazoan sequences (4, 5). Afterwards, we used our own metazoan
58 reference dataset manually curated by phylogeny to annotate the metazoan sequences
59 (available at figshare <https://dx.doi.org/10.6084/m9.figshare.3475007.v1>). With the
60 exception of the supergroups Arthropoda, Chordata and Annelida, in which the
61 database was also phylogenetically curated at the subphylum level, the phylogenetic
62 tree built to determine the taxonomic affiliation only confirmed assignment at the
63 phylum level. OTUs were primarily assigned to a group when they had an e-value
64 below 10^{-100} against a reference sequence. Additionally, OTUs with higher BLAST e-
65 values and represented by more than 100 reads from three or more samples were also
66 considered if they were phylogenetically placed within metazoans by performing
67 Maximum likelihood trees through RAxML 7.2.8 (6) . We ended with 372,934
68 representative metazoan reads and 1076 OTUs from 137 samples (Table S2), 103
69 distinct samples, 34 of them in duplicate (same RNA/DNA extraction but separate
70 PCR and sequencing reactions).

71 *Reproducibility of PCR and 454 sequencing*

72 In order to evaluate the reproducibility of the different replicates, we selected
73 duplicated samples from different templates (8 DNA, 9 RNA), different sampling
74 sites (2 Barcelona, 7 Naples and 8 Oslo), in which at least one of the duplicates
75 contain more than 100 metazoan reads (n=17). Each duplicate (same nucleic acid
76 template and separate PCR and 454 sequencing) was selected from our 34 duplicated
77 samples out of 137. We calculated the linear regressions by plotting OTU abundances
78 in each duplicate (3). The pyroread ratio between duplicates varied from 0.01 to 0.94.
79 We plotted the statistics of linear regression, R² coefficients and regression slopes,
80 over the pyroread ratio between duplicates (Fig. S8). R² coefficients (Fig. S8) were
81 high (0.91 on average) and independent of pyrotag ratios. Thus, we confirmed that the
82 read distribution among different OTUs were found at similar relative abundances in
83 each duplicate. Slopes regression and pyroread ratio present a linear growth (R² of
84 0.95; slope of 0.97), confirming that OTU abundances increased proportionally with
85 the number of pyrotags in the sample as described in Massana et al. 2015 (3).

86 *Diversity and distribution analysis*

87 The metazoan OTU table obtained was processed for community analysis using
88 QIIME (7). Beta-diversity analyses including PCA and Jackknife clustering were
89 performed with Unifrac (8). The OTU tree used as input for Unifrac was constructed
90 after aligning the OTUs with Mothur (9). A subset of aligned sequences from our
91 homemade database was used as a reference for Mothur input. Then, a maximum
92 likelihood tree was generated with RAxML 7.2.8 and using GTRCATI as the
93 evolutionary model. A hundred repeated runs on distinct starting trees were carried

94 out to select the tree with the best topology and 100 bootstrap replicates were
95 performed using the same evolutionary model.

96 *Phylogenetic and diversity analysis of BioMarks V4 sequence tags belonging to a*
97 *novel metazoan group*

98 In order to phylogenetically place the short reads assigned to the novel metazoan
99 group (MAME 1; MArine METazoan group 1), we performed a RAxML-EPA analysis
100 (10). First, we built a metazoan reference tree using the longest putative MAME 1
101 sequence (1878 bp) found by BLAST at NCBI nt nr database (KC582969), as a
102 unique MAME 1 representative. Metazoan 18S rRNA gene sequences were
103 downloaded from GenBank (Table S3) and aligned using a MAFFT 7 E-INS-i
104 strategy (11). The resulting alignment was checked by eye with Geneious 8.0.4 (12),
105 and the ambiguously aligned positions deleted, resulting in a total of 1472 nucleotide
106 positions. Bayesian inference analysis was conducted with MrBayes 3.2.6 (13) using
107 the GTR + Γ +invariant model of evolution running 6,000,000 generations. Maximum
108 likelihood trees were generated with RAxML 7.2.8, using GTRCATI as the
109 evolutionary model. To place the MAME 1 group within tunicates, an additional
110 alignment was constructed with all tunicate sequences available and a phylogenetic
111 tree was inferred using the same strategy. Tunicate sequences were mostly taken from
112 Tsagkogeorga et al. 2009 (14) who had an alignment of 110 sequences (95 from
113 tunicates) occupying 1746 nucleotide positions. All the alignments and trees are
114 available at figshare (<https://dx.doi.org/10.6084/m9.figshare.3475007.v1>).

115 Next, we searched for MAME 1 – like sequences in other metabarcoding studies. In
116 particular, we downloaded 487 marine environmental 18S amplicon datasets from
117 NCBI's SRA (March 2016) using fastq-dump from SRA-toolkit with -R option,

118 which selects the high quality reads (SRA Handbook). We performed a BLAST
119 search over the SRA dataset using KC582969 as a query and an e-value cut-off of e-
120 100, retrieving 3677 putative MAME 1 reads from 104 SRA runs. Before processing
121 them, we used PEAR (16) to merge all the Illumina pair-end reads retrieved. Next, we
122 used usearch v8.1.861 for quality filtering, dereplication, clustering (97%) and
123 chimera checking using SILVA SSU 119 (5) as a reference. We ended up with 14
124 putative MAME 1 OTUs representing 3597 reads. We also performed a BLAST
125 search (cut-off e value of e-10) against the Tara Oceans database (17), retrieving 58
126 putative MAME 1 OTUs representing 123,779 reads.

127 We aligned all the MAME 1 – like short-read OTUs retrieved in the previous step and
128 the ones from BioMarks with the representative sequences used for the metazoan and
129 tunicate reference trees using the MAFFT strategy described earlier. After discarding
130 sequences that did not align properly, we ended up with 69 MAME 1 OTUs (3 from
131 BioMarKs, 14 from SRA and 52 from TARA), as well as the NCBI sequences
132 KC582969 and HQ869055. Ambiguous positions were removed from the alignment
133 checked by eye with Geneious 8.0.4 (12). The metazoan alignment yielded 1514
134 nucleotide positions, while the tunicate-specific alignment generated 1707 positions.
135 Finally, we used RAxML-EPA to place the short reads in both the metazoan and the
136 tunicate-specific datasets.

137 After the OTU assignments, we built an OTU table with the 69 MAME 1 group
138 OTUs. We used QIIME to analyze their read abundance and distribution across
139 different geographical locations, depths and size fractions. The OTU table and all the
140 alignments and trees are available as supplementary information at figshare
141 (<https://dx.doi.org/10.6084/m9.figshare.3475007.v1>).

142 **Supplementary References**

- 143 1. Stoeck T, et al. (2010) Multiple marker parallel tag environmental DNA
144 sequencing reveals a highly complex eukaryotic community in marine anoxic
145 water. *Mol Ecol* 19(Suppl. 1):21–31.
- 146 2. Logares R, et al. (2012) Diversity patterns and activity of uncultured marine
147 heterotrophic flagellates unveiled with pyrosequencing. *ISME J* 6(10):1823–
148 1833.
- 149 3. Massana R, et al. (2015) Marine protist diversity in European coastal waters
150 and sediments as revealed by high-throughput sequencing. *Environ Microbiol*
151 17(10):4035–4049.
- 152 4. Guillou L, et al. (2013) The Protist Ribosomal Reference database (PR2): A
153 catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated
154 taxonomy. *Nucleic Acids Res* 41(D1):597–604.
- 155 5. Quast C, et al. (2013) The SILVA ribosomal RNA gene database project:
156 improved data processing and web-based tools. *Nucleic Acids Res*
157 41(D1):D590–D596.
- 158 6. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based
159 phylogenetic analyses with thousands of taxa and mixed models.
160 *Bioinformatics* 22(21):2688–2690.
- 161 7. Caporaso JG, et al. (2010) QIIME allows analysis of high-throughput
162 community sequencing data. *Nat Methods* 7(5):335–336.
- 163 8. Hamady M, Lozupone CA, Knight R (2010) Fast UniFrac: facilitating high-

- 164 throughput phylogenetic analyses of microbial communities including analysis
165 of pyrosequencing and PhyloChip data. *ISME J* 4(1):17–27.
- 166 9. Schloss PD, et al. (2009) Introducing mothur: Open-source, platform-
167 independent, community-supported software for describing and comparing
168 microbial communities. *Appl Environ Microbiol* 75(23):7537–7541.
- 169 10. Berger SA, Krompass D, Stamatakis A (2011) Performance, accuracy, and
170 Web server for evolutionary placement of short sequence reads under
171 maximum likelihood. *Syst Biol* 60(3):291–302.
- 172 11. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for
173 rapid multiple sequence alignment based on fast Fourier transform. *Nucleic
174 Acids Res* 30(14):3059–3066.
- 175 12. Kearse M, et al. (2012) Geneious Basic: An integrated and extendable desktop
176 software platform for the organization and analysis of sequence data.
177 *Bioinformatics* 28(12):1647–1649.
- 178 13. Ronquist F, et al. (2012) Mrbayes 3.2: Efficient bayesian phylogenetic
179 inference and model choice across a large model space. *Syst Biol* 61(3):539–
180 542.
- 181 14. Tsagkogeorga G, et al. (2009) An updated 18S rRNA phylogeny of tunicates
182 based on mixture and secondary structure models. *BMC Evol Biol* 16(9:187):1–
183 16.
- 184 15. SRA Handbook Available at: <https://www.ncbi.nlm.nih.gov/books/NBK47528/>
185 [Accessed March 1, 2016].

- 186 16. Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: A fast and accurate
187 Illumina Paired-End reAd mergeR. *Bioinformatics* 30(5):614–620.
- 188 17. de Vargas C, et al. (2015) Eukaryotic plankton diversity in the sunlit ocean.
189 *Science* 348(6237):1261605–1261605.
- 190 18. Paps J, Baguña J, Riutort M (2009) Lophotrochozoa internal phylogeny: new
191 insights from an up-to-date analysis of nuclear ribosomal genes. *Proc R Soc B*
192 *Biol Sci* 276(1660):1245–1254.
- 193 19. Paps J, Baguña J, Riutort M (2009) Bilaterian phylogeny: A broad sampling of
194 13 nuclear genes provides a new lophotrochozoa phylogeny and supports a
195 paraphyletic basal acoelomorpha. *Mol Biol Evol* 26(10):2397–2406.
- 196 20. Zhang Z-QQ (2013) Animal biodiversity: An update of classification and
197 diversity in 2013. *Zootaxa* 3703(1):5.

198 **Supplementary Figure Legends and Tables**

199 **Fig. S1: BioMarKs sampling sites.** Map indicating the sampling locations where the
200 data were collected and the summary of the dataset characteristics. Map retrieved
201 from Wikimedia Commons
202 (https://commons.wikimedia.org/wiki/File:Blank_map_Europe_without_borders.png)
203 CC-BY-SA-4.0,3.0,2.5,2.0,1.0 (<https://creativecommons.org/licenses/by-sa/4.0/>).

204 **Fig. S2: Box plot distribution of relative metazoan abundance compared with all**
205 **eukaryotes.** Relative abundance of metazoans compared to all eukaryotes in (a)
206 different oxic pelagic fractions, (b) different sites and in (c) different depths. Note that
207 data is provided from just one sample in the anoxic sediments.

208 **Fig. S3: Jackknife clustering analysis of phylogenetic composition of the samples.**
209 The chart represents the relative abundance within metazoan phyla in each sample.
210 Samples from extracellular DNA and the ones with less than 100 reads (DNA+RNA)
211 were removed from the analysis.

212 **Fig. S4: Principal component analysis of the samples.** Samples from extracellular
213 DNA and the ones with less than 100 reads (DNA+RNA) were removed from this
214 analysis. Analyses are shown for (a) size fraction, (b) depth and (c) site.

215 **Fig. S5: MAME 1 phylogenetic position and ecological distribution.** (a) Metazoan
216 18S rRNA phylogenetic tree placing the novel metazoan group MAME 1. The tree
217 was inferred using RaxML-EPA from the 18S rRNA gene sequence including
218 representatives from all metazoan groups. Metazoan super-group nomenclature is
219 based on Paps et al. 2009a and b (18, 19). The nodal support values marked with a
220 dot correspond to maximum likelihood 100-replicate bootstrap support and Bayesian
221 posterior probabilities. (b) Worldwide distribution of MAME 1 group. World map
222 within BioMarks data or within public repositories. Dot size represents the number of
223 reads found shown on a log₂ scale. Bar charts show the distribution of MAME 1 reads
224 by depth and fraction.

225 **Fig. S6: Read distribution of shared OTUs among water column and benthic**
226 **environments.** Each dot represents an OTU. Axes indicate whether reads belong to
227 the water column or the sediment. Colors indicate the taxonomy of the OTU.

228 **Fig. S7: Comparison of number of OTUs found, number of described metazoan**
229 **species and number of 18S rRNA metazoan sequences in NCBI.** (a) Total number
230 of OTUs from 18S rRNA retrieved in our dataset (blue bars) compared to the number
231 of described species for each metazoan phylum (20) (red bars). (b) Total number of
232 OTUs from 18S rRNA retrieved from Genbank (blue bars) compared to the number of
233 described species for each metazoan phylum (20) (red bars). The number of 18S
234 rRNA sequences from NCBI was obtained from the following search for each
235 phylum: "*txid33208[Organism:exp] (18S OR SSU) NOT (mitochondrial OR*
236 *mitochondria)*". (B) MAME 1 distribution.

237 **Fig. S8: Linear regression statistics of the read distribution between the**
238 **duplicated samples analysed (n=17).** The figure shows on the Y axis R² coefficients
239 (blue dots) and slope values (read dots) plotted over the pyroread ratio between
240 duplicates (X axis).

241 **Table S1.** Description of the sampling sites.

242

Site	Coordinates	Distance to coast (Km)	Max. Depth (m) ¹	Sampling date	DCM (m)	Temperature Surface (°C)	Temperature DCM (°C)	Temperature Sediment (°C)	Salinity Surface (PSU)	Salinity Sediment (PSU)	Chla (µg/l) ²	[NO ₃] ⁻ Surface/DCM (µg/l)	[PO ₄ ³⁻] Surface/DCM (µg/l)	Total Phosphorus Surface/DCM (µg/l)
Blanes	41° 40' N 2° 48' E	1.0	20	2/2010	N/A	12.5	N/A	12.6	37.5	38.2	1.0	2 / N/A	7 / N/A	13 / N/A
Gijon	43° 40' N 5° 35' W	12.0	110	9/2010	40	20.2	14.0	12.0	35.7	36.6	7.0	2 / 26	3 / 4	10 / 12
Naples 2009	40° 48' N 14° 15' E	4.0	75	10/2010	23	22.8	22.4	14.6	37.7	37.9	1.4	16 / 0	1 / 1	22 / 16
Naples 2010	40° 48' N 14° 15' E	4.0	76	5/2010	35	19.2	15.5	14.0	37.2	37.9	1.2	<2 / <2	4 / 3	14 / 8
Oslo 2009	59° 16' N 10° 43' E	1.5	100	09/2010	8	15.0	15.0	8.0	25.0	35.0	3.2	9 / 1	4 / 3	22 / 21
Oslo 2010	59° 16' N 10° 43' E	1.5	100	06/2010	9	15.0	12.5	6.0	22.0	35.0	1.9	<2 / <2	3 / 2	12 / 11
Roscoff	48° 46' N 3° 57' W	5.0	60	4/2010	N/D	9.9	N/D	9.9	34.9	34.9	0.5	87 / N/D	12 / N/D	29 / (N/D)
Varna	43° 10' N 28° 50' E	40.0	400	5/2010	40	21.5	9.5	8.5	16.0	22.0	8.0	2 / 2	4 / 3	11 / 11

243 Surface is considered as < 5 m depth. N/A= not applicable, N/D= no data.

244 ¹ Maximum depth of the water column.

245 ² Maximum Chlorophyll a concentration in the water column measured with fluorometry (fluorometer attached to a CTD).

Table S26 Summary of samples including the total number of eukaryotic reads after quality control and the total number of metazoan reads after all the filtering process and the OTU assignment.

248

Description	All eukaryotes				Metazoans			
	Samples	Total reads	DNA reads	RNA reads	Samples	Total reads	DNA reads	RNA reads
Site								
Blanes	11	94366	48813	45553	11	59626	35719	23907
Gijon	4	50178	29116	21062	4	233	199	34
Naples	46	600756	266549	334207	44	131168	70170	60998
Oslo	44	406563	224876	181687	44	93635	60637	32998
Roscoff	9	55567	41861	13706	9	17127	14731	2396
Varna	25	250977	122751	132006	25	71145	39550	31595
Size Fraction (μm)								
0-0.2 (eDNA)	8	44564	44564	n/a	8	7879	7879	n/a
0.8-3	38	439410	208067	231343	38	37575	22981	14594
3-20	36	394879	194915	199964	36	23570	15949	7621
20-2000	33	379910	187405	192505	31	222149	127190	94959
Depth								
Subsurface	63	624138	336875	287263	23	81761	46996	34765
DCM	45	536884	262120	274764	62	134470	85545	48925
Anoxic	7	101004	37328	63676	45	105932	63027	42905
Sediments	24	216013	111294	104409	7	50771	25438	25333
Template								
DNA	74	746245	n/a	n/a	72	220766	n/a	n/a
RNA	65	728221	n/a	n/a	65	152168	n/a	n/a
Total	n/a	1474466	746245	728221	n/a	3792934	220766	152168

249
250
251
252

Table S3. Summary of the sequences from GenBank used to place new MAME 1 group within metazoans in Figure 5. The summary includes the accession number of the 18S rRNA gene sequence, the specie that it belongs and its taxonomy.

Taxonomy	Specie	Accession Number
Porifera	<i>Leucosolenia sp.</i>	AJ622898
	<i>Oscarella carmela</i>	EU702422
	<i>Aphrocallistes beatrix</i>	FM946127
	<i>Petrosia sp.</i>	DQ927321
Ctenophora	<i>Mertensia ovum</i>	AF293679
	<i>Pleurobrachia bachei</i>	AF293677
	<i>Mnemiopsis leidyi</i>	AF293700
Placozoa	<i>Trichoplax sp.</i>	Z22783
Cnidaria	<i>Aurelia aurita</i>	AY039208
	<i>Coryne pusilla</i>	Z86107
	<i>Alcyonium gracillimum</i>	Z92902
	<i>Parazoanthus axinellae</i>	U42453
	<i>Nematostella vectensis</i>	AF254382
	<i>Actinia equina</i>	AJ133552
	<i>Anemonia sulcata</i>	X53498
Acoelomorpha	<i>Paratomella rubra</i>	AF102892
	<i>Nemertoderma westbladi</i>	AF327726
Xenoturbellida	<i>Xenoturbella bocki</i>	AY291292
Hemichordata	<i>Saccoglossus pusillus</i>	AF236800
	<i>Balanoglossus carnosus</i>	D14359
	<i>Glossobalanus minutus</i>	AF119089
	<i>Ptychodera flava</i>	AF278681
Echinodermata	<i>Endoxocrinus parrae</i>	Z80951
	<i>Antedon serrata</i>	D14357
	<i>Strongylocentrotus purpuratus</i>	L28055
	<i>Asthenosoma owstoni</i>	Z37118
	<i>Aquilonastra coronata</i>	AB084566
	<i>Astropecten latespinosus</i>	AB084546
Cephalochordata	<i>Branchiostoma floridae</i>	M97571
Craniata	<i>Urobatis jamaicensis</i>	AY049861
	<i>Mitsukurina owstoni</i>	AY049840
Tunicata	<i>Oikopleura sp.</i>	AB013015
	<i>Oikopleura labradoriensis</i>	FM244869
	<i>Ascidia sydneiensis</i>	AF165819
	<i>Ciona intestinalis</i>	AB013017
	<i>Clavelina meridionalis</i>	FM244840
	<i>Pyrostremma spinosum</i>	HQ015379
	<i>Thalia sp.</i>	AB859895
	<i>Salpidae sp.</i>	HQ015377
	<i>Halocynthia igaboja</i>	AY903925
	<i>Cnemidocarpa humilis</i>	FM244859

	<i>Molgula occidentalis</i>	FM244850
	<i>Molgula provisionalis</i>	L12434.2
MAME 1	<i>Uncultured Eukaryote</i>	KC582969
	<i>Uncultured Eukaryote</i>	HQ869055
Kinorhyncha	<i>Pycnophyes kielensis</i>	U67997
Priapulida	<i>Halicryptus spinulosus</i>	AF342790
	<i>Priapulus caudatus</i>	Z38009
Nematomorpha	<i>Gordius albopunctatus</i>	U88337
	<i>Neochordodes occidentalis</i>	AF421768
	<i>Paragordius tricuspidatus</i>	AF421771
Chaetognatha	<i>Eukrohnia bathypelagica</i>	DQ351896
	<i>Parasagitta setosa</i>	DQ351900
	<i>Sagitta bipunctata</i>	DQ351890
Nematoda	<i>Enoplus brevis</i>	U88336
	<i>Desmodora ovigera</i>	Y16913
	<i>Catanema sp.</i>	Y16912
Tardigrada	<i>Milnesium tardigradum</i>	U49909
Arthropoda	<i>Colossendeis sp.</i>	AF005440
	<i>Pandinus imperator</i>	AY210831
	<i>Limulus polyphemus</i>	L81949
	<i>Lithobius obscurus</i>	AF334271
	<i>Orthoporus sp.</i>	AY210829
	<i>Heterothrips arisaemae</i>	KC512970
	<i>Ctenolepisma longicaudata</i>	AY210811
	<i>Triops longicaudatus</i>	AF144219
	<i>Orchesellides sinensis</i>	KC236251
	<i>Squilla empusa</i>	L81946
Gnathostomulida	<i>Haplognathia simplex</i>	DQ079931
Bryozoa	<i>Fron dipora verrucosa</i>	FJ409612
Gastrotricha	<i>Paraturbanella teissieri</i>	JF357661
Entoprocta	<i>Barentsia benedeni</i>	U36272
Cycliophora	<i>Symbion pandora</i>	AY218106
Micrognathozoa	<i>Limnognathia maerski</i>	AJ487046
Rotifera	<i>Philodina roseola</i>	AF154567
	<i>Brachionus plicatilis</i>	U49911
	<i>Lecane bulla</i>	AF154566
	<i>Asplanchna sieboldi</i>	AF092434
Platyhelminthes	<i>Catenula sp.</i>	AJ012532
	<i>Stenostomum leucops</i>	D85095
	<i>Macrostomum hystricinum</i>	AF051329
	<i>Haplopharynx rostratus</i>	AJ012511
	<i>Pseudoceros tritriatus</i>	AJ228794
	<i>Thysanozoon brocchii</i>	D85096
	<i>Planocera multitentaculata</i>	D83383

	<i>Discocelis tigrina</i>	U70078
	<i>Notoplana australis</i>	D85097
Nemertea	<i>Amphiporus ochraceus</i>	AY039668
	<i>Cerebratulus lacteus</i>	AY145368
	<i>Lineus ruber</i>	AY039672
Mollusca	<i>Rhabdus rectius</i>	AF120523
	<i>Lima lima</i>	AF120533
	<i>Pteria hirundo</i>	AF120532
	<i>Nuculana minuta</i>	DQ279938
	<i>Yoldia myalis</i>	AF207643
Brachiopoda	<i>Glottidia palmeri</i>	U12647
	<i>Lingula anatina</i>	X81631
	<i>Phoronis australis</i>	U36271
Annelida	<i>Urechis caupo</i>	AF342805
	<i>Aspidosiphon misakie</i>	AF119090
	<i>Phascolosoma granula</i>	X79874
	<i>Dero digitata</i>	DQ459984
	<i>Acanthobdella peledin</i>	AY040680
	<i>Glossiphonia complana</i>	AF099943
	<i>Erpobdella octoculata</i>	AF099949

253
254
255

Table S4. Summary of reads and OTUs obtained within all metazoan phyla in BioMarks data.

Metazoan Phyla	OTUs	Reads	Metazoan Phyla	OTUs	Reads
Porifera	13	262	Priapulida	2	117
Ctenophora	10	23,371	Kinorhyncha	3	2,291
Cnidaria	55	18,627	Nematoda	247	8,932
Acoelomorpha	31	3,069	Tardigrada	3	8
Gastrotricha	23	1,540	Arthropoda		
Gnathostomulida	1	33	Myriapoda	1	190
Rotifera	12	3,576	Hexapoda	3	12
Bryozoa	12	1,055	Crustacea	370	190,872
Phoronida	1	306	Chelicerata	7	575
Platyhelminthes	33	1,540	Chaetognatha	21	14,739
Nemertea	7	622	Xenoturbellida	1	15
Mollusca	53	9,907	Echinodermata	16	3,206
Annelida			Hemichordata	2	123
Polychaeta	97	34,693	Craniata	14	1,961
Clitellata	1	5	Tunicata	30	37,982
<i>Sipuncula</i>	3	681	MAME 1	3	1,860

256
257
258

Table S5. Summary of the 20 most abundant metazoan OTUs in the water column taking into account the number of RNA reads.

Metazoan group	OTU	RNA reads	% ^a within metaz.	% ^b within group	CIM ^c	CIM BLAST % ID ^d	CIM taxonomy
Crustacea	6	10746	9.8%	22.0%	HM997070	100%	<i>Paracalanus parvus</i> (Calanoida, Copepod)
	2	7944	7.3%	16.3%	JX995318	100%	<i>Calanus helgolandicus</i> (Calanoida, Copepod)
	21	4834	4.4%	9.9%	JX995321	100%	<i>Pseudocalanus elongatus</i> (Calanoida, Copepod)
	11	3545	3.3%	7.3%	JX995298	100%	<i>Centropages typicus</i> (Calanoida, Copepod)
	15	3093	2.8%	6.3%	HM997062	100%	<i>Temora discaudata</i> (Calanoida, Copepod)
	84	2796	2.6%	5.7%	HM997079	98%	<i>Clausocalanus arcuicornis</i> (Calanoida, Copepod)
	224	2679	2.5%	5.5%	HM997079	99%	<i>Clausocalanus arcuicornis</i> (Calanoida, Copepod)
	22	2219	2.0%	4.5%	JF781540	98%	<i>Oithona sp.</i> (Cyclopoida, Copepod)
	8	1865	1.7%	3.8%	GU969179	100%	<i>Oithona Similis</i> (Cyclopoida, Copepod)
Total		39721	36.4%	81.2%			
Tunicata	4	21436	19.6%	56.6%	AB013014	100%	<i>Oikopleura dioica</i> (Appendicularian)
	39	1723	1.6%	4.6%	AY116613	100%	<i>Oikopleura dioica</i> (Appendicularian)
	169	1082	1.0%	2.9%	AB013012	100%	<i>Doliolum nationalis</i> (Doliolid)
Total		24241	22.2%	64.0%			
Ctenophora	7	12031	11.0%	83.5%	AF100944	100%	<i>Pleurobrachia pileus</i> (Typhlocoela)
	52	2108	1.9%	14.6%	AF293700	100%	<i>Lamproteus cruentiventer</i> (Cyclocoela)
Total		22387	12.9%	98.2%			
Chaetognatha	49	3736	3.4%	42.3%	DQ351879	99%	<i>Krohnitta pacifica</i> (Saggitioidea)
	80	3285	3.0%	37.2%	DQ351877	99%	<i>Flaccisagitta enflata</i> (Saggitioidea)
Total		7021	6.4%	79.5%			
Cnidaria	79	2114	1.9%	46.3%	AY039208	100%	<i>Aurelia aurita</i> (Scyphozoa)
	50	1887	1.7%	41.3%	DQ080014	100%	<i>Lilyopsis rosea</i> (Hydrozoa)
Total		4001	3.7%	87.6%			
Rotifera	116	2573	2.4%	86.9%	DQ297711	99%	<i>Notommata cordonella</i> (Ploimida)
MAME 1	102	1093	1.0%	65.1%	HQ869055*	95%	<i>Uncultured eukaryote</i> *

259
260
261
262
263
264

a. Percentage of RNA reads within the metazoans in the water column.

b. Percentage of RNA reads within the correspondent in the water column.

c. Accession number of the Closest Identified Match.

d. BLAST identity.

* The Closest Identified Match match has a very low identity (less than 90%). It is indicated instead, the Closest Environmental Match (CEM).

265 **Table S6.** Summary of the 20 most abundant metazoan OTUs in the sediments taking into account
 266 the number of RNA reads.
 267

Metazoan group	OTU	RNA reads	% ^a within metaz.	% ^b within group	CIM ^c	CIM BLAST % ID ^d	CIM taxonomy
Polychaeta	61	1823	5.2%	22.7%	EU340097	99%	<i>Aurospio foodbancsia</i> (Spionida)
	89	1742	5.0%	21.7%	JF903633	100%	<i>Prosphaerosyllis magnoculata</i> (Aciculata)
	88	1423	4.1%	17.7%	JN936464	92%	<i>Paralysippe annectens</i> (Scolecida)
	26	643	1.9%	8.0%	AF412798	99%	<i>Parougia sp.</i> (Aciculata)
	36	495	1.4%	6.2%	JN936464	92%	<i>Paralysippe annectens</i> (Scolecida)
	69	389	1.1%	4.8%	AY838852	100%	<i>Ninoe nigripes</i> (Aciculata)
	229	388	1.1%	4.8%	AY611455	100%	<i>Polydora giardi</i> (Scolecida)
Total		6903	19.8%	85.8%			
Crustacea	68	2900	8.3%	36.7%	AB076635	98%	<i>Limnocythere sp.</i> (Ostracoda)
	211	802	2.3%	10.2%	AB076621	89%	<i>Kotoracythere inconspicua</i> (Ostracoda)
	273	458	1.3%	5.8%	EU380309	97%	<i>Itunella muelleri</i> (Copepoda)
	216	458	1.3%	5.8%	AY627016	100%	<i>Brayda sp.</i> (Copepoda)
	76	399	1.2%	5.1%	AB076631	100%	<i>Leptocythere lacertosa</i> (Ostracoda)
	260	326	0.9%	4.1%	EU380293	97%	<i>Paramphiascella fulvofasciata</i> (Copepoda)
Total		5343	15.3%	67.6%			
Mollusca	18	5389	15.5%	82.9%	DQ279940	100%	<i>Abra nitida</i> (Bivalvia)
	67	847	2.4%	13.0%	EF489348	100%	<i>Scaphander lignarius</i> (Gastropoda)
Total		6236	17.9%	95.9%			
Platyhelminthes	74	1427	4.1%	48.3%	FJ715296	99%	<i>Microstomum papillosum</i> (Rhabditophora)
Echinodermata	51	1159	3.3%	92.8%	AJ011142	99%	<i>Amphiura chiajei</i> (Ophiurida)
Nematoda	176	624	1.8%	12.5%	AJ966473	94%	<i>Anaplectus sp.</i> (Chromadorea)
Sipuncula	340	500	1.4%	91.2%	AF519248	100%	<i>Phascolion strombus</i> (Golfingiida)
Cnidaria	50	337	1.0%	67.4%	DQ080014	100%	<i>Lilyopsis rosea</i> (Hydrozoa)

268
 269
 270
 271

- a. Percentage of RNA reads within the metazoans in the sediments.
 b. Percentage of RNA reads within the correspondent in the sediments.
 c. Accession number of the Closest Identified Match.
 d. BLAST identity.

272 **Table S7.** Summary of the OTUs, whose RNA reads from small fractions are suspected
 273 to come from gametes.
 274

Distribution	Metazoan group	OTU	Reads ^c within small fraction	% ^d within dataset	CIM	CIM BLAST % ID.	CIM taxonomy
Small and large fractions^a	Ctenophora	7	11201	50.3%	AF100944	100%	<i>Pleurobrachia pileus</i> (Typhlocoela)
	Cnidaria	79	1589	7.1%	AY039208	100%	<i>Aurelia aurita</i> (Scyphozoa)
	Total		12790	57.5%			
Small exclusive^b	Polychaeta	88	340	1.5%	JN936464	92%	<i>Paralysippe annectens</i> (Scolecida)
	Ctenophora	52	212	1.0%	KJ754158	100%	<i>Mnemiopsis leidy</i> (Cyclocoela)
	Ctenophora	5798	104	0.5%	KJ754154	98%	<i>Pleurobrachia brunnea</i> (Typhlocoela)
	Ctenophora	715	77	0.4%	HG931678	96%	<i>Beroe ovata</i> (Cyclocoela)
	Polychaeta	19	70	0.3%	AY611452	88%	<i>Hydroides novegica</i> (Palpata)
	Total		803	3.6%			
TOTAL^e			13593	3.2%^e			

- 275 a. OTUs that are present in pico/nano and micro/meso fractions.
 276 b. The 5 most abundant OTUs exclusive from the pico/nano fractions.
 277 c. Number of RNA reads within the pico/nano fractions.
 278 d. Percentage of RNA reads within the pico/nano fractions for metazoans.
 279 e. Percentage of RNA reads within the pico/nano fractions for all eukaryotes.
 280

Dataset specs

Templates

DNA

RNA

Sample fractions

Picoplankton (0.8–3 μm)

Nanoplankton (3–20 μm)

Micro/mesoplankton (20–2,000 μm)

Sediments (total)

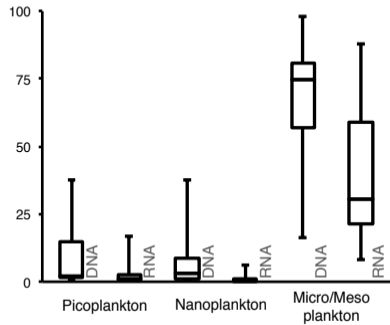
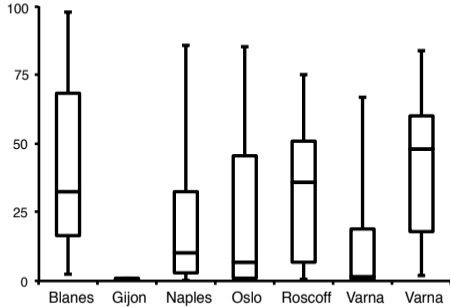
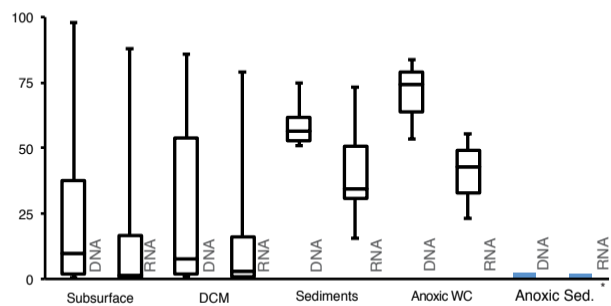
Depth

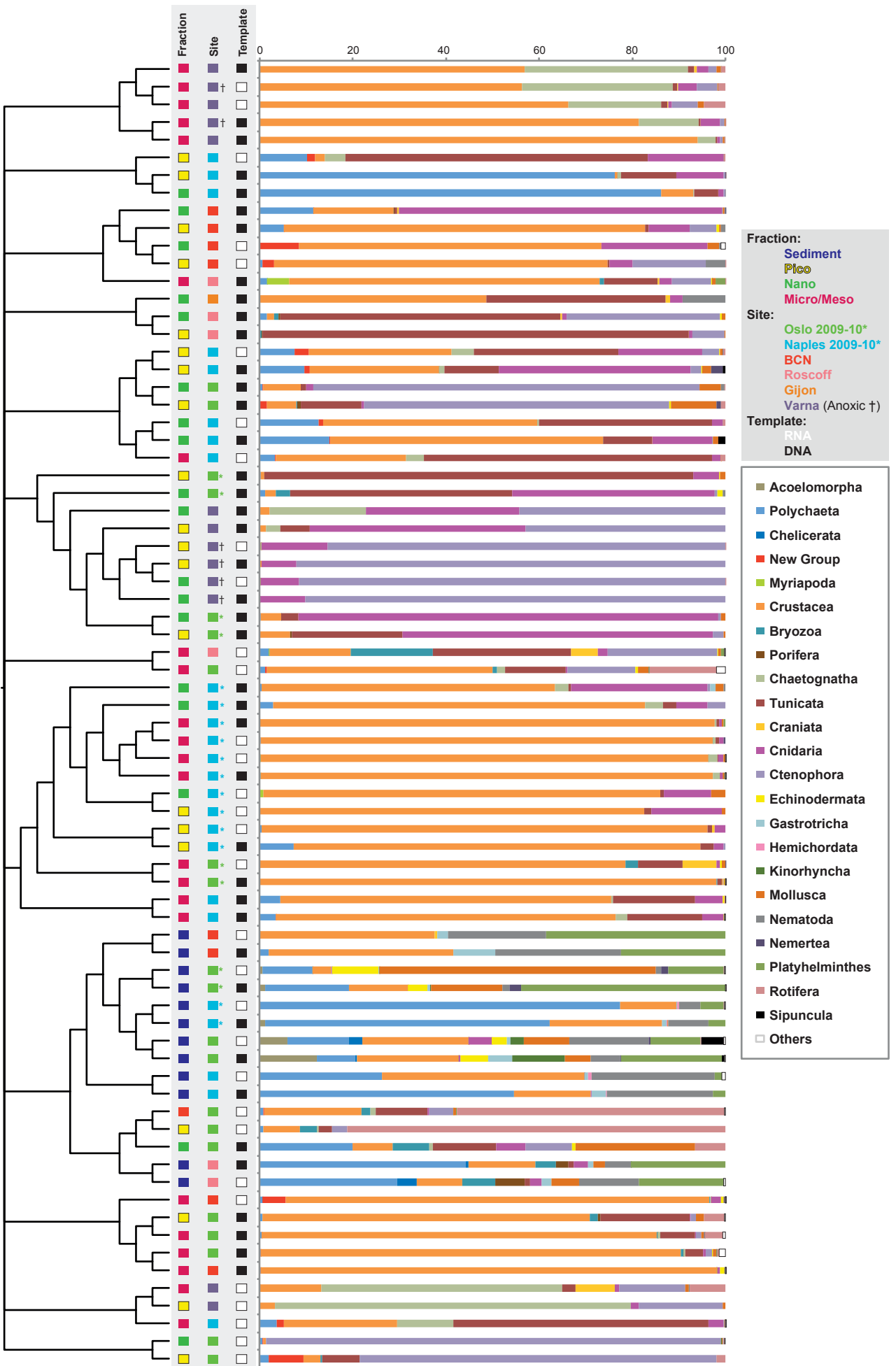
Subsurface

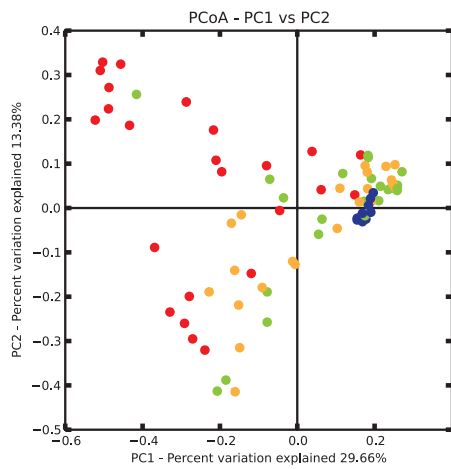
DCM

Sediments

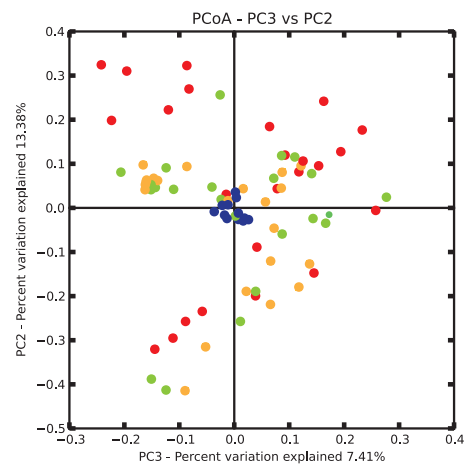


(a)**(b)****(c)**



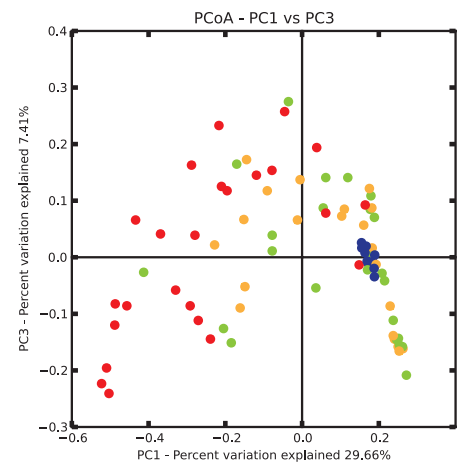
(a)

Picoplankton

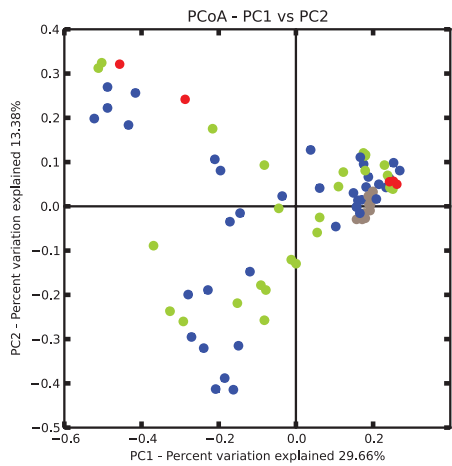


Nanoplankton

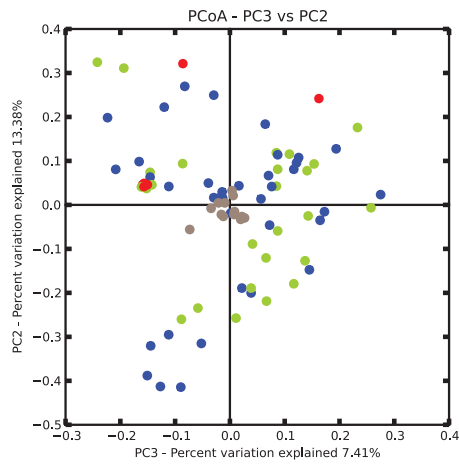
Micro/Mesoplankton



Sediment

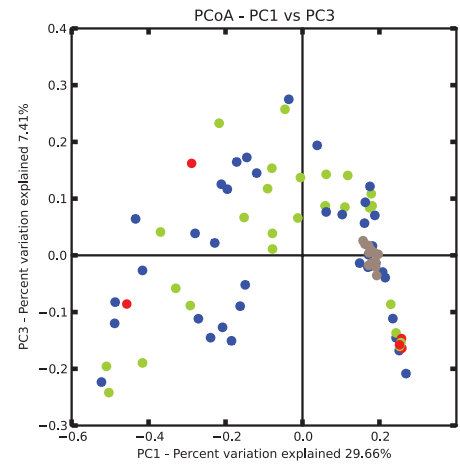
(b)

Subsurface

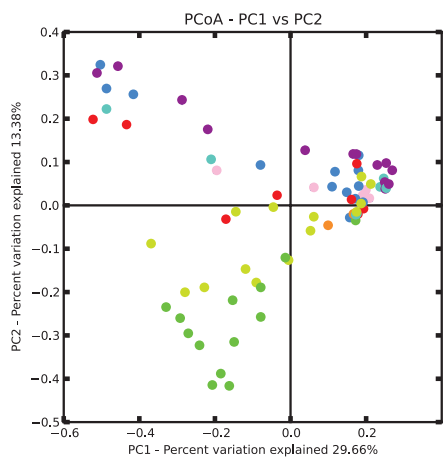


DCM

Anoxic



Sediment

(c)

Blanes

Gijon

Naples_2009

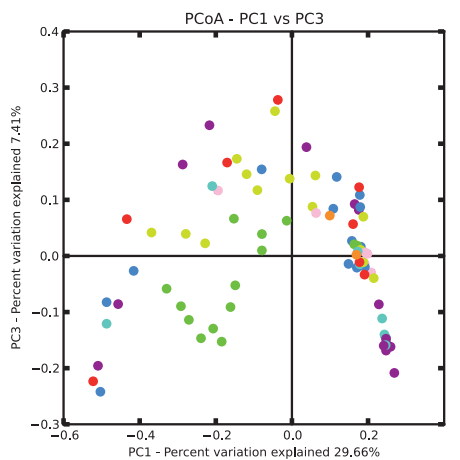
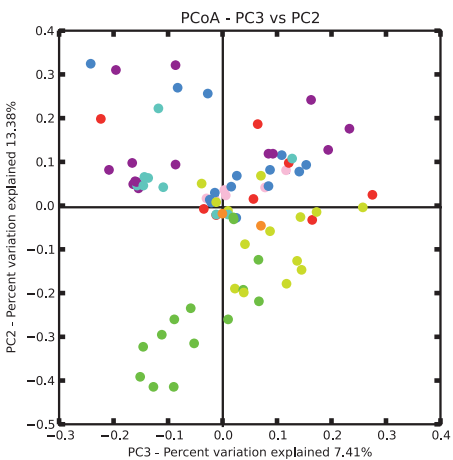
Naples_2010

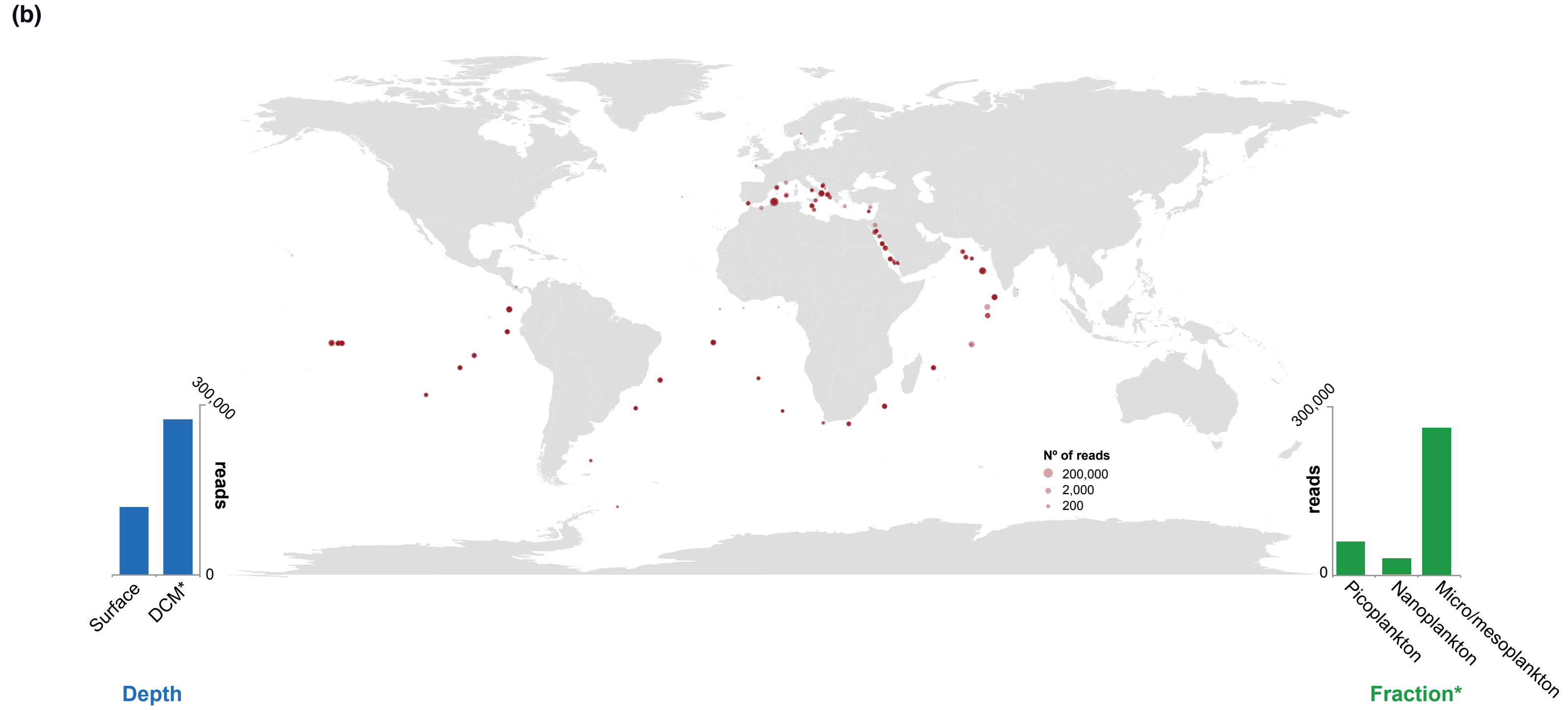
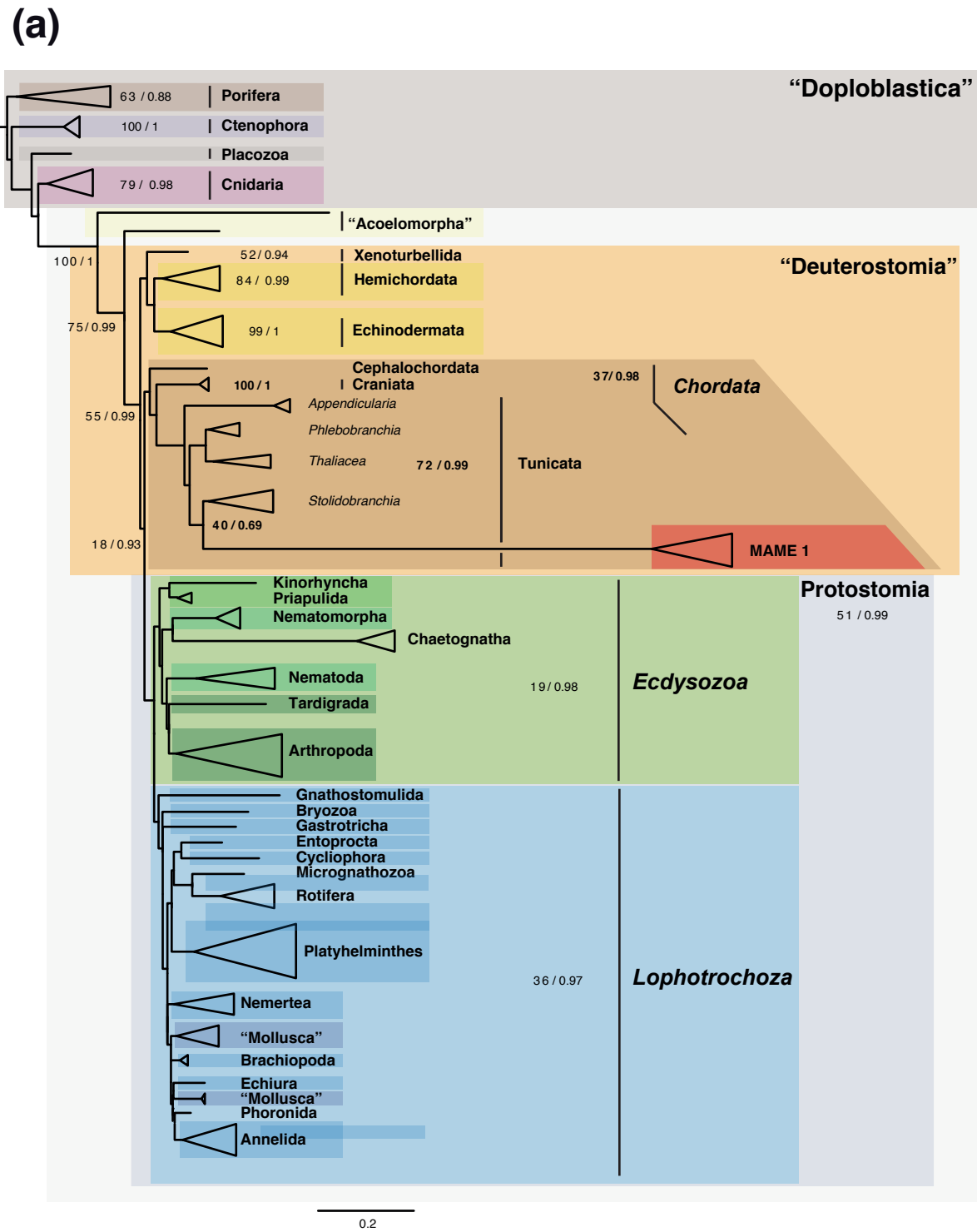
Oslo_2009

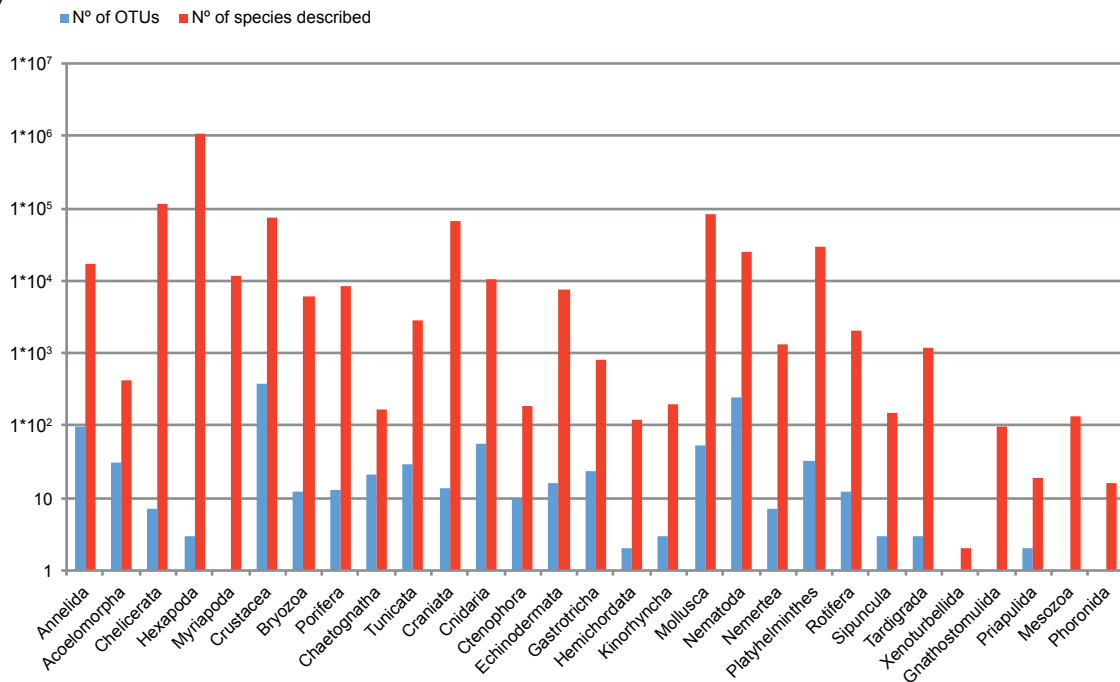
Oslo_2010

Roscoff

Varna





(a)**(b)**