Metatranscriptomics reveals a shift in microbial community composition and function during summer months in a coastal marine environment

Ben J. G. Sutherland[1,2,3*ǂ], Jan F. Finke[2,4*ǂ], Robert Saunders[5], Snehal Warne[5], Angela D. Schulze[1], Jeff H. T. Strohm[1], Amy M. Chan[2], Curtis A. Suttle[2,6,7,8] and Kristina M. Miller[1]

1 Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, BC, Canada V9T 6N7

2 Department of Earth, Ocean and Atmospheric Sciences, The University of British Columbia, Vancouver, BC, Canada V6T 1Z4

3 Sutherland Bioinformatics, Lantzville, BC, Canada V0R 2H0

4 Hakai Institute, PO Box 309, Heriot Bay, BC, Canada, V0P 1H0

5 RKS Labs, Parksville, BC, Canada V9P 2A2

6 Department of Microbiology and Immunology, The University of British Columbia, Vancouver, BC, Canada V6T 1Z3

7 Department of Botany, The University of British Columbia, Vancouver, BC, Canada V6T 1Z4

8 Institute for the Oceans and Fisheries, The University of British Columbia, Vancouver, BC, Canada, V6T1Z4

ǂ these authors contributed equally to this work.

*Author for correspondence: BJGS

Current Address:

Sutherland Bioinformatics, Lantzville BC, Canada, V0R 2H0

Sutherland.Bioinformatics@protonmail.com

*Author for correspondence: JFF

Department of Earth, Ocean and Atmospheric Sciences, The University of British Columbia, Vancouver, BC, Canada, V6T 1Z4

Email: jfinke@eoas.ubc.ca

Data Deposition: Raw metatranscriptome sequence data has been uploaded to SRA under BioProject PRJNA738385, BioSample SAMN19728638-SAMN19728657.

Abstract

Temperate coastal marine waters are often thermally stratified from spring through fall but can be dynamic and disrupted by tidal currents and wind-driven upwelling. These mixing events introduce deeper, cooler water with a higher partial pressure of CO_2 (pCO2), and its associated microbial communities to the surface. Anecdotally, these events impact shellfish hatcheries and farms, warranting improved understanding of changes in composition and activity of marine microbial communities in relation to environmental processes. To characterize both compositional and functional changes associated with abiotic factors, here we generate a reference metatranscriptome from the Strait of Georgia over representative seasons and analyze metatranscriptomic profiles of the microorganisms present within intake water containing different pCO_2 levels at a shellfish hatchery in British Columbia from June through October. Abiotic factors studied include pH, temperature, alkalinity, aragonite, calcite and pCO_2. Community composition changes were observed to occur at broad taxonomic levels, and most notably to vary with temperature and pCO_2. Functional gene expression profiles indicated a strong difference between early (June-July) and late summer (August-October) associated with viral activity. The taxonomic data suggests this could be due to the termination of cyanobacteria and phytoplankton blooms by viral lysis in the late season. Functional analysis indicated fewer differentially expressed transcripts associated with abiotic variables (e.g., pCO_2) than with the temporal effect. Microbial composition and activity in these waters varies with both short-term effects observed alongside abiotic variation as well as long-term effects observed across seasons. The analysis of both taxonomy and functional gene expression simultaneously in the same samples by environmental RNA (eRNA metatranscriptomics) provided a more comprehensive view for monitoring water bodies than either would in isolation.

Keywords: metatranscriptome assembly; metatranscriptomics; microbial community; oyster; pCO2; seasonality

Introduction

Metatranscriptomics characterizes expressed genes (i.e., RNA transcripts) that are present in an environmental sample. These transcripts may be within the cell, or in the extracellular space. While metagenomics profiles the taxonomy of a sample, metatranscriptomics can profile the biological functions that are present in the sample and are active or differentially accumulated in particular environments (Aguiar-Pulido et al. 2016). Metatranscriptomics can also be used to deduce taxonomic information for dominant taxa in communities (Shi et al. 2011; Neves et al. 2017; Marcelino et al. 2019), in particular with longer transcriptome contigs, which are expected to produce correct taxonomic assignments (Shakya et al. 2019). Metatranscriptomics is typically conducted through next-generation sequencing, which has the benefit of identifying novel genes and functions not known to be used by the identified taxa (Gilbert et al. 2008). Millions of marine microorganisms and viruses occur within a millilitre of seawater and their abundance and community composition varies with regions and environmental variables (Gilbert et al. 2011; Wigington et al. 2016; Finke et al. 2017; Yu et al. 2018; Gregory et al. 2019). A large scale global study found indications that the functional gene content in marine microbial samples is largely shaped by taxonomic composition (Salazar et al. 2019). However, the detailed distribution of generated community functions and their link to environmental variables remains to be studied (Moran 2015). Altogether, the environmental metatranscriptomic approach has strong potential for profiling active processes and community composition under changing conditions in marine molecular ecology (Moran et al. 2013; Cristescu 2019).

Coastal water microbial communities are shaped by widely varying environmental conditions, including temperature, salinity, and especially pCO_2 and pH. (Salisbury et al. 2008; Joint et al. 2011; Ray et al. 2012; Lv et al. 2016; Lee et al. 2017). These are factors that are impacted by seasons, tides, or even biological activity. Microbes tolerate pH fluctuations under regular conditions and co-vary with water bodies in the short-term (Joint et al. 2011), but generally adjust in the long term. Microbial, bacterial and protist communities show significant responses to high pCO_2 concentrations (Ray et al. 2012; Zhang et al. 2015; Thomson et al. 2016). In the past, average pCO_2 concentrations have increased from 280 parts per million (ppm) (Friedlingstein et al. 2019) to currently 400 ppm (Blunden and Arndt 2020) and are deemed to increase to an average of 1000 ppm by the end of this century, with an associated rise of sea surface temperature by 1ºC and drop of pH by 0.29 (Kirtman 2013; Pachauri 2014; Bindoff 2019).

Many microorganisms are expected to vary with temperature and/ or pCO_2, and this may include shellfish pathogens such as *Vibrio* (Gomez-Leon et al. 2005; Green et al., 2019), *Roseovarius* (Maloy et al. 2007), and *Mikrocytos* (Carnegie et al. 2003), as well as harmful algae (Landsberg 2002) and viruses (Renault et al. 2014). By the nature of their intertidal habitat and filter feeding lifestyle, in particular oysters are affected by varying environmental conditions and microbiomes (Wegner et al. 2013; Lokmer et al. 2016;

93 Cho 2019). As a consequence, changes in microbial communities are expected to affect oyster hatcheries.
94 Interestingly, for detecting an oyster parasite, environmental RNA (eRNA) was found to be superior to
95 detect the species than when using environmental DNA (eDNA) (Merou et al. 2019). The impact of
96 temperature increases and pCO_2 on overall microbial communities and microbial activity is difficult to
97 predict, but is expected to be significant with wide-ranging consequences. This serves as a motivation to
98 develop comprehensive ways to improve our understanding of changes in marine microbial community
99 composition and activity in response to environmental processes.
100 Here we compare the microbial composition, active genes, and enriched functional categories of
101 differentially expressed genes within samples of intake water at a shellfish hatchery over a range of pCO_2
102 levels, temperatures, and months. We apply comparative metatranscriptomics by first developing a *de novo*
103 metatranscriptome assembly using the samples in the study. We take a taxonomic approach to view the
104 community composition of each sample. In parallel, we quantify the relative expression levels of each
105 transcript among the samples and perform differential gene expression analysis in relation to environmental
106 metadata. Collectively, this work profiles hatchery intake water across 20 dates over a two-year period.
107 Using environmental metadata, multivariate clustering, differential expression analysis, and functional
108 enrichment analysis, we characterize these different sampling events in relation to date of collection and
109 environmental parameters.
110

111 **Materials and Methods**

112 *Water collection, RNA extraction and environmental variables*

113 Water samples were collected from the ocean water intake at a Pacific oyster hatchery in Qualicum Bay,
114 central East Coast Vancouver Island, British Columbia over a two year period, with the majority of samples
115 collected in 2015 (Supplemental File S1). On each collection day, six litres of water were taken and filtered
116 through sterile 0.22 µm cellulose acetate filters (Millipore, Burlington, MA). Filters were stored at -80ºC.
117 Temperature was measured by a mercury thermometer, salinity with a refractometer and pH with a glass
118 probe pH meter (Jenco, San Diego, CA). Alkalinity was measured with a HI901 titrator (Hanna Instruments,
119 Smithfield, RI). pCO_2 was determined with a LI840A infrared gas analyzer (Li-Cor, Lincoln, NB).
120 Aragonite and calcite were calculated with the CO2SYS.BAS program
121 (https://github.com/jamesorr/CO2SYS-Excel/blob/master/CO2sys_mod.bas). The filters were then used
122 for RNA extraction by the Power Water RNA isolation kit (MoBio, Carlsbad, CA) following
123 manufacturer's instructions, including the alternate lysis step. Output total RNA was depleted for ribosomal
124 RNA and prepped for RNA-seq using the Scriptseq Complete Gold (Epidemiology) kit (Illumina, San
125 Diego, CA). A total of ~50 ng of ribosomal depleted RNA was used as an input to transcriptome libraries.
126 Individual barcoded libraries were randomly multiplexed into groups of four samples using equimolar

127 quantities and sequenced on a MiSeq using a v3 600 reagent kit (Illumina, San Diego, CA) to generate
128 paired-end 250 bp reads. Environmental variables were analyzed in a principal component analysis (PCA)
129 using the Vegan package (Oksanen et al. 2016) in the R environment (R Core Team, 2022), where missing
130 values were imputed with unit averages.
131
132 *Bioinformatics and metatranscriptome assembly*
133 Raw and quality trimmed sequence data were inspected with FastQC (Andrews 2010) and MultiQC (Ewels
134 et al. 2016). Quality trimming was conducted to remove low quality reads and adapters with Cutadapt
135 (Martin 2011) using flags -q 20 to remove < Q20 data from the 3'-end of the read and -m 50 to remove
136 reads shorter than 50 bp. Results were output as an interleaved fastq. Putative ribosomal RNA (rRNA) reads
137 were removed using SortMeRNA (Kopylova, Noé, and Touzet 2012) using all suggested rRNA databases,
138 and therefore enriching for putative messenger RNA (mRNA) *in silico*.
139 Two different approaches were taken to assemble the reference metatranscriptome. First, a
140 reference metatranscriptome was assembled using the metatranscriptome-optimized assembler IDBA-tran
141 (Peng 2012) with a majority of the samples (16 / 20 samples; 191,303,324 reads), not including all samples
142 due to computational constraints. Samples used for this approach comprised an equal number of samples
143 from both ends of the pCO_2 range present in the collections. This assembly was conducted with default
144 settings and run on 56 threads. Second, a reference metatranscriptome was generated by first assembling
145 each of the 20 libraries individually using IDBA-tran, and then merging these 20 assemblies into a single
146 assembly using CD-HIT-EST (Li and Godzik 2006). CD-HIT-EST merged contigs at 95% similarity, and
147 dedupe.sh of BBTools (Bushnell, Rood, and Singer 2017) was used to de-duplicate with default parameters
148 (i.e., the 'merged assembly approach'). These assemblies were compared based on the total number of
149 contigs, total length, multi-mapping proportions, and mapping percentages to select the best assembly.
150 Reads for each sample were aligned against the reference metatranscriptome using Bowtie2
151 (Langmead and Salzberg 2012) in end-to-end mode allowing for multi-mappings. A maximum of 40
152 alignments were retained for each read. Alignments were then filtered to remove low quality mappings (i.e.,
153 retain mapq ≥ 2). Retained alignments were quantified using eXpress (Roberts et al. 2013). Effective counts
154 from eXpress were output into a table in R, and imported into edgeR (Robinson and Oshlack 2010).
155 Filtering was conducted to only retain contigs against which at least five reads mapped in the sample with
156 the fewest reads (i.e., 3.86 counts per million; CPM), and requiring that the contig was represented at this
157 CPM level or higher in at least five samples. Retained transcripts were normalized for library size using the
158 TMM normalization method of edgeR v.3.28.1.
159

Taxonomic community analysis

The expressed transcripts of the metatranscriptome were annotated for taxonomic identity using BLASTn (Camacho et al. 2009) against the nt database of NCBI, retaining a maximum of 100 alignments and descriptions per record. Best annotations were selected based on the e-value and a minimum cut-off at $E < 10^{-5}$. Best match phylogenetic lineages of annotated transcripts were extracted with a custom python tool (see Data Accessibility) based on the subject sequence id using the ranked lineage database (NCBI), and exporting different levels of the taxonomy. The taxonomic overview and characterization used all expressed genes, at the kingdom, phylum and order level, and further analysis was conducted at the genus level. A canonical correlation analysis (CCA) of genus abundances and environmental variables, as well as ANOVA tests for significance of regressions were performed with the vegan package in R. To compare community composition at different lineage levels, pairwise distances were calculated using the Bray-Curtis dissimilarity measure in vegan, and Mantel tests were performed using the ade4 package (Dray and Dufour 2007). Linear models of genus abundance versus pCO_2 concentration and t-tests for genus abundance between seasons were conducted in R. Significant regressions were defined by $p < 0.05$, adjusted p-values were determined using the Benjamin-Hochberg procedure.

Differential expression analysis

To view detailed gene expression trends among samples, log2 CPM values were used for multidimensional scaling (MDS) plots. The plotMDS function of *limma* v.3.42.2 (Ritchie et al. 2015) was used to generate MDS plots using both the leading log2-fold-change as well as a PCA (gene.selection = common). Samples were grouped into binary groups for pCO_2 (low/normal versus high) and season for differential expression analysis. High pCO_2 was considered when the value was greater than 700 ppm, and low/normal was considered less than 700 ppm. Early summer was considered as June through July, and late summer was considered for August through October (see Table 1). Expression levels for each transcript were analyzed in a generalized linear model (i.e., glmFit and glmLRT) in edgeR to analyze the effect of pCO_2 and the effect of early vs. late summer, and their interaction. Genes with pairwise $p \leq 0.05$ after Benjamini-Hochberg multiple test correction were considered differentially expressed.

To annotate transcripts with functions, expressed transcripts were assigned UniProt descriptions and identifiers by using BLASTx (Altschul 1997) to align contigs against the Swiss-Prot database (UniProt 2017) using the pipeline go_enrichment (Eric Normandeau, see Data Accessibility), with flags --max_target_seqs 1 in outfmt 6 format, and only retaining hits with $E < 10^{-5}$. The UniProt identifier was used as an input for Gene Ontology (GO) enrichment analysis in DAVID bioinformatics (Huang, Sherman, and Lempicki 2009), using differentially expressed lists compared against all expressed genes in the metatranscriptome for those transcripts annotated with UniProt identifiers.

Results

Sampling and environmental conditions

Intake water at the commercial oyster hatchery on east coast Vancouver Island, BC was sampled on four separate days in 2014 (June-August) and on 16 days in 2015 (June-November), with the total samples collected being seasonally balanced between the early summer (i.e., June-July, n = 11) and late summer (i.e., August-October, n = 9). Environmental variables were measured from the intake water, including pH, temperature, salinity, alkalinity, aragonite, calcite and pCO_2, as shown in Table 1 (for complete data see Supplemental File S1).

The focal variable of this study, pCO_2, ranged from 81-1060 ppm (average = 588 +/- 290 ppm). Following IPCC assessments (Gattuso 2014) we classify samples into low $pCO_2 < 400$ ppm (n = 6), medium pCO_2 400-700 ppm (n = 4), and high $pCO_2 > 700$ ppm (n = 10) samples. A two dimensional PCA of environmental conditions shows variability among samples, separating them by low, medium and high pCO_2 conditions (Figure 1). Samples S25 and S35, having concentrations of 397 and 386 ppm pCO_2 cluster close to the medium pCO_2 samples. Samples S22, S26 and S36 have almost identical environmental conditions. The PCA describes 77% of total variation in the first two principal components, with 52% explained by PC1 and 25% by PC2. The most influential variables in the PCA are pCO_2 and pH, separating the samples by their pCO_2 classes. The contributions of pCO_2 and pH are inverse, as are salinity and temperature. Calcium carbonate ($CaCO_3$), aragonite and calcite differ in their variation from the main axes described by pH/ pCO_2 and salinity/ temperature. Notably, the grouping of samples by environmental variables does not display clustering by sampling months and the associated early vs. late summer classification (Figure 1).

Sequencing and assembly

The metatranscriptome libraries each yielded on average 14.7 M (s.d. = 2.8 M) paired-end reads. Depletion of rRNA in library preparation removed most of the rRNA from a majority of the samples, although for six of the 20 samples, 30-65% of sequenced library remained as rRNA (Figure S1). Residual rRNA was removed *in silico*, and is primarily comprised of bacterial 23s and eukaryotic 28s rRNA, but also archaeal rRNA.

The reference metatranscriptome was assembled from a total of 191,303,324 mRNA reads (33,050,961,278 bp) originating from pooling the non-rRNA data from 16 of the samples (n = 8 from each of low pCO_2 and high pCO_2). This input produced a final assembly of 8,003,896 contigs (total length = 2,468,090,451 bp; longest contig: 103,407 bp; N50 = 272 bp; number contigs > 500 bp = 600,691). This assembly was compared to other assemblies that used fewer libraries, or those that were individually assembled by sample then subsequently merged (i.e., 'merged assembly'; see Methods). The collectively-

228 assembled contigs with the highest number of input samples show fewer multi-mapping reads than did the
229 merged assembly. The percentage of reads aligning a single time increases substantially until four libraries
230 were added, then tapers off to not increase notably with eight, 12, or 16 libraries. The addition of more
231 libraries after four libraries did not substantially increase the percentage of reads mapping, but it also did
232 not increase redundancy as evaluated by multi-mapping. The 16 library collectively-assembled assembly
233 also has a similar number of contigs and total length to the other assemblies (see Supplemental Results;
234 Figure S2). Therefore, this collectively-assembled assembly was chosen to be used for all downstream
235 functional analyses (i.e., the 'final reference metatranscriptome assembly').
236 Aligning reads from individual samples against the final reference metatranscriptome assembly
237 resulted in an average alignment rate per sample of 56% (median = 57%; min. = 37%; max = 68%), with
238 an average of 40% of the total alignments per sample with both read pairs aligning concordantly a single
239 time. On average per sample, 12% of reads align concordantly more than once, which may indicate
240 remaining redundancy in the assembly. Applying low expression filters removed the majority of contigs
241 from the metatranscriptome, retaining 32,866 contigs with CPM ≥ 3.86 in at least 5 of 20 samples.
242
243 *Taxonomic analysis*
244 To infer the community compositions of samples, the taxonomic lineages for expressed transcripts were
245 analyzed. Taxonomic assignment was successful for 19,315 of the 32,866 expressed transcripts (59%), but
246 the taxonomic resolution varied among contigs. For example, 89% of the annotated transcripts resolve to
247 at least the phylum level, 80% to class, 79% to genus and only 30% to the species level. The data includes
248 some expected non-microbial taxa such as fish or molluscs and some suspected false taxonomic
249 assignments, but is vastly dominated by microbes. A total of 708 genera in 104 classes and 58 phyla are
250 annotated, but only 10 microbial phyla are present at an abundance level of greater than 0.5% of the total
251 reads. When combined, these 10 phyla account for over 95% of all reads, yet several other phyla are
252 represented in the data as well (Supplemental File S2). These 10 dominant phyla show an abundance of
253 reads assigned to bacteria (~74%), specifically *Proteobacteria*, *Bacteriodetes*, *Firmicutes*, *Actinobacteria*
254 and *Cyanobacteria*. Overall, fewer reads are assigned to archaea (~11%), viruses (~9%) and eukaryotes
255 (~1.6%). *Archaea* are represented by *Thaumarcheota* and *Euryarcheota*. Viruses are largely in the class
256 *Caudovirales* (*Uroviricota*) which includes, for example, bacteriophages and cyanophages, and the
257 *Nucleocytoviricota* that represent the nucleocytoplasmic large DNA viruses (NCLDV). Eukaryotes are in
258 the phylum *Bacillariophyta.* Figure 2 shows the variation in community composition of these ten phyla
259 among samples, *Proteobacteria* and *Bacteroidetes* are clearly dominating in most samples, especially in
260 the early season. *Cyanobacteria* and *Bacillariophyta* are also mostly present in early season samples. In

261 late season there is an increase in relative abundance for the *Thaumarchaeota* and *Euryarchaeota*, but
262 especially the *Uroviricota* and *Nucleocytoviricota* viruses.
263 The overall variation in community composition among samples was evaluated based on Bray-
264 Curtis similarity in a pairwise distance analysis. When compared at different taxonomic levels, variations
265 in community compositions are congruent, and this is true for comparisons of genus to class (R=0.97), class
266 to phylum (R=0.91) and genus to phylum (R=0.90), all showing significant congruence (Mantel test $p \leq$
267 0.01). A CCA of taxa composition for the 708 genera and environmental variables produces a significant
268 model (P=0.013), explaining 56% of the total variation in taxon composition through environmental
269 variables (Figure 3). Of that variation the first dimension (CCA1) describes 39% of the variation and the
270 second dimension (CCA2) describes 27%. The relationship between sample composition as indicated by
271 the sample labels and the genera (dots) are shown, where taxa are coloured according to the corresponding
272 four kingdoms. Generally, kingdoms are distributed across the CCA, but eukaryotes show some grouping
273 with samples S21, S15 and S13. Viruses show grouping with late summer samples S31, S34, S36 and S37.
274 The effect and strength of environmental variables on the sample composition is indicated by vectors.
275 Temperature and salinity, and pH and pCO_2 describe the predominant dimensions of variation. A stepwise
276 regression determined that temperature (p=0.001) and pCO_2 (p=0.028) are significant environmental
277 variables affecting the composition of samples. Additionally, early vs. late summer is significant (p=0.023)
278 in separating samples by composition. Correlating all 708 genera to pCO_2, our main environmental variable
279 of interest, revealed 67 genera showing a linear correlation of their $\log_{10}$ transformed abundances to pCO_2
280 concentrations ($p < 0.05$; Supplemental File S3), although of these, only two are significant based on
281 multiple test corrected values (i.e., *Methyloceanibacter* and *Deferribacter*). The genera with the highest
282 linear correlations are predominantly in the *Proteobacteria*, *Cyanobacteria,* and *Picornavirales* and
283 *Caudovirales*. Similarly, 93 genera show significant variation between the early and late season samples,
284 especially *Caudovirales* and *Algavirales* (Supplemental File S4). Of these, 23 are significant based on
285 multiple test corrected values, including *Caudovirales*. The top ten genera for both models are summarized
286 in Table 2.
287
288 *Gene expression analysis*
289 After the low expression filter was applied, in total 22,121 (67%) of the expressed contigs (n = 32,866)
290 were functionally annotated with UniProt identifiers (BLASTx E < 10-5). All gene expression analysis was
291 conducted using both the annotated and unknown transcripts except for functional enrichment analysis,
292 which depends on the UniProt identifier.
293 Using the filtered expression data, an unsupervised multidimensional scaling plot (MDS plot) was
294 used to group samples by similarity in gene expression (Figure 4). Based on gene expression signatures this

295 plot indicates similarity among samples from a common season, irrespective of sampling year, where
296 samples from June-July (early summer) are separated from August-October (late summer/fall). Early season
297 samples (blue) and late season samples (red) separated on PC1 (Figure 4), with one sample being an
298 exception, S19, which groups outside of its season. Notably, S19 was the only sample from 2014 that was
299 collected in August or later (collected Aug. 6Th, 2014; Supplemental File S1). No grouping within the MDS
300 plot was observed for an effect of year nor of technical aspects of sample handling.

301 Based on the observations of the effect of pCO_2 as a key environmental variable on taxonomic
302 composition and a focal variable of the study, a differential expression analysis was conducted using pCO_2
303 separated into either high or medium/low levels, and early summer vs. late summer as binary explanatory
304 variables, as well as their interaction. There is a larger effect of early vs. late summer than of pCO_2, where
305 2,765 transcripts are found differentially expressed based on early vs. late, and 720 based on pCO_2. Of these,
306 45 are differentially expressed in both contrasts (Supplemental File S5). Of the transcripts affected by
307 season, 318 are over-expressed in early summer, and 2447 are over-expressed in late summer/early fall. Of
308 the transcripts affected by pCO_2, 553 are over-expressed by high pCO_2 and 167 are under-expressed. There
309 are no transcripts showing a significant interaction effect of sampling period and pCO_2.

310 Differentially expressed transcripts were used for Gene Ontology (GO) enrichment analysis (Table
311 3). Annotated transcripts overexpressed in the early summer ($n = 318$) are enriched for metabolic processes
312 and biosynthesis (e.g., cellular macromolecule metabolic process; $n = 32$ genes; $p = 0.01$), and those
313 overexpressed in the late summer are most notably enriched with viral process (biological process; $n = 53$;
314 $p << 0.0001$) and virion (cellular component; $n = 30$; $p << 0.0001$). The viral process category is mainly
315 enriched with transcripts annotated from phage taxa ($n = 40$ of 53 transcripts; Supplemental File S6).

316 Genes overexpressed in high pCO_2 include cobalamin biosynthetic process ($n = 5$, $p = 0.003$),
317 organic substance biosynthetic process ($n = 81$, $p = 0.004$), and DNA replication ($n = 17$, $p =0.005$). Genes
318 with lower expression at high pCO_2 include protein maturation ($n = 6$, $p << 0.001$) and response to abiotic
319 stimulus ($n = 6$, $p < 0.001$).
320

321 # Discussion

322 *Profiled characteristic conditions indicate short-term environmental fluctuations*

323 The collected samples reflect typical intake water of oyster hatcheries from standard aquaculture practices
324 on East Coast Vancouver Island (Helm 2004), spanning over several months and two years. The variables
325 were analyzed in a PCA to understand the covariation of environmental variables and how they shape the
326 sampling conditions. The variation in environmental data describes the characteristic water flow and carbon
327 chemistry of intertidal oyster habitats and the region of sampling (Strait of Georgia) (Dickinson et al. 2012;
328 Ianson et al. 2016); salinity and temperature are inversely affected by influxes of cold seawater with

329 relatively high salinity and warmer fresh water with lower salinities. Water with higher pCO_2 concentrations
330 is characterized by lower pH values and the combined effect of pH, pCO_2, and salinity is reflected in the
331 $CaCO_3$ concentrations, with calcium being one of the salts in seawater. This relationship shows in the
332 sample separation by low, medium and high pCO_2 values and the force loadings of environmental variables.
333 Calcite and aragonite are both derivatives of $CaCO_3$, varying based on the pCO_2 and the pH of water (Doney
334 2010), and their influence is thus mediated between these two variables. The clustering of samples observed
335 when considering environmental data highlights the impact of pCO_2 and salinity and suggests that
336 variations in environmental conditions observed in the scale of the present study (two years) are driven by
337 short term influxes of fresh or marine waters through tidal cycles and upwelling of deep waters.
338
339 *Taxonomic composition varies with abiotic factors and early and late summer*
340 The microbial community analysis here is based on sequences with taxonomic assignment. However, we
341 expect this to be a representative description, with the possible exception of largely uncharacterized taxa in
342 reference databases. Derived from taxonomic assignments of contigs and their respective expression levels,
343 community similarity among samples proved to be congruent across different taxonomic levels. This
344 validates the approach of using taxonomic assignments down to the genus level. Importantly, it also
345 indicates that shifts in community composition happen at higher taxonomic levels and not just at the genus
346 or species level. Contigs assigned to non-microbial taxa or terrestrial taxa are expected to be either the
347 product of dispersed tissue or cells in the water column or alternatively incorrect taxonomic assignment.
348 Across samples the microbial communities are clearly dominated by *Proteobacteria*, *Bacteriodetes*,
349 *Firmicutes*, *Actinobacteria* and *Cyanobacteria*. Both eukaryote and archaea microorganisms are notably
350 less represented, and even matched in abundance by viral sequences. The described bacterial phyla are
351 commonly found to be dominant in marine (Sunagawa et al. 2015) and coastal waters (Yung et al. 2016;
352 Yu et al. 2018), and constitute the microbiome of oysters themselves (Trabal et al. 2012; Lokmer et al. 2015;
353 Lokmer et al. 2016; Dubé et al. 2019; Stevick et al. 2019). In particular, *Proteobacteria* (*Vibrionales*) and
354 *Bacteriodetes* (*Flavobacteria*) represent common marine microbes and pathogens (Gomez-Leon et al. 2005;
355 Schulze et al. 2006; Paillard et al. 2008; Chen et al. 2017). Observations by Stevick and co-workers (Stevick
356 et al. 2019) found that *Cyanobacteria* were among the dominant phyla in the oyster rearing water
357 communities and *Synechococcus* is a ubiquitous cyanobacterial genus in coastal environments (Partensky
358 et al. 1999; Tai and Palenik 2009). Similarly, the abundant eukaryote *Bacillariophyta* and *Chlorophyta*,
359 both present in the samples taken here, are common phytoplankton in coastal waters (Worden, Nolan, and
360 Palenik 2004; Armbrust 2009).
361 Matching the microbial community, the dominant *Caudovirales* includes general bacteriophages
362 and specifically cyanophages (Weinbauer and Rassoulzadegan 2004). The present results are thus mirroring

363 the general ubiquity of heterotrophic bacteria, but also indicating their lysis. The *Nucleocytoviricota* include
364 giant viruses commonly infecting protists (Fischer et al. 2010; VanEtten et al. 2010), but also
365 phycodnaviruses infecting eukaryote phytoplankton (e.g., *Chlorophyta*), both being abundantly present in
366 coastal waters (VanEtten et al. 1982, 2002). Taken together, the observation of a low presence of
367 cyanobacteria sequences and *Bacillariophyta* sequences alongside high abundance of plankton viruses and
368 phages in some late season samples, may indicate the lysis of cyanobacteria and algae blooms later in the
369 season. This observation would match phytoplankton bloom patterns described for the northern Strait of
370 Georgia where diatoms (*Bacillariophyta*) and prasinophytes (*Chlorophyta*) are mostly dominant with only
371 periodic cyanobacteria blooms (Del Bel Belluz et al. 2021). Additionally, the general decrease in
372 *Proteobacteria* in the late season samples is observed alongside increased presence of phage (*Uroviricota*)
373 activity, which also fits with the understanding that bacterial blooms are terminated by viral lysis. As well,
374 viruses show a correlation to late summer samples in the canonical correspondence analysis and the t-test
375 following multiple test correction (i.e., *Caudovirales*).

376 In the linear models of taxa to pCO_2, our main environmental variable of interest, 67 correlations
377 are significant based on individual tests ($p < 0.05$), but only two are significant based on multiple test
378 corrected p-values, and so the 65 taxa should be considered as putatively characteristic of changes in pCO_2
379 due to water influx. Future studies may indicate the broader reliability of the finding in these taxa. Several
380 of these 67 taxa, including *Proteobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Firmicutes* and *Deferribacteres*
381 match bacterial taxa with abundances that are associated with tidal cycles and salinity (Lee et al. 2017;
382 Chen et al. 2019). There remains a large portion of unexplained variation in the taxonomic data when all
383 taxa are considered. The CCA shows that only about 50% of the variation among microbial communities
384 could be explained by the combined environmental variables. The effect of environmental variables on the
385 taxonomic community data in the CCA matches their interplay in the PCA based on environmental data
386 (Figure 1). The CCA also confirms that early vs. late summer sampling time has a significant effect on the
387 community composition, most notably for viruses, although overall to a lesser effect than observed in the
388 functional gene expression results (see below). We could not identify the abiotic variables responsible for
389 the early vs. late summer effect, which indicates other variables that may be involved in community
390 differences are missing (e.g., nutrients or irradiance levels, among others).

391 In any case, the observed effect and explanatory levels of environmental variables matches previous
392 studies, where temperature has been shown to be a main variable correlated to community composition of
393 coastal marine bacterioplankton (El-Swais et al. 2015; Yung et al. 2016; Yu et al. 2018). Further, Yu and
394 co-workers also established the effect of pH that corresponds to our observations on the influence of pCO_2
395 on community composition. Therefore, although some abiotic or biotic variables may be missing from the
396 study, the results fit with other studies on abiotic factors influencing microbial communities.

397 Overall, the data show that the microbial communities in coastal waters vary at high taxonomic
398 levels with season. The communities further vary alongside with environmental variables that are driven
399 by tidal cycles and upwelling. As demonstrated, metatranscriptome data can be used to monitor the presence
400 of a wide range of microbes and putative pathogens in seawater, expanding on the use of specific microbe
401 probing (e.g., Merou et al. 2019).
402
403 *Differential RNA transcript abundance reveals functional variation between early and late summer*
404 Environmental variables are not only expected to alter microbial community composition, but also to
405 influence the function of cells on the transcriptomic level. Further, a change in community composition
406 does not necessarily indicate a change in functions within the community if one taxon is replaced by a
407 similar taxon with similar functions. A study comparing taxonomic and functional composition in marine
408 microbial samples revealed a disconnect between taxonomy and functions (Louca et al. 2016). In the
409 functional transcriptome analysis of the present study, early and late summer sampling shows a clear effect
410 on overall transcriptome expression, but environmental variables including pCO_2 do not show clear
411 groupings of samples in the gene expression data. Consequently, similar to the community composition
412 results, here functional variation is expected to have been influenced by factors outside of the measured
413 environmental variables (e.g., nutrients and irradiance levels, among others). Observing this pattern in gene
414 expression despite having samples spread out over two years suggests some consistency to this seasonal
415 trend. However, the grouping of the only August sample (S19) from 2014 not with the late summer but
416 rather with the early summer unlike the August samples in 2015 may indicate a difference in the exact
417 timing of the putatively seasonal shift in expression. Without more samples from 2014 after the August
418 sample (S19), it is not possible to tell if the late season signature would appear again, but the early samples
419 from both years consistently group. To determine if this late season signature persists in multiple years, and
420 when the switch from early to late occurs each year, would require future work.
421 The early vs. late summer effect is most notably enriched for transcripts involved in phage viral
422 activity, matching the findings in the taxonomic analysis. Viruses are recognized as major components in
423 marine ecosystems (Suttle, 2007; Brussaard et al. 2008), but their replication dynamics remain elusive.
424 However, in the North Pacific Ocean it has been observed that viral productivity and abundance is higher
425 in summer (July) than in winter (Jan-Feb) (Gainer et al. 2017), leading these researchers to conclude that
426 seasonality is an important consideration to understand viral dynamics (Jiang and Paul 1994; Tsai et al.
427 2013). A study within a Korean bay over different seasons found that the number of reads and unique
428 species of viruses identified differed depending on month; the most viral reads were found in March and
429 December and the fewest reads in June and September (Hwang 2017). It is recognized that viral
430 composition changes depend seasonally on a range of factors including temperature, salinity, dissolved

431 oxygen, primary production and nutrient concentrations (Brum et al. 2015; Fuhrman et al. 2015). Generally,
432 viral abundance, specifically for dsDNA viruses, has been found to correlate with nutrient concentration,
433 as well as heterotrophic bacteria abundance (Wigington et al. 2016; Finke et al. 2017). Viruses infecting
434 bacteria (i.e., bacteriophages) are typically considered among the dominant group of viruses in marine
435 environments (Breitbart et al. 2002; Steward et al. 2013). In the Korean bay study, 73% of the viral reads
436 were from bacteriophage, and 26% were from algal viruses, with only 1% involving other viruses (Hwang
437 et al. 2017). The effect of pCO_2 was also investigated in the differential expression analysis, but this had a
438 lesser effect than that the temporal effect observed. Interestingly, in the present study, overexpression of
439 protein chaperones was observed at low pCO_2, which is the opposite to previous observations where
440 metatranscriptomic responses of overexpressed chaperonin transcripts was observed in high pCO_2
441 mesocosms (Gilbert et al. 2008).

442 Marine microbial communities are often comprised of a few dominant species and many rare ones,
443 and results from the Tara Oceans Expedition found approximately 37 k bacterial and archaeal species, 100
444 k protist groups and 5.5 k double-stranded bacterial and archaeal virus populations (summarized in Moran
445 2015). The present study is successful in assessing this high degree of diversity in gene expression profiles
446 among samples. As is suggested from the community composition and transcript activity of the present
447 study, previous metatranscriptome studies have also found a substantial increase in viral transcripts after a
448 phytoplankton bloom, presumably due to infected cells in lytic stage (Gilbert et al. 2008).

449 Overall, taxonomic and transcriptomic profiling complemented each other and when combined
450 provided a comprehensive view of the changes observed in this study. This matches the observation by
451 Salazar et al. (2019) that the correlation between taxonomic composition and functional composition is
452 variable. The community composition analysis provided more information on taxa associated with abiotic
453 factors, and the functional analysis highlighted the large effect of viral activity that differed over time during
454 the summer months. Being able to assess viral activity alongside microbial data is a major advantage of the
455 metatranscriptomic (eRNA) approach applied here to assess marine microbial communities.
456

457 **Conclusions**

458 In the present study, microbial communities were derived from metatranscriptome data, which avoided
459 primer bias that can occur in amplicon-based approaches, and captures data from all superkingdoms. The
460 observed variability among microbial communities was found to be associated with both temperature and
461 pCO_2, as well as other changes that occurred between early and late summer. Changing environmental
462 conditions may occur over short time scales through upwelling or currents, or may have large, structured
463 changes that occur temporally. These temporal and abiotic factors appeared to be disconnected and were
464 two different trends occurring in the marine environment. The functional gene expression analysis pointed

465 to a strong difference in viral activity moving into the late season, and a much lesser effect of abiotic factors
466 such as pCO_2, temperature and salinity. Together, these analyses provide community composition and
467 functional gene differences associated with abiotic factors and time, and likely captured viral termination
468 of bacterial, cyanobacterial and algal blooms later in the season. Being able to indiscriminately assess
469 eukaryote, prokaryote and viral taxa, describing a combined bloom and viral lysis across several host virus
470 systems proved to be the advantage of using metatranscriptome analysis. The improved understanding of
471 changes in marine microbial communities with abiotic factors helps to monitor water bodies.
472 Metatranscriptomics of eRNA allowed the characterization of both community and gene expression activity
473 changes simultaneously, providing a comprehensive view of these dynamic water bodies.
474

475 **Acknowledgements**

476 This project would not have been possible without the involvement, contributions, and sample provision
477 from the industry partner, Island Scallops (ISL; CEO Robert Saunders). J. Finke and B. Sutherland were
478 supported during this work by a grant from the Gordon and Betty Moore Foundation (GBMF, Grant number
479 5600) awarded to C. Suttle and K. Miller. This project was supported by an Aquaculture Collaborative
480 Research and Development Program (ACRDP) grant from Fisheries and Oceans, Canada, awarded to K.
481 Miller (Grant number P-14-02-001). The authors declare that they do not have any conflicts of interest in
482 this work.
483

484 **Data Accessibility**

485 Supplemental files including the reference metatranscriptome (fasta), quantified transcript expression levels
486 (.csv), and annotation for the functional and taxonomic analyses (.txt) are available on FigShare:
487 https://doi.org/10.6084/m9.figshare.18128966.v1
488 Pipeline for analyzing metatranscriptome data: https://github.com/bensutherland/eRNA_taxo
489 Pipeline used for annotating transcripts: https://github.com/enormandeau/go_enrichment
490 Script to extract taxonomic lineages:
491 https://github.com/janfelix/bioinformatic_tools/blob/master/taxid2lineage.py
492

References

Aguiar-Pulido, Vanessa, Wenrui Huang, Victoria Suarez-ulloa, Trevor Cickovski, Kalai Mathee, and Giri Narasimhan. 2016. "Approaches for Microbiome Analysis" *Evolutionary Bioinformatics* 12: 5–16. https://doi.org/10.4137/EBO.S36436.TYPE.

Andrews, S. 2010. "FastQC: A quality control tool for high throughput sequence data" Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Armbrust, E. Virginia. 2009. "The Life of Diatoms in the World's Oceans." *Nature* 459 (7244): 185–92. https://doi.org/10.1038/nature08057.

Bindoff, Nathaniel L. 2019. "Changing Ocean, Marine Ecosystems, and Dependent Communities. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate."

Blunden, J., and D. S. Arndt. 2020. "State of the Climate in 2019." *Bulletin of the American Meteorological Society*. Vol. 101. https://doi.org/10.1175/2020BAMSStateoftheClimate.1.

Breitbart, Mya, Peter Salamon, Bjarne Andresen, Joseph M. Mahaffy, Anca M. Segall, David Mead, Farooq Azam, and Forest Rohwer. 2002. "Genomic Analysis of Uncultured Marine Viral Communities." *Proceedings of the National Academy of Sciences of the United States of America* 99 (22): 14250–55. https://doi.org/10.1073/pnas.202488399.

Brum, Jennifer R., J. Cesar Ignacio-Espinoza, Simon Roux, Guilhem Doulcier, Silvia G. Acinas, Adriana Alberti, Samuel Chaffron, et al. 2015. "Patterns and Ecological Drivers of Ocean Viral Communities." *Science* 348 (6237). https://doi.org/10.1126/science.1261498.

Brussaard, Corina P D, Steven W Wilhelm, Frede Thingstad, Markus G Weinbauer, Gunnar Bratbak, Mikal Heldal, Susan a Kimmance, et al. 2008. "Global-Scale Processes with a Nanoscale Drive: The Role of Marine Viruses." The ISME Journal 2 (6): 575–78. https://doi.org/10.1038/ismej.2008.31.

Bushnell, Brian, "BBMap – sourceforge.net/projects/bbmap/" as per: https://jgi.doe.gov/data-and-tools/bbtools/

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. 2009. "BLAST+: Architecture and Applications." *BMC Bioinformatics* 10: 1–9. https://doi.org/10.1186/1471-2105-10-421.

Carnegie, Ryan B, Gary R Meyer, Janice Blackbourn, Nathalie Cochennec-laureau, Franck C J Berthe, and Susan M Bower. 2003. "Molecular Detection of the Oyster Parasite Mikrocytos Mackini , and a Preliminary Phylogenetic Analysis." *Diseases of Aquatic Organisms* 54: 219–27. https://doi.org/10.3354/dao054219

Chen, Shicheng, Jochen Blom, Thomas P. Loch, Mohamed Faisal, and Edward D. Walker. 2017. "The Emerging Fish Pathogen Flavobacterium Spartansii Isolated from Chinook Salmon: Comparative Genome Analysis and Molecular Manipulation." *Frontiers in Microbiology* 8 (NOV). https://doi.org/10.3389/fmicb.2017.02339.

Chen, Xiaowei, Wei Wei, Jianning Wang, Hongbo Li, Jia Sun, Ruijie Ma, Nianzhi Jiao, and Rui Zhang. 2019. "Tide Driven Microbial Dynamics through Virus-Host Interactions in the Estuarine Ecosystem." *Water Research* 160: 118–29. https://doi.org/10.1016/j.watres.2019.05.051.

Cho, Anna. 2019. "Characterizing the Microbiome of Farmed Pacific Oysters from British Columbia and Alaska." https://doi.org/10.14288/1.0376801.

Cristescu, Melania E. 2019. "Can Environmental RNA Revolutionize Biodiversity Science?" *Trends in Ecology and Evolution* 34 (8): 694–97. https://doi.org/10.1016/j.tree.2019.05.003.

Del Bel Belluz, Justin Del, M. Angelica Peña, Jennifer M. Jackson, and Nina Nemcek. 2021. "Phytoplankton Composition and Environmental Drivers in the Northern Strait of Georgia (Salish Sea), British Columbia, Canada." *Estuaries and Coasts*. https://doi.org/10.1007/s12237-020-00858-2.

Dickinson, G. H., A. V. Ivanina, O. B. Matoo, H. O. Portner, G. Lannig, C. Bock, E. Beniash, and I. M. Sokolova. 2012. "Interactive Effects of Salinity and Elevated CO2 Levels on Juvenile Eastern Oysters, Crassostrea Virginica." *Journal of Experimental Biology* 215 (1): 29–43. https://doi.org/10.1242/jeb.061481.

Doney, Scott C. 2010. "The Growing Human Footprint on Coastal and Open-Ocean Biogeochemistry." *Science* 328 (5985): 1512–16. https://doi.org/10.1126/science.1185198.

Dray, S, and A B Dufour. 2007. "The Ade4 Package: Implementing the Duality Diagram for Ecologists." *Journal of Statistical Software* 22. http://dx.doi.org/10.18637/jss.v022.i04

Dubé, Caroline Eve, Chin Long Ky, and Serge Planes. 2019. "Microbiome of the Black-Lipped Pearl Oyster Pinctada Margaritifera, a Multi-Tissue Description with Functional Profiling." *Frontiers in Microbiology* 10 (JULY): 1–17. https://doi.org/10.3389/fmicb.2019.01548.

El-Swais, Heba, Katherine A Dunn, Joseph P Bielawski, William K W Li, and David A Walsh. 2015. "Seasonal Assemblages and Short-Lived Blooms in Coastal North-West Atlantic Ocean Bacterioplankton." *Environmental Microbiology* 17 (10): 3642–61. https://doi.org/10.1111/1462-2920.12629.

Ewels, Philip, Måns Magnusson, Sverker Lundin, and Max Käller. 2016. "MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report." *Bioinformatics* 32 (19): 3047–48. https://doi.org/10.1093/bioinformatics/btw354.

Finke, Jan F., Brian P.V. Hunt, Christian Winter, Eddy C. Carmack, and Curtis A. Suttle. 2017. "Nutrients and Other Environmental Factors Influence Virus Abundances across Oxic and Hypoxic Marine Environments." *Viruses* 9 (6): 1–15. https://doi.org/10.3390/v9060152.

Fischer, Matthias G, Michael J Allen, William H Wilson, and Curtis A Suttle. 2010. "Giant Virus with a Remarkable Complement of Genes Infects Marine Zooplankton." *PNAS*, 1–6. https://doi.org/10.1073/pnas.1007615107/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1007615107.

Fitzer, Susan C., Vernon R. Phoenix, Maggie Cusack, and Nicholas A. Kamenos. 2014. "Ocean Acidification Impacts Mussel Control on Biomineralisation." *Scientific Reports* 4. https://doi.org/10.1038/srep06218.

Friedlingstein, Pierre, Matthew W Jones, Michael O Sullivan, Robbie M Andrew, Judith Hauck, Glen P Peters, Wouter Peters, Julia Pongratz, Stephen Sitch, and Corinne Le Quéré. 2019. "Global Carbon Budget 2019," *Earth Syst. Sci. Data* 1783–1838.

Fuhrman, Jed A, Jacob A Cram, and David M Needham. 2015. "Marine Microbial Community Dynamics and Their Ecological Interpretation." *Nature Publishing Group* 13 (3): 133–46. https://doi.org/10.1038/nrmicro3417.

Gainer, P. Jackson, Helena L. Pound, Alyse A. Larkin, Gary R. LeCleir, Jennifer M. DeBruyn, Erik R. Zinser, Zackary I. Johnson, and Steven W. Wilhelm. 2017. "Contrasting Seasonal Drivers of Virus Abundance and Production in the North Pacific Ocean." *PLoS ONE* 12 (9): 1–18. https://doi.org/10.1371/journal.pone.0184371.

Gilbert, Jack A., Dawn Field, Ying Huang, Rob Edwards, Weizhong Li, Paul Gilna, and Ian Joint. 2008. "Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities." *PLoS ONE* 3 (8). https://doi.org/10.1371/journal.pone.0003042.

Gilbert, Jack A, Joshua A Steele, J Gregory Caporaso, Lars Steinbrück, Jens Reeder, Ben Temperton, Susan Huse, et al. 2012. "Defining Seasonal Marine Microbial Community Dynamics." *The ISME Journal* 6 (2): 298–308. https://doi.org/10.1038/ismej.2011.107.

Gomez-Leon, J, L Villamil, M L Lemos, B Novoa, and A Figueras. 2005. "Isolation of Vibrio Alginolyticus and Vibrio Splendidus from Aquacultured Carpet Shell Clam (Ruditapes Decussatus) Larvae Associated with Mass Mortalities." *Applied and Environmental Microbiology* 71 (1): 98–104. https://doi.org/10.1128/AEM.71.1.98.

Green, Timothy J., Nachshon Siboni, William L. King, Maurizio Labbate, Justin R. Seymour, and David Raftos. 2019. "Simulated Heat Wave Alters Abundance and Structure of *Vibrio* Populations Associated with the Pacific Oyster Resulting in Mass Mortality Events." *Microbial Ecology* 77 : 736-747. https://doi.org/10.1007/s00248-018-1242-9

Gregory, Ann C., Ahmed A. Zayed, Nádia Conceição-Neto, Ben Temperton, Ben Bolduc, Adriana Alberti, Mathieu Ardyna, et al. 2019. "Marine DNA Viral Macro- and Microdiversity from Pole to Pole." *Cell* 177 (5): 1109-1123.e14. https://doi.org/10.1016/j.cell.2019.03.040.

Helm, Michael M. 2004. “Hatchery Culture of Bivalves.” Rome.
Huang, Da Wei, Brad T. Sherman, and Richard A. Lempicki. 2009. “Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources.” *Nature Protocols* 4 (1): 44–57. https://doi.org/10.1038/nprot.2008.211.
Hwang, Jinik, So Yun Park, Mirye Park, Sukchan Lee, and Taek Kyun Lee. 2017. “Seasonal Dynamics and Metagenomic Characterization of Marine Viruses in Goseong Bay, Korea.” *PLoS ONE* 12 (1): 1–15. https://doi.org/10.1371/journal.pone.0169841.
Ianson, Debby, Susan E. Allen, Benjamin L. Moore-Maley, Sophia C. Johannessen, and Robie W. Macdonald. 2016. “Vulnerability of a Semienclosed Estuarine Sea to Ocean Acidification in Contrast with Hypoxia.” *Geophysical Research Letters* 43 (11): 5793–5801. https://doi.org/10.1002/2016GL068996.
Jiang, S C, and J H Paul. 1994. “Seasonal and Diel Abundance of Viruses and Occurrence of Lysogeny/Bacteriocinogeny in the Marine Environment.” *Marine Ecology Progress Series* 104: 163–72. http://dx.doi.org/10.3354/meps104163
Kim, Bo Mi, Kyobum Kim, Ik Young Choi, and Jae Sung Rhee. 2017. “Transcriptome Response of the Pacific Oyster, Crassostrea Gigas Susceptible to Thermal Stress: A Comparison with the Response of Tolerant Oyster.” *Molecular and Cellular Toxicology* 13 (1): 105–13. https://doi.org/10.1007/s13273-017-0011-z.
Kirtman, Ben. 2013. “Near-Term Climate Change: Projections and Predictability. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.”
Kopylova, Evguenia, Laurent Noé, and Hélène Touzet. 2012. “SortMeRNA: Fast and Accurate Filtering of Ribosomal RNAs in Metatranscriptomic Data.” *Bioinformatics* 28 (24): 3211–17. https://doi.org/10.1093/bioinformatics/bts611.
Landsberg, Jan H. 2010. “The Effects of Harmful Algal Blooms on Aquatic Organisms” *Reviews in Fisheries Sciences* 1262: 113–390. https://doi.org/10.1080/20026491051695.
Langmead, Ben, and Steven L. Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.
Lee, Eunhee, Doyun Shin, Sung Pil Hyun, Kyung Seok Ko, Hee Sun Moon, Dong Chan Koh, Kyoochul Ha, and Byung Yong Kim. 2017. “Periodic Change in Coastal Microbial Community Structure Associated with Submarine Groundwater Discharge and Tidal Fluctuation.” *Limnology and Oceanography* 62 (2): 437–51. https://doi.org/10.1002/lno.10433.
Li, Weizhong, and Adam Godzik. 2006. “Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences.” *Bioinformatics* 22 (13): 1658–59. https://doi.org/10.1093/bioinformatics/btl158.
Lokmer, Ana, M Anouk Goedknegt, David W Thieltges, Dario Fiorentino, Sven Kuenzel, John F Baines, and K Mathias Wegner. 2016. “Spatial and Temporal Dynamics of Pacific Oyster Hemolymph Microbiota across Multiple Scales” *Front. Microbiol.* 7 (August): 1–18. https://doi.org/10.3389/fmicb.2016.01367.
Lokmer, Ana, and Karl Mathias Wegner. 2015. “Hemolymph Microbiome of Pacific Oysters in Response to Temperature, Temperature Stress and Infection.” *ISME Journal* 9 (3): 670–82. https://doi.org/10.1038/ismej.2014.160.
Louca, S, L Wegener Parfrey, and M Doebeli. 2016. “Decoupling Function and Taxonomy in the Global Ocean Microbiome.” *Science (New York, N.Y.)* 353 (6305): 1272–77. https://doi.org/10.1126/science.aaf4507
Lv, Xiaofei, Bin Ma, Junbao Yu, Scott X. Chang, Jianming Xu, Yunzhao Li, Guangmei Wang, Guangxuan Han, Guan Bo, and Xiaojing Chu. 2016. “Bacterial Community Structure and Function Shift along a Successional Series of Tidal Flats in the Yellow River Delta.” *Scientific Reports* 6 (November): 1–10. https://doi.org/10.1038/srep36550.
Maloy, Aaron P, Susan E Ford, Richard C Karney, and Katherine J Boettcher. 2007. “Roseovarius Crassostreae , the Etiological Agent of Juvenile Oyster Disease (Now to Be Known as Roseovarius

Oyster Disease) in Crassostrea Virginica." *Aquaculture* 269: 71–83. https://doi.org/10.1016/j.aquaculture.2007.04.008.

Marcelino, Vanesa R., Laszlo Irinyi, John Sebastian Eden, Wieland Meyer, Edward C. Holmes, and Tania C. Sorrell. 2019. "Metatranscriptomics as a Tool to Identify Fungal Species and Subspecies in Mixed Communities." *IMA Fungus* 10. https://doi.org/10.1186/s43008-019-0012-8.

Martin, M. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads." *EMBnet* 7 (10): 2803–9. https://doi.org/10.14806/ej.17.1.200

Merou, Nicolas, Cyrielle Lecadet, Stephane Pouvreau and Isabelle Arzul. 2019. "An eDNA/eRNA-based approach to investigate the life cycle of non-cultivable shellfish micro-parasites: the case of Bonamia ostreae, a parasite of the European flat oyster Crassostrea edulis." *Microbial Biotechnology* 0 (0): 1-12. doi:10.1111/1751-7915.13617.

Moran, Mary Ann. 2015. "The Global Ocean Microbiome." *Science* 350 (6266). https://doi.org/10.1126/science.aac8455.

Moran, Mary Ann, Brandon Satinsky, Scott M. Gifford, Haiwei Luo, Adam Rivers, Leong Keat Chan, Jun Meng, et al. 2013. "Sizing up Metatranscriptomics." *ISME Journal* 7 (2): 237–43. https://doi.org/10.1038/ismej.2012.94.

Neves, Andre L.A., Fuyong Li, Bibaswan Ghoshal, Tim McAllister, and Le L. Guan. 2017. "Enhancing the Resolution of Rumen Microbial Classification from Metatranscriptomic Data Using Kraken and Mothur." *Frontiers in Microbiology* 8 (DEC): 1–13. https://doi.org/10.3389/fmicb.2017.02445.

Oksanen, Jari, Guilleome F Blanchet, Roeland Kindt, Pierre Legendre, Peter R Minchin, R B O'Hara, Gavin L Simpson, Peter Solymos, M Henry H Stevens, and Helene Wagner. 2016. "Vegan: Community Ecology Package. R Package Version 2.3-3." 2016. http://cran.r-project.org/package=vegan.

Pachauri, Rajendra K. 2014. "Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change." *Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.

Partensky, F, J Blanchot, and D Vaulot. 1999. "DifferentiaI Distribution and Ecology of Prochlorococcus and Synechococcus in Oceanic Waters : A Review." *Bulletin de l'Institute Oceanographic Monaco*.

R Core Team. 2022. "R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria." 2015. http://www.r-project.org.

Ray, Jessica L, Birte Topper, Shu An, Anna Silyakova, Joachim Spindelbock, Runar Thyrhaug, Michael S DuBow, T Frede Thingstad, Ruth-Anne Sandaa. 2012. "Effect of Increased pCO2 on Bacterial Assemblage Shifts in Response to Glucose Addition in Fram Strait Seawater Mesocosms." *FEMS Microbiology Ecology* https://doi.org/10.1111/j.1574-6941.2012.01443.x.

Renault, Tristan, Anne Lise Bouquet, Julien Thomas Maurice, Coralie Lupo, and Philippe Blachier. 2014. "Ostreid Herpesvirus 1 Infection among Pacific Oyster (Crassostrea Gigas) Spat: Relevance of Water Temperature to Virus Replication and Circulation Prior to the Onset of Mortality." *Applied and Environmental Microbiology* 80 (17): 5419–26. https://doi.org/10.1128/AEM.00484-14.

Ritchie, Matthew E., Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. 2015. "Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies." *Nucleic Acids Research* 43 (7): e47. https://doi.org/10.1093/nar/gkv007.

Roberts, Adam, Harvey Feng, and Lior Pachter. 2013. "Fragment Assignment in the Cloud with EXpress-D." *BMC Bioinformatics* 14 (1): 1–9. https://doi.org/10.1186/1471-2105-14-358.

Robinson, Mark D., and Alicia Oshlack. 2010. "A Scaling Normalization Method for Differential Expression Analysis of RNA-Seq Data." *Genome Biology* 11 (3). https://doi.org/10.1186/gb-2010-11-3-r25

Salisbury, Joseph, Mark Green, Chris Hunt, and Janet Campbell. 2008. "Coastal Acidification by Rivers: A Threat to Shellfish?" *EOS* 89 (50): 513. https://doi.org/10.1029/2008EO500001

Salazar, Guillem, Lucas Paoli, Adriana Alberti, Jaime Huerta-Cepas, Hans-Joachim Ruscheweyh, Miguelangel Cuenca, Christopher M. Field, Luis Pedro Coelho, Corinne Cruaud, Stefan Engelen,

Ann C. Gregory, Karine Labadie, Claudie Marec, Eric Pelletier, Marta Royo-Llonch, Simon Roux, Pablo Sánchez, Hideya Uehara, Ahmed A. Zayed, Georg Zeller, Margaux Carmichael, Céline Dimier, Joannie Ferland, Stefanie Kandels, Marc Picheral, Sergey Pisarev, Julie Poulain, Silvia G. Acinas, Marcel Babin, Peer Bork, Emmanuel Boss, Chris Bowler, Guy Cochrane, Colomban de Vargas, Michael Follows, Gabriel Gorsky, Nigel Grimsley, Lionel Guidi, Pascal Hingamp, Daniele Iudicone, Olivier Jaillon, Stefanie Kandels-Lewis, Lee Karp-Boss, Eric Karsenti, Fabrice Not, Hiroyuki Ogata, Stephane Pesant, Nicole Poulton, Jeroen Raes, Christian Sardet, Sabrina Speich, Lars Stemmann, Matthew B. Sullivan, Shinichi Sunagawa, Patrick Wincker, Silvia G. Acinas, Marcel Babin, Peer Bork, Chris Bowler, Colomban de Vargas, Lionel Guidi, Pascal Hingamp, Daniele Iudicone, Lee Karp-Boss, Eric Karsenti, Hiroyuki Ogata, Stephane Pesant, Sabrina Speich, Matthew B. Sullivan, Patrick Wincker, Shinichi Sunagawa. 2005. "Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome."*Cell* 179 (5): 1068-1083. https://doi.org/10.1016/j.cell.2019.10.014.

Schulze, Angela D., Abayomi O. Alabi, Adele R. Tattersall-Sheldrake, and Kristina M. Miller. 2006. "Bacterial Diversity in a Marine Hatchery: Balance between Pathogenic and Potentially Probiotic Bacterial Strains." *Aquaculture* 256 (1–4): 50–73. https://doi.org/10.1016/j.aquaculture.2006.02.008.

Shakya, Migun, Chien Chi Lo, and Patrick S.G. Chain. 2019. "Advances and Challenges in Metatranscriptomic Analysis." *Frontiers in Genetics* 10 (SEP): 1–10. https://doi.org/10.3389/fgene.2019.00904.

Shi, Yanmei, Gene W. Tyson, John M. Eppley, and Edward F. Delong. 2011. "Integrated Metatranscriptomic and Metagenomic Analyses of Stratified Microbial Assemblages in the Open Ocean." *ISME Journal* 5 (6): 999–1013. https://doi.org/10.1038/ismej.2010.189.

Stevick, Rebecca J., Saebom Sohn, Tejashree H. Modak, David R. Nelson, David C. Rowley, Karin Tammi, Roxanna Smolowitz, Kathryn Markey Lundgren, Anton F. Post, and Marta Gómez-Chiarri. 2019. "Bacterial Community Dynamics in an Oyster Hatchery in Response to Probiotic Treatment." *Frontiers in Microbiology* 10 (MAY): 1–13. https://doi.org/10.3389/fmicb.2019.01060.

Steward, Grieg F, Alexander I Culley, Jaclyn A Mueller, Elisha M Wood-Charlson, Mahdi Belcaid, and Guylaine Poisson. 2013. "Are We Missing Half of the Viruses in the Ocean ?" *The ISME Journal* 7: 672–79. https://doi.org/10.1038/ismej.2012.121.

Sunagawa, Shinichi, Luis Pedro Coelho, Samuel Chaffron, Jens Roat Kultima, Karine Labadie, Guillem Salazar, Bardya Djahanschiri, et al. 2015. "Structure and Function of the Global Ocean Microbiome." *Science* 348 (6237): 1–10. https://doi.org/10.1126/science.1261359.

Suttle, Curtis A. 2007. "Marine Viruses--Major Players in the Global Ecosystem." Nature Reviews. Microbiology 5 (10): 801–12. https://doi.org/10.1038/nrmicro1750.

Tai, V, B Palenik. 2009. "Temporal variation of Synechococcus clades at a coastal Pacific Ocean monitoring site." ISME Journal. 3 (8): 903-915. https://doi.org/10.1038/ismej.2009.35.

Thomson, Paul G, Andrew T Davidson, and Lynsey Maher. 2016. "Increasing CO 2 Changes Community Composition of Pico- and Nano-Sized Protists and Prokaryotes at a Coastal Antarctic Site" *MEPS* 554: 51–69. https://doi.org/10.3354/meps11803.

Trabal, Natalia, José M Mazón-suástegui, Ricardo Vázquez-juárez, Felipe Asencio-valle, Enrique Morales-bojórquez, and Jaime Romero. 2012. "Molecular Analysis of Bacterial Microbiota Associated with Oysters (Crassostrea Gigas and Crassostrea Corteziensis) in Different Growth Phases at Two Cultivation Sites." *Microb Ecol* 64: 555–69. https://doi.org/10.1007/s00248-012-0039-5.

Tsai, A. Y., G. C. Gong, and J. Hung. 2013. "Seasonal Variations of Virus-and Nanoflagellate-Mediated Mortality of Heterotrophic Bacteria in the Coastal Ecosystem of Subtropical Western Pacific." *Biogeosciences* 10 (5): 3055–65. https://doi.org/10.5194/bg-10-3055-2013.

VanEtten, J L, M V Graves, W Boland, and N Delaroque. 2002. "Phycodnaviridae – Large DNA Algal Viruses." *Archives of Virology* 147: 1479–1516. https://doi.org/10.1007/s00705-002-0822-6.

VanEtten, J L, L C Lane, and D D Dunigan. 2010. "DNA Viruses: The Really Big Ones (Giruses)." *Annu Rev Microbiol* 64: 83–99. https://doi.org/10.1146/annurev.micro.112408.134338.DNA.

VanEtten, J L, R H Meints, D Kuczmarski, D E Burbank, and K Lee. 1982. "Viruses of Symbiotic Chlorella-like Algae Isolated from Paramecium Bursaria and Hydra Viridis." *Proceedings of the National Academy of Sciences* 79 (12): 3867–71. https://doi.org/10.1073/pnas.79.12.3867.

Wang, Qing, Ruiwen Cao, Xuanxuan Ning, Liping You, Changkao Mu, Chunlin Wang, Lei Wei, Ming Cong, Huifeng Wu, and Jianmin Zhao. 2016. "Effects of Ocean Acidification on Immune Responses of the Pacific Oyster Crassostrea Gigas." *Fish and Shellfish Immunology* 49: 24–33. https://doi.org/10.1016/j.fsi.2015.12.025.

Wegner, Karl Mathias, Nils Volkenborn, Hannes Peter, and Alexander Eiler. 2013. "Disturbance Induced Decoupling between Host Genetics and Composition of the Associated Microbiome." *BMC Microbiology* 13 (1). https://doi.org/10.1186/1471-2180-13-252.

Weinbauer, Markus G., and Fereidoun Rassoulzadegan. 2004. "Are Viruses Driving Microbial Diversification and Diversity?" *Environmental Microbiology* 6 (1): 1–11. https://doi.org/10.1046/j.1462-2920.2003.00539.x.

Wigington, Charles H, Derek Sonderegger, Corina P D Brussaard, Alison Buchan, Jan F Finke, Jed A Fuhrman, Jay T Lennon, et al. 2016. "Marine Virus and Microbial Cell Abundances." *Nature Microbiology*, no. January: 4–11. https://doi.org/10.1038/nmicrobiol.2015.24.

Worden, Alexandra Z, Jessica K Nolan, and Brian Palenik. 2004. "Assessing the Dynamics and Ecology of Marine Picophytoplankton : The Importance of the Eukaryotic Component." *Limnology and Oceanography* 49 (1): 168–79. https://doi.org/10.4319/lo.2004.49.1.0168

Yu, Shu Xian, Yun Long Pang, Yin Chu Wang, Jia Lin Li, and Song Qin. 2018. "Distribution of Bacterial Communities along the Spatial and Environmental Gradients from Bohai Sea to Northern Yellow Sea." *PeerJ* 2018 (1): 1–21. https://doi.org/10.7717/peerj.4272.

Yung, Cheuk Man, Christopher S. Ward, Katherine M. Davis, Zackary I. Johnson, and Dana E. Hunt. 2016. "Insensitivity of Diverse and Temporally Variable Particle-Associated Microbial Communities to Bulk Seawater Environmental Parameters." *Applied and Environmental Microbiology* 82 (11): 3431–37. https://doi.org/10.1128/AEM.00395-16.

Zhang, Linlin, Li Li, Ximing Guo, Gary W. Litman, Larry J. Dishaw, and Guofan Zhang. 2015. "Massive Expansion and Functional Divergence of Innate Immune Genes in a Protostome." *Scientific Reports* 5: 1–11. https://doi.org/10.1038/srep08693.

Zhao, Xuelin, Hong Yu, Lingfeng Kong, and Qi Li. 2012. "Transcriptomic Responses to Salinity Stress in the Pacific Oyster Crassostrea Gigas." *PLoS ONE* 7 (9). https://doi.org/10.1371/journal.pone.0046244.

Tables

Table 1. Summary statistics of ocean chemistry variables. pCO_2 was the target abiotic variable of the study, but the effect of the other variables was also considered in the context of metatranscriptome profiles. For calcite and aragonite, the saturation states are shown.

	Min	**Max**	**Mean**	**Median**	**Std. dev.**
pH	7.56	8.35	7.8	7.78	0.22
Temperature (ºC)	11.4	16.7	14.4	14	1.5
Salinity (‰)	28	29	28.2	28	0.41
Alkalinity (µmol/kg)	795.1	2634.2	1501.1	1400	465.1
Aragonite (Ω)	0.34	2.72	0.93	0.65	0.69
Calcite (Ω)	0.54	3.15	1.39	1.04	0.86
pCO_2 (ppm)	81.2	1059.9	588.3	606.6	289.9

Table 2. Top ten significant genera of linear models against low, medium, and high pCO_2 concentration and T-test of early vs. late summer. P-values express the model significance and P-adj. is adjusted p-value with Benjamini-Hochberg. Statistic values indicate positive and negative correlations.

Linear model genus vs. pCO_2

R2	P-value	P-adj	Slope	Genus	Order	Phylum	Superkingdom
0.718	<0.001	0.038	0.001	*Methyloceanibacter*	*Rhizobiales*	*Proteobacteria*	*Bacteria*
0.597	<0.001	0.038	-0.002	*Deferribacter*	*Deferribacterales*	*Deferribacteres*	*Bacteria*
0.552	<0.001	0.063	-0.001	*Gossypium*	*Malvales*	*Streptophyta*	*Eukaryota*
0.614	0.001	0.098	-0.002	*Thioalkalivibrio*	*Chromatiales*	*Proteobacteria*	*Bacteria*
0.694	0.001	0.102	-0.002	*Mitrocomella*	*Leptothecata*	*Cnidaria*	*Eukaryota*
0.466	0.001	0.102	-0.002	*Sogarnavirus*	*Picornavirales*	*Pisuviricota*	*Viruses*
0.460	0.001	0.102	-0.001	*Methylovorus*	*Nitrosomonadales*	*Proteobacteria*	*Bacteria*
0.447	0.001	0.108	-0.001	*Methylophilus*	*Nitrosomonadales*	*Proteobacteria*	*Bacteria*
0.462	0.001	0.108	-0.001	*Aeromonas*	*Aeromonadales*	*Proteobacteria*	*Bacteria*
0.470	0.002	0.155	0.001	*Steinhofvirus*	*Caudovirales*	*Uroviricota*	*Viruses*

T-test genus vs. early and late summer

Statistic	P-value	P-adj	Stderr	Genus	Order	Phylum	Superkingdom
7.053	<0.001	0.001	0.121	*Coregonus*	*Salmoniformes*	*Chordata*	*Eukaryota*
6.884	<0.001	0.002	0.118	*Biomphalaria*		*Mollusca*	*Eukaryota*
4.129	0.001	0.030	0.169	*Leucotheavirus*	*Caudovirales*	*Uroviricota*	*Viruses*
3.759	0.004	0.088	0.207	*Nesterenkonia*	*Micrococcales*	*Actinobacteria*	*Bacteria*
3.713	0.002	0.051	0.217	*Glaciecola*	*Alteromonadales*	*Proteobacteria*	*Bacteria*
3.693	0.002	0.062	0.164	*Emcibacter*	*Emcibacterales*	*Proteobacteria*	*Bacteria*
3.614	0.002	0.062	0.125	*Glaesserella*	*Pasteurellales*	*Proteobacteria*	*Bacteria*
3.566	0.004	0.095	0.217	*Nephromyces*	*Nephromycida*	*Apicomplexa*	*Eukaryota*
3.560	0.006	0.110	0.194	*Tannerella*	*Bacteroidales*	*Bacteroidetes*	*Bacteria*
3.221	0.006	0.110	0.112	*Borrelia*	*Spirochaetales*	*Spirochaetes*	*Bacteria*

790 **Table 3**. **Gene Ontology (GO) enrichment for pCO_2 and season.** GO enrichment analysis indicates the
791 effect of pCO_2 and season on several biological processes, most notably a highly significant enrichment for
792 viral processes in the late season. Columns shown are the GO Term, count in gene list, p-value for
793 enrichment test, count in background list, and fold enrichment. Full GO enrichment results are presented in
794 Supplemental Materials.

	GO Term	Count	P-value	Count (background)	Fold Enrich.
Overexpr. in high pCO2	GO:0009236~cobalamin biosynthetic process	5	0.0030	18	7.9
	GO:0006260~DNA replication	17	0.0048	226	2.1
	GO:0019058~viral life cycle	8	0.015	75	3.0
	GO:0039693~viral DNA genome replication	4	0.020	17	6.7
Overexpr. in low pCO2	GO:0051604~protein maturation	6	6.1E-06	30	21.0
	GO:0006950~response to stress	10	0.0032	344	3.1
	GO:0006457~protein folding	5	0.0067	83	6.3
Overexpr. in early season	GO:0003735~structural constituent of ribosome	8	0.0059	134	3.6
Overexpr. in late season	GO:0016032~viral process	53	7.3E-42	91	9.7
	GO:0006260~DNA replication	55	1.8E-20	226	4.1
	GO:0051701~interaction with host	23	5.1E-16	45	8.5
	GO:0046718~viral entry into host cell	17	1.2E-13	27	10.5

795

Figures

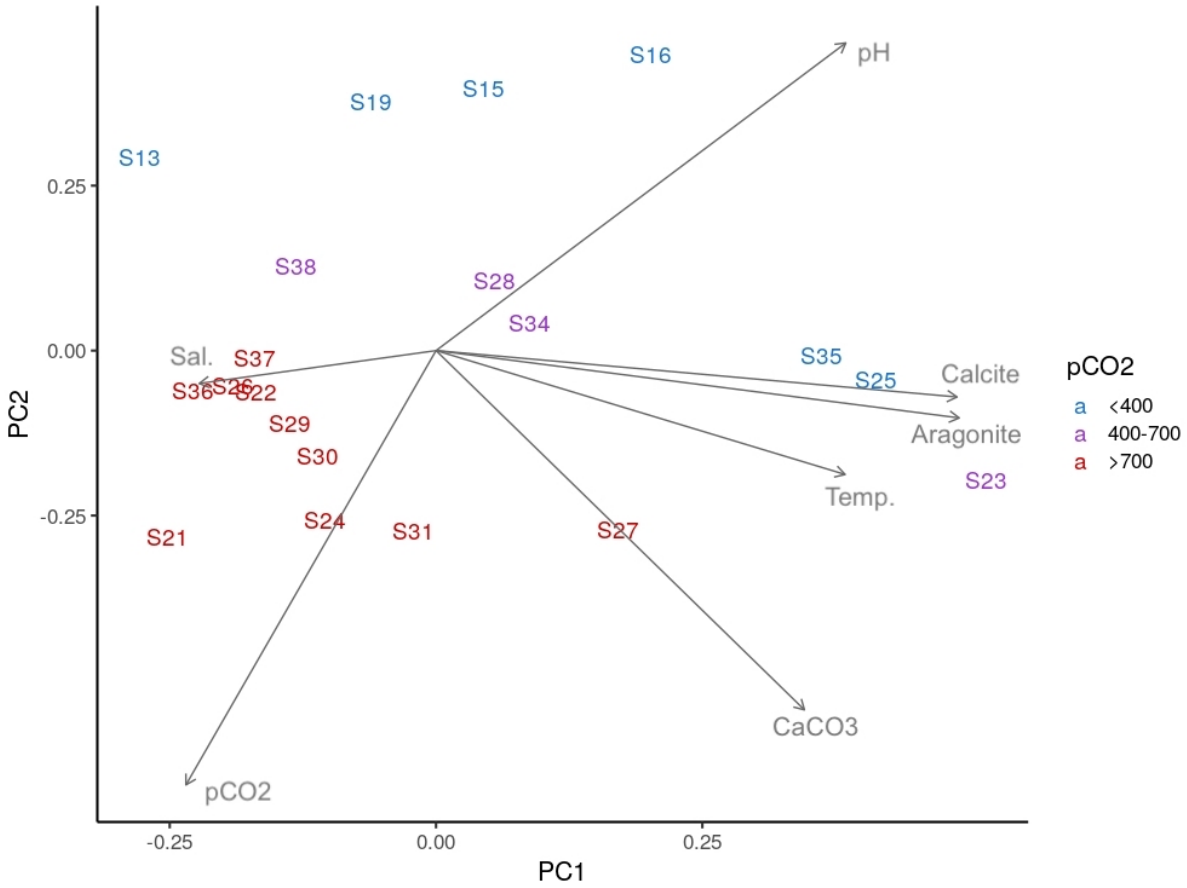


Figure 1. Principal Components Analysis (PCA) of samples based on environmental conditions. Samples are indicated as labels, their corresponding pCO_2 classification is indicated by colour (blue=low, violet=medium, red=high). Arrow direction and length indicate the relative effect and strength of environmental variables. Here the positions of samples are entirely from environmental conditions, not based on taxonomic or gene expression data.

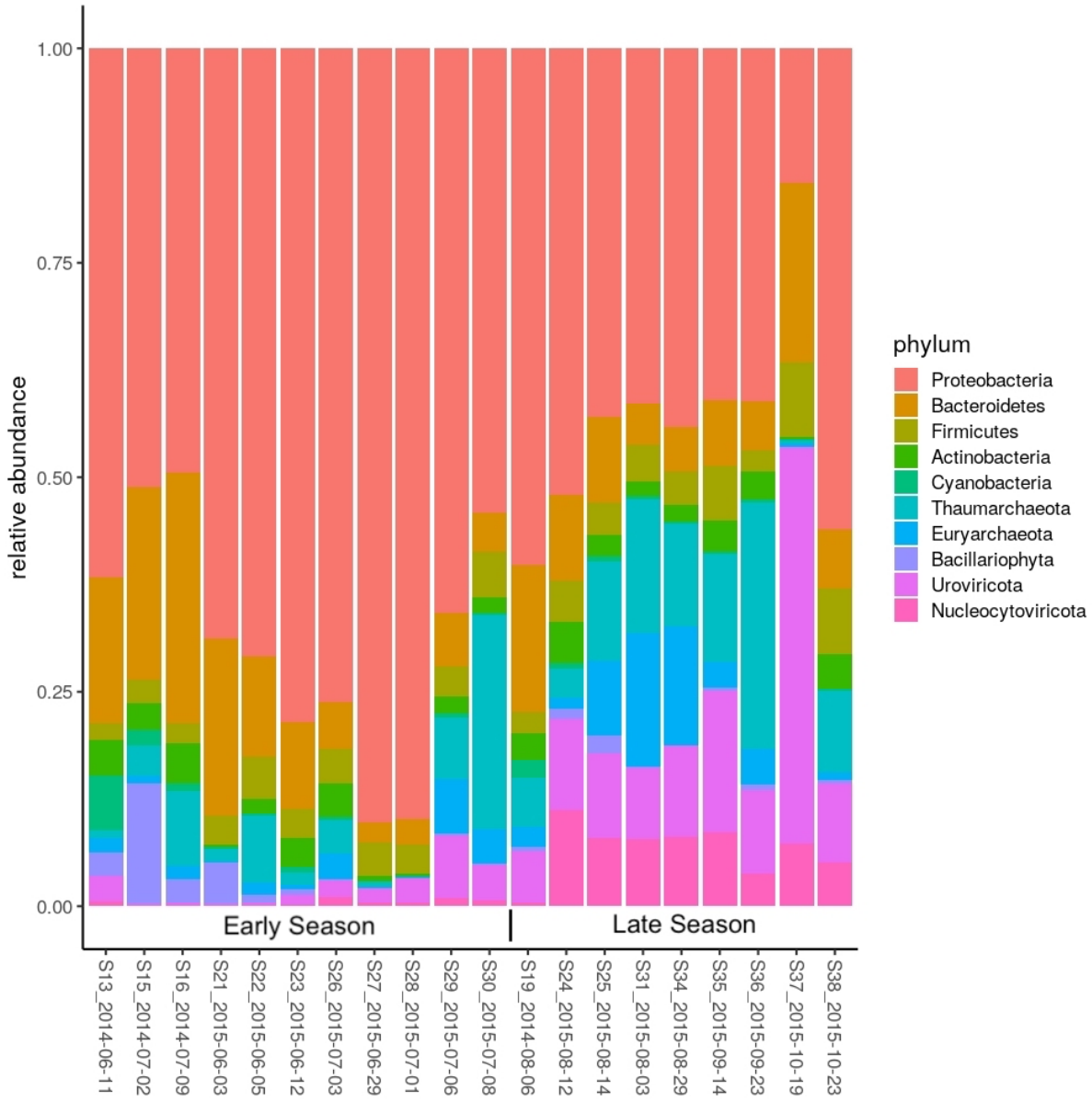


Figure 2. Community composition of the top 10 phyla across samples. Stacked bar-plot of the relative abundance (total number of alignments to taxa) across samples for microbial phyla above 0.5% total abundance, phyla are indicated by color coding in legend, samples are arranged by early season vs. late season.

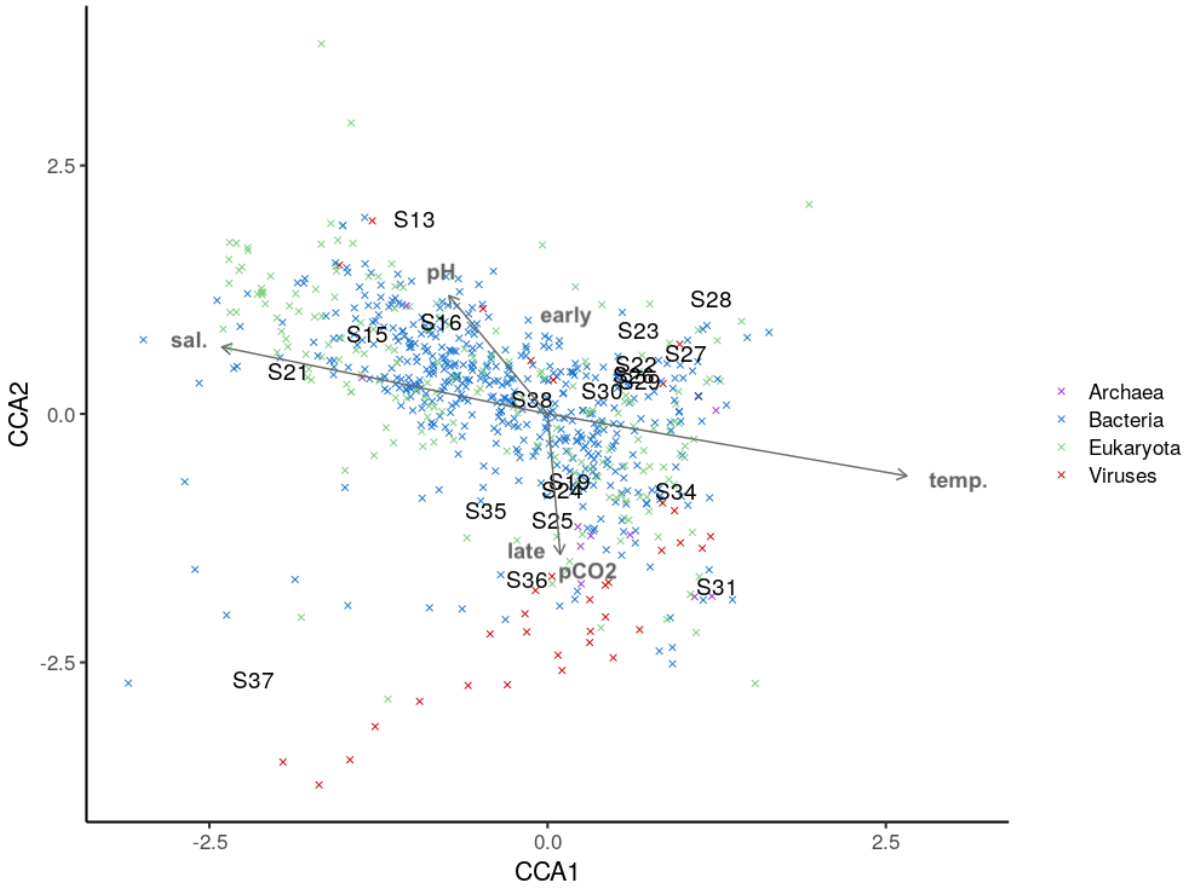


Figure 3. Canonical correspondence analysis (CCA) showing the effect of environmental variables on the variation in relative abundance of genera among samples. Samples are shown as labels and the relative association of genera is shown as crosses. Crosses for genera are color coded by their corresponding superkingdom. The effect of environmental variables on the composition of genera in the samples is indicated by grey arrows, the arrow length corresponds to scaled effect strength. The effect of early and late summer is indicated in grey labels.

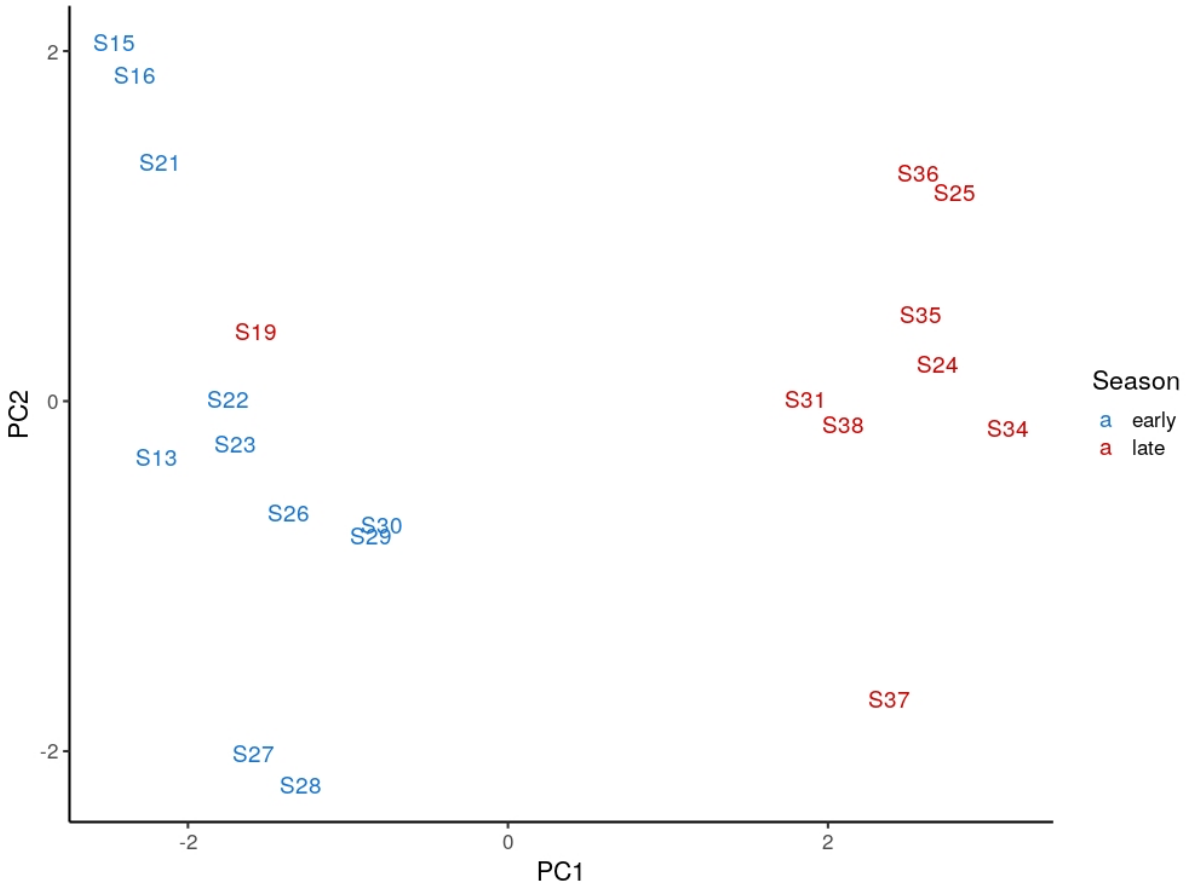
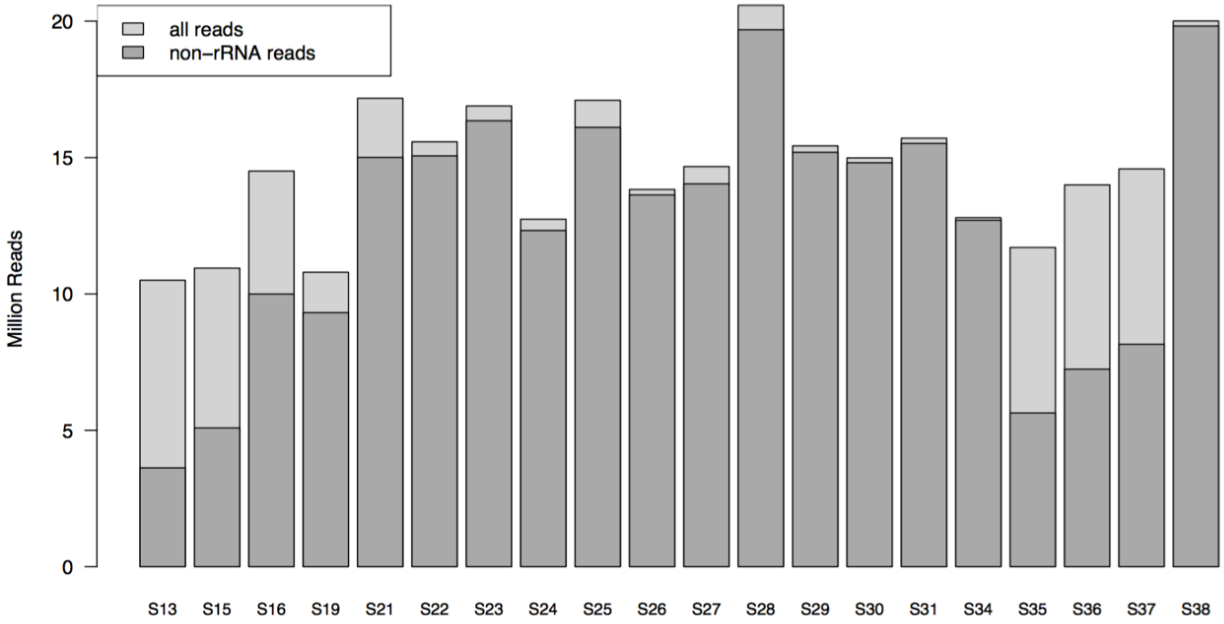


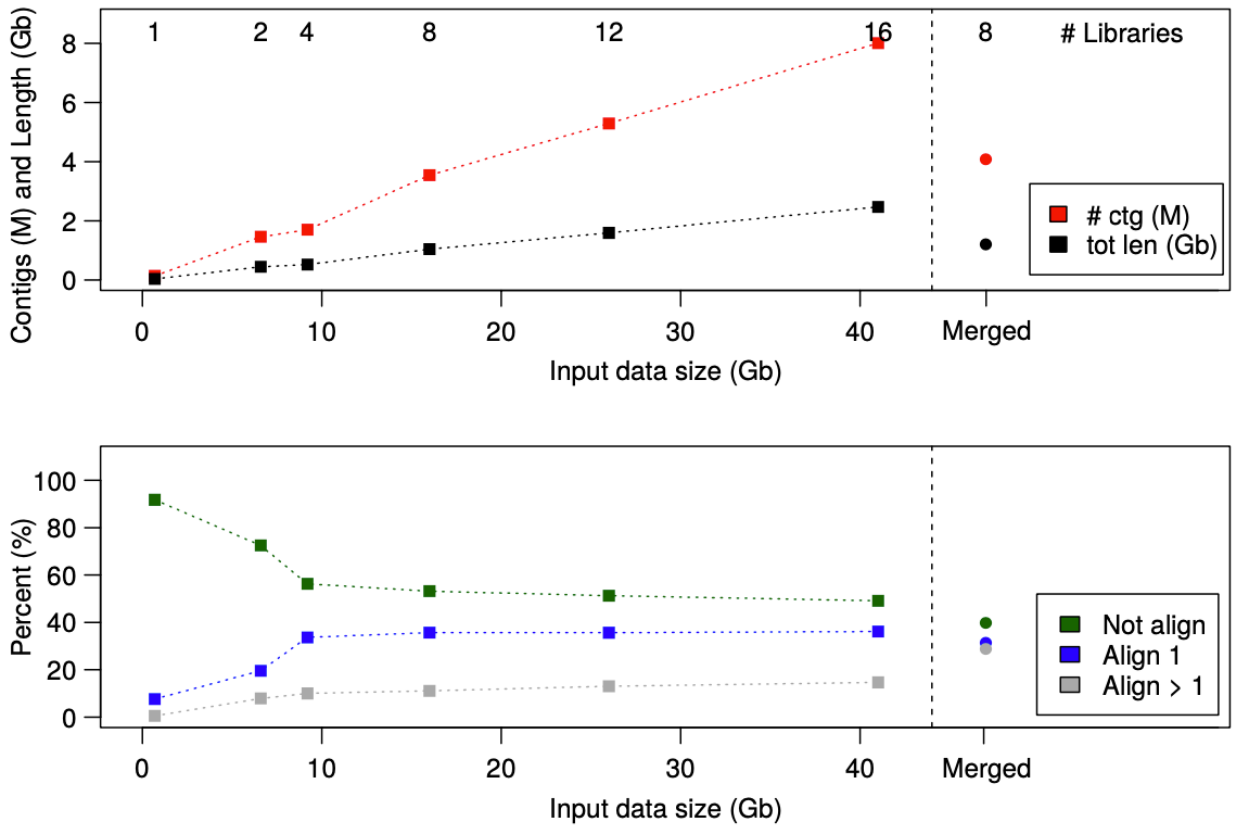
Figure 4. Unsupervised multidimensional scaling (MDS) plot on samples based on gene expression. Dimension 1 explains the most variation, separating the late and early season samples. Samples are labeled by sample number and colors indicate early vs. late season. Samples S13, S15, S16, and S19 are from 2014, and the rest are from 2015. Full details on samples can be viewed in Supplemental File S1.

822 **Supplementary Materials**



823

824 **Supplemental Figure 1. Proportion of non-rRNA and rRNA reads per sample.** Results indicate that
825 for a majority of the samples, the rRNA depletion was successful, enriching the amount of messenger RNA
826 for analysis.



Supplemental Figure 2. Assessment of reference metatranscriptome assemblies. (A) Total contig numbers (red) increase with input data size linearly, and so does total length of output assembly (black). The individual assembly of eight libraries that were subsequently merged together (shown to the right of the hatched line) has a similar number of contigs and length as the eight libraries assembled simultaneously. (B) The percentage of unaligned reads decreases as more libraries are added, initially with a rapid decrease until four libraries are added, and then with a more gentle slope as additional libraries are added. The benefit of increasing from eight to 16 libraries is not as evident in the percentage of reads aligned, suggesting that once four or eight libraries are assembled together, not much benefit is added by increasing the number of libraries in this dataset. Interestingly, the merged assembly has slightly fewer unaligned reads, but double the number of multi-mapping reads (align > 1, grey), indicating a strong amount of redundancy is still present in the merged assembly that is not present in the 16 library simultaneous assembly.

Supplemental Data

Supplemental File S1. Complete environmental and metadata for all samples.

Supplemental File S2. Overview of lineage data assigned to contigs.

Supplemental File S3. Full list of genera with significant linear models to pCO_2 concentrations.

Supplemental File S4. Full list of genera with significant different abundances between early and late summer samples.

Supplemental File S5. Full differential gene expression analysis including genes differentially expressed between pCO2 levels and between seasons, as well as genes found in both comparisons.

Supplemental File S6. Full Gene Ontology analysis including GO enrichment for each differentially expressed gene list for Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). Viral transcripts identified in BP in the season differential analysis are also included.