# 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, there have been concerns that these changes in water quality as well as in microbial composition and activity may be involved in mass mortality events of Pacific oysters (*Crassostrea gigas*) on the East Coast of Vancouver Island, British Columbia. Therefore, improved understanding of the composition and microbial activity of marine waters associated with seasons and abiotic variables may be useful in managing these mortality events. To characterize both compositional and functional changes associated with abiotic factors, here we generate a reference metatranscriptome from the Strait of Georgia over the 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 seasonal changes, as captured in the samples analyzed here. The analysis of both taxonomy and functional gene expression simultaneously in the same samples (i.e., metatranscriptomics) provided a more comprehensive view for monitoring water bodies than either would in isolation.


**Keywords**: metatranscriptome assembly; metatranscriptomics; microbial community, ocean acidification; oyster; pCO2; seasonality; transcriptomics

# 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-generational sequencing, which has the benefit of identifying novel genes and functions not known to be used by the identified taxa (Gilbert et al. 2008). Considering that millions of marine microorganisms and viruses occur within a millilitre of seawater (Wigington et al. 2016; Finke et al. 2017), these assemblages should be considered as a collective in their generated functions rather than being restricted to the functions known to be provided by a single cell or a single taxon (Moran 2015). For detecting an oyster parasite, eRNA was found to detect a species substantially longer than when using eDNA (Merou 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). Due to these reasons, 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).

Globally, the Pacific oyster *Crassostrea gigas* aquaculture industry is dependent both on hatchery-reared commercial oysters as well as naturalized oysters that can be moved onto farms (Sutherland et al. 2020). This is a valuable industry that brings jobs to small communities; in British Columbia, Canada, the Pacific oyster aquaculture industry was recently estimated to produce 14.8 M CAD in revenue per annum (Sun and Hallin 2018). Pacific oyster hatcheries and farms have both experienced large-scale mortality events globally in recent years, including in France (Soletchnik et al. 2005), Australia (Li et al. 2007), California (Burge et al. 2007) and British Columbia (Cassis et al. 2011). Mass mortality events often occur at the spat lifestage (i.e., juveniles attached to substrate; Gomez-Leon et al. 2005; Garcia et al. 2006). Mortalities of adult oysters nearly ready for harvest are also an issue (Soletchnik et al. 2005; Solomieu et al. 2015; Green et al. 2019). The mortalities may be caused by a variety of biotic and abiotic stressors. In many cases, the causes of mass mortality events remain elusive.

In some cases oyster mortalities have been linked to various disease causing agents (Renault et al. 2001; Gomez-Leon et al. 2005; Maloy et al. 2007; Solomieu et al. 2015; Green et al. 2019), including aquatic shellfish pathogens such as *Vibrio* (Gomez-Leon et al. 2005; Paillard et al. 2008), *Roseovarius* (Maloy et al. 2007), and *Mikrocytos* (Carnegie et al. 2003). Oyster Herpes Virus (OsHV-1)-specific

infections have been identified as contributing to massive mortality events in France (Renault et al. 2001; Davison et al. 2009; Renault et al. 2014; Arzul et al. 2017; Martenot et al. 2017). Harmful algae blooms have also been associated with oyster mortalities (Landsberg 2010; Cassis et al. 2011). By the nature of their intertidal habitat and filter feeding lifestyle, oysters are exposed to varying environmental conditions and microbial assemblages (De Schryver and Vadstein 2014; Lokmer et al. 2016; Cho 2019). It has also been recognized that the oyster microbiome is in constant exchange with the pool of exogenous environmental microorganisms (Wegner et al. 2013; Lokmer et al. 2016; Cho A and Finke JF, *unpublished*), which has the potential to introduce pathogens to the oyster. The cause of oyster mortalities may be polymicrobial, where stress leads to an initial infection by an agent such as a virus, and then subsequent bacterial and eukaryote secondary infections can occur. Abiotic correlates to the above microorganisms and environmental indicators associated with their presence may prove to be a useful tool to monitor oyster populations on farms and hatcheries.

Abiotic perturbations such as temperature, $pCO_2$, and salinity, can alter oyster metabolism and growth, can trigger mortality (Zhao et al. 2012; Dickinson et al. 2012; Wang et al. 2016; Kim et al. 2017) and can impact microbial assemblages (Ray and To 2012). Coastal water microbial communities form under widely varying environmental conditions, including ranges of salinity and pH, factors that are impacted by seasons, tides, or even biological activity (Salisbury et al. 2008; Joint et al. 2011; Lv et al. 2016; Lee et al. 2017). Microbes tolerate pH fluctuations under regular conditions and co-vary with water bodies in the short-term (Joint et al. 2011), but adjust in the long term. Microbial, bacterial and protist communities show significant responses to high $pCO_2$ concentrations (Ray and To 2012; Zhang et al. 2015; Thomson et al. 2016). Average $pCO_2$ concentrations have increased from 280 ppm before the industrialization (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). The impact of ocean acidification on overall microbial activity and assemblages is difficult to predict, but the interactions between environmental variables, microbes, and oyster mortalities are important avenues of research.

Here we compare the microbial composition, active genes, and enriched functional categories of differentially expressed genes within samples of intake water at a shellfish hatchery over a range of $pCO_2$ levels, temperatures, and months. We apply comparative metatranscriptomics by first developing a *de novo* metatranscriptome assembly using the samples in the study. We take a taxonomic approach to view the community composition of each sample. In parallel, we quantify the relative expression levels of each transcript among the samples and perform differential gene expression analysis in relation to environmental metadata. Collectively, this work profiles hatchery intake water across 20 dates over a

two-year period. Using environmental metadata, multivariate clustering, differential expression analysis, and functional enrichment analysis, we characterize these different sampling events in relation to date of collection and water environmental parameters.

## Materials and Methods

*Water collection, RNA extraction and environmental variables*

Water samples were collected from the ocean water intake at a Pacific oyster hatchery in Qualicum Bay, central East Coast Vancouver Island, British Columbia over a two year period, with the majority of samples collected in 2015 (Supplemental File S1). On each collection day, six litres of water were taken and filtered through sterile 0.22 µm PVDF filters (Millipore, Burlington, MA). Filters were stored at -80ºC. Temperature was measured by a mercury thermometer, salinity with a refractometer and pH with a glass probe pH meter (Jenco, San Diego, CA). Alkalinity was measured with a HI901 titrator (Hanna Instruments, Smithfield, RI), $pCO_2$ was determined with a LI840A infrared gas analyzer (Li-Cor, Lincoln, NB), and aragonite and calcite were calculated with the CO2SYS.BAS program (https://github.com/jamesorr/CO2SYS-Excel/blob/master/CO2sys_mod.bas). The filters were then used for RNA extraction by the Power Water RNA isolation kit (MoBio, Carlsbad, CA) following manufacturer's instructions, including the alternate lysis step. Output total RNA was depleted for ribosomal RNA and prepped for RNA-seq using the Scriptseq Complete Gold (Epidemiology) kit (Illumina, San Diego, CA). A total of ~50 ng of ribosomal depleted RNA was used as an input to transcriptome libraries. Individual libraries were randomly pooled into groups of four samples using equimolar quantities and sequenced on a MiSeq v3 600 reagent kit (Illumina, San Diego, CA) to generate paired-end 250 bp reads. Environmental variables were analyzed in a principal component analysis (PCA) using the Vegan package (Oksanen et al. 2016) in the R environment (R Core Team, 2022), where missing values were imputed with unit averages.

*Bioinformatics and metatranscriptome assembly*

Raw and quality trimmed sequence data were inspected with FastQC (Andrews 2010) and MultiQC (Ewels et al. 2016). Quality trimming was conducted to remove low quality reads and adapters with Cutadapt (Martin 2011) using flags -q 20 to remove < Q20 data from the 3'-end of the read and -m 50 to remove reads shorter than 50 bp. Results were output as an interleaved fastq. Putative ribosomal RNA (rRNA) reads were removed using SortMeRNA (Kopylova, Noé, and Touzet 2012) using all suggested rRNA databases, and therefore enriching for putative messenger RNA (mRNA) *in silico*.

Two different approaches were taken to assemble the reference metatranscriptome. First, a reference metatranscriptome was assembled using the metatranscriptome-optimized assembler IDBA-tran

(Peng 2012) with a majority of the samples (16 / 20 samples; 191,303,324 reads), not including all samples due to computational constraints. Samples used for this approach comprised an equal number of samples from both ends of the $pCO_2$ range present in the collections. This assembly was conducted with default settings and run on 56 threads. Second, a reference metatranscriptome was generated by first assembling each of the 20 libraries individually using IDBA-tran, and then merging these 20 assemblies into a single assembly using CD-HIT-EST (Li and Godzik 2006). CD-HIT-EST merged contigs at 95% similarity, and dedupe.sh of BBTools (Bushnell, Rood, and Singer 2017) was used to de-duplicate with default parameters (i.e., the 'merged assembly approach'). These assemblies were compared based on the total number of contigs, total length, multi-mapping proportions, and mapping percentages to select the best assembly.

Reads for each sample were aligned against the reference metatranscriptome using Bowtie2 (Langmead and Salzberg 2012) in end-to-end mode allowing for multi-mappings. A maximum of 40 alignments were retained for each read. Alignments were then filtered to remove low quality mappings (i.e., mapq $\geq$ 2). Retained alignments were quantified using eXpress (Roberts et al. 2013). Effective counts from eXpress were output into a table in R, and imported into edgeR (Robinson and Oshlack 2010). Filtering was conducted to only retain contigs against which at least five reads mapped in the sample with the fewest reads (i.e., 3.86 counts per million; CPM), and requiring that the contig was represented at this CPM level or higher in at least five samples. Retained transcripts were normalized for library size using the TMM normalization method of edgeR v.3.28.1.

*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, 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 were conducted in R, where significant regressions were defined by $p < 0.05$.

*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 approximately 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-assembled contigs with the highest number of input samples show fewer multi-mapping reads than did the merged assembly. The percentage of reads aligning a single time increases substantially until four libraries were added, then tapers off to not increase notably with eight, 12, or 16 libraries. Although the addition of more libraries after four libraries did not substantially increase the percentage of reads mapping, but also did not increase redundancy as evaluated by multi-mapping. The 16 library collectively-assembled assembly also has a similar number of contigs and total length to the other assemblies (see Supplemental Results; Figure S2). Therefore, this collectively-assembled assembly was chosen to be used for all downstream functional analyses (i.e., the 'final reference metatranscriptome assembly').

Aligning reads from individual samples against the final reference metatranscriptome assembly resulted in an average alignment rate per sample of 56% (median = 57%; min. = 37%; max = 68%), with an average of 40% of the total alignments per sample with both read pairs aligning concordantly a single time. On average per sample, 12% of reads align concordantly more than once, which may indicate

remaining redundancy in the assembly. Applying low expression filters removed the majority of contigs from the metatranscriptome, retaining 32,866 contigs with CPM ≥ 3.86 in at least 5 of 20 samples.

*Taxonomic analysis*

In order to infer the community compositions of samples, the taxonomic lineages for expressed transcripts were analyzed. Taxonomic assignment was successful for 19,315 of the 32,866 expressed transcripts (59%), but the taxonomic resolution varied among contigs. For example, 89% of the annotated transcripts resolve to at least the phylum level, 80% to class, 79% to genus and only 30% to the species level. Some occurring taxa are non-microbial or suspected false taxonomic assignment, and are left out for this analysis. A total of 708 genera in 104 classes and 58 phyla are annotated, only 10 microbial phyla are present at an abundance level of greater than 0.5% of the total reads. When combined, these 10 phyla account for over 95% of all reads, yet several other phyla are represented in the data as well (Supplemental File S2). These 10 dominant phyla show an abundance of reads assigned to bacteria (~74%), specifically *Proteobacteria*, *Bacteriodetes*, *Firmicutes*, *Actinobacteria* and *Cyanobacteria*. Overall, fewer reads are assigned to archaea (~11%), viruses (~9%) and eukaryotes (~1.6%). *Archaea* are represented by *Thaumarcheota* and *Euryarcheota*. Viruses are largely in the class *Caudovirales* (*Uroviricota*) which includes, for example, bacteriophages and cyanophages, and the *Nucleocytoviricota* that represent the nucleocytoplasmic large DNA viruses (NCLDV). Eukaryotes are in the phylum *Bacillariophyta.* Figure 2 shows the variation in community composition of these ten phyla among samples, *Proteobacteria* and *Bacteroidetes* are clearly dominating in most samples, especially in the early season. *Cyanobacteria* and *Bacillariophyta* are also mostly present in early season samples. In late season there is an increase in relative abundance for the *Thaumarchaeota* and *Euryarchaeota*, but especially the *Uroviricota* and *Nucleocytoviricota* viruses.

The overall variation in community composition among samples was evaluated based on Bray-Curtis similarity in a pairwise distance analysis. When compared at different taxonomic levels, variations in community compositions are congruent, and this is true for comparisons of genus to class (R=0.97), class to phylum (R=0.91) and genus to phylum (R=0.90), all showing significant congruence (Mantel test $p \leq 0.01$). A CCA of taxa composition for the 708 genera and environmental variables produces a significant model (P=0.013), explaining 56% of the total variation in taxon composition through environmental variables (Figure 3). Of that variation the first dimension (CCA1) describes 39% and the second dimension (CCA2) describes 27% of the variation. The relationship between sample composition as indicated by the sample labels and the genera (dots) are shown, taxa are coloured according to the corresponding four kingdoms. Generally, kingdoms are distributed across the CCA, but eukaryotes show some grouping with samples S21, S15 and S13. Viruses show grouping with late summer samples S31,

S34, S36 and S37. The effect and strength of environmental variables on the sample composition is indicated by vectors. Temperature and salinity, and pH and $pCO_2$ describe the predominant dimensions of variation. A stepwise regression determined that temperature ($p=0.001$) and $pCO_2$ ($p=0.028$) are significant environmental variables affecting the composition of samples. Additionally, early vs. late summer is significant ($p=0.023$) in separating samples by composition. Correlating all 708 genera to $pCO_2$, our main environmental variable of interest, revealed 67 genera that have a significant ($p<0.05$) linear correlation of their $\log_{10}$ transformed abundances to $pCO_2$ concentrations (Supplemental File S3). The genera with significant linear correlations are predominantly in the *Proteobacteria*, *Cyanobacteria,* and *Picornavirales* and *Caudovirales*. Similarly, 93 genera show significant variation between the early and late season samples, especially *Caudovirales* and *Algavirales* (Supplemental File S4). The top ten genera for both models are summarized in Table 2.

*Gene expression analysis*

After the low expression filter was applied, in total 22,121 (67%) of the expressed contigs (n = 32,866) were functionally annotated with UniProt identifiers (BLASTx E < 10-5). All gene expression analysis was conducted using both the annotated and unknown transcripts except for functional enrichment analysis, which depends on the UniProt identifier.

Using the filtered expression data, an unsupervised multidimensional scaling plot (MDS plot) was used to group samples by similarity in gene expression (Figure 4). Based on gene expression signatures this plot indicates similarity among samples from a common season, where samples from June-July (early summer) are separated from August-October (late summer/fall), with early season samples (blue) and late season samples (red) separated on PC1 (Figure 4). One sample is an exception, S19, which groups outside of its season. Notably, S19 was the only sample from 2014 that was collected in August or later (collected Aug. 6$^{Th}$, 2014; Supplemental File S1). No effect of year nor technical aspects of sample handling are observed.

Based on the observations of the effect of $pCO_2$ as a key environmental variable on taxonomic composition and a focal variable of the study, a differential expression analysis was conducted using $pCO_2$ separated into either high or medium/low levels, and early summer vs. late summer as binary explanatory variables, as well as their interaction. There is a larger effect of early vs. late summer than of $pCO_2$, where 2,765 transcripts are found differentially expressed based on early vs. late, and 720 based on $pCO_2$. Of these, 45 are differentially expressed in both contrasts (Supplemental File S5). Of the transcripts affected by season, 318 are over-expressed in early summer, and 2447 are over-expressed in late summer/early fall. Of the transcripts affected by $pCO_2$, 553 are over-expressed by high $pCO_2$ and 167 are

under-expressed. There are no transcripts showing a significant interaction effect of sampling period and $pCO_2$.

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

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

# Discussion

*Profiled characteristic conditions indicate short-term environmental fluctuations*

The collected samples reflect typical intake water of oyster hatcheries from standard aquaculture practices on East Coast Vancouver Island (Helm 2004), spanning over several months and two years. The variables were analyzed in a PCA to understand the covariation of environmental variables and how they shape the sampling conditions. The variation in environmental data describes the characteristic water flow and carbon chemistry of intertidal oyster habitats and the region of sampling (Strait of Georgia) (Dickinson et al. 2012; Ianson et al. 2016); salinity and temperature are inversely affected by influxes of cold seawater with relatively high salinity and warmer fresh water with lower salinities. Water with higher $pCO_2$ concentrations are characterized by lower pH values and the combined effect of pH, $pCO_2$, and salinity is reflected in the $CaCO_3$ concentrations, with calcium being one of the salts in seawater. This relationship shows in the sample separation by low, medium and high $pCO_2$ values and the force loadings of environmental variables. Calcite and aragonite are both derivatives of $CaCO_3$, varying based on the $pCO_2$ and the pH of water (Doney 2010), and their influence is thus mediated between these two variables. The clustering of samples observed when considering environmental data highlights the impact of $pCO_2$ and salinity and suggests that variations in environmental conditions observed in the scale of the present study (two years) are driven by short term influxes of fresh or marine waters through tidal cycles and upwelling of deep waters.

*Taxonomic composition varies with abiotic factors and early and late summer*

The microbial community analysis here is based on sequences with taxonomic assignment. However, we expect this to be a representative description, with the possible exception of largely uncharacterized taxa in reference databases. Derived from taxonomic assignments of contigs and their respective expression levels, community similarity among samples proved to be congruent across different taxonomic levels. This supports the approach of using taxonomic assignments down to the genus level. Importantly, it also indicates that shifts in community composition happen at higher taxonomic levels and not just at the genus or species level. Contigs assigned to non-microbial taxa or terrestrial taxa are expected to be either the product of dispersed tissue or cells in the water column or alternatively incorrect taxonomic assignment.

Across samples the microbial communities are clearly dominated by *Proteobacteria*, *Bacteriodetes*, *Firmicutes*, *Actinobacteria* and *Cyanobacteria*. Both eukaryote and archaea microorganisms are notably less represented, and even matched in abundance by viral sequences. The described bacterial phyla are commonly found to be dominant in marine (Sunagawa et al. 2015) and coastal waters (Yung et al. 2016; Yu et al. 2018), and constitute the microbiome of oysters themselves (Trabal et al. 2012; Lokmer et al. 2015; Lokmer et al. 2016; Dubé et al. 2019; Stevick et al. 2019). *Proteobacteria* (*Vibrionales*) and *Bacteriodetes* (*Flavobacteria*) in particular represent common marine microbes and pathogens (Gomez-Leon et al. 2005; Schulze et al. 2006; Paillard et al. 2008; Chen et al. 2017). Observations by Stevick and co-workers (Stevick et al. 2019) found that *Cyanobacteria* were among the dominant phyla in the oyster rearing water communities and *Synechococcus* is a ubiquitous cyanobacterial genus in coastal environments (Partensky et al. 1999; Tai and Palenik 2009). Similarly, the abundant eukaryote *Bacillariophyta* and *Chlorophyta*, both present in the samples taken here, are common phytoplankton in coastal waters (Worden, Nolan, and Palenik 2004; Armbrust 2009).

Matching the microbial community, the dominant *Caudovirales* includes general bacteriophages and specifically cyanophages (Weinbauer and Rassoulzadegan 2004). The present results are thus mirroring the general ubiquity of heterotrophic bacteria, but also indicating their lysis. The *Nucleocytoviricota* include giant viruses commonly infecting protists (Fischer et al. 2010; VanEtten et al. 2010), but also phycodnaviruses infecting eukaryote phytoplankton (e.g. *Chlorophyta*), both being abundantly present in coastal waters (VanEtten et al. 1982, 2002). Taken together, the observation of a low presence of cyanobacteria sequences and *Bacillariophyta* sequences alongside high abundance of plankton viruses and phages in some late season samples, may indicate the lysis of cyanobacteria and algae blooms later in the season. This observation would match phytoplankton bloom patterns described for the northern Strait of Georgia where diatoms (*Bacillariophyta*) and prasinophytes (*Chlorophyta*) are mostly dominant with only periodic cyanobacteria blooms (Del Bel Belluz et al. 2021). Additionally, the

general decrease in *Proteobacteria* in the late season samples is observed alongside increased presence of phage (*Uroviricota*) activity, which also fits with the understanding that bacterial blooms are terminated by viral lysis. As well, viruses show a correlation to late summer samples in the canonical correspondence analysis and the T-test.

The 67 taxa showing significant linear correlations to $pCO_2$, our main environmental variable of interest, can be considered characteristic for changes in $pCO_2$ due to water influx. Several of these taxa, including *Proteobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Firmicutes* and *Deferribacteres* match bacterial taxa with abundances that are associated with tidal cycles and salinity (Lee et al. 2017; Chen et al. 2019). There remains a large portion of unexplained variation in the taxonomic data when all taxa are considered. The CCA shows that only about 50% of the variation among microbial communities could be explained by the combined environmental variables. The effect of environmental variables on the taxonomic community data in the CCA matches their interplay in the PCA based on environmental data (Figure 1). The CCA also confirms that early vs. late summer sampling time has a significant effect on the community composition, although to a lesser effect than observed in the functional gene expression results (see below). We could not identify the abiotic variables responsible for the early vs. late summer effect, which indicates other variables may be involved in community differences are missing (e.g. nutrients or irradiance levels, among others).

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

Overall, the data show that the microbial communities confronting oysters in hatchery intake water vary at high taxonomic levels with season. The communities further vary alongside with environmental variables that are driven by tidal cycles and currents in coastal waters. As demonstrated, metatranscriptome data can be used to monitor the presence of a wide range of microbes and putative pathogens in seawater, expanding on the use of specific microbe probing (e.g. Merou et al. 2019).

*Differential RNA transcript abundance reveals functional variation between early and late summer*

Environmental variables are not only expected to alter microbial community composition, but also to influence the function of cells on the transcriptomic level. Further, a change in community composition does not always indicate a change in functions within the community if one taxon is replaced by a similar

taxon with similar functions. A study comparing taxonomic and functional composition in marine microbial samples revealed a disconnect between taxonomy and functions (Louca et al. 2016). In the functional transcriptome analysis of the present study, early and late summer sampling shows a clear effect on overall transcriptome expression, but environmental variables including $pCO_2$ do not show clear groupings of samples in the gene expression data. Consequently, similar to the community composition results, here functional variation is expected to have been influenced by factors outside of the measured environmental variables (e.g., nutrients and irradiance levels, among others). The lack of an annual pattern in gene expression suggests some stability to this temporal trend over the two years analyzed, however, the grouping of the only August sample (S19) from 2014 not with the late summer but rather with the early summer unlike the August samples in 2015 may indicate a difference in the exact timing of the shift in composition observed in each year.

The early vs. late summer effect is most notably enriched for transcripts involved in phage viral activity. Viral dynamics are an underappreciated component of the global ocean carbon cycle. However, in the North Pacific Ocean it has been observed that viral productivity and abundance is higher in summer (July) than in winter (Jan-Feb) (Gainer et al. 2017), leading these researchers to conclude that seasonality is an important consideration to understand viral dynamics (Jiang and Paul 1994; Tsai et al. 2013). A study within a Korean bay over different seasons found that the number of reads and unique species of viruses identified differed depending on month; the most viral reads were found in March and December and the fewest reads in June and September (Hwang 2017). It is recognized that viral composition changes depend seasonally on a range of factors including temperature, salinity, dissolved oxygen, primary production and nutrient concentrations (Brum et al. 2015; Fuhrman et al. 2015). Generally, viral abundance, specifically for dsDNA viruses, has been found to correlate with nutrient concentration, as well as heterotrophic bacteria abundance (Wigington et al. 2016; Finke et al. 2017). Viruses infecting bacteria (i.e., bacteriophages) are typically considered among the dominant group of viruses in marine environments (Breitbart et al. 2002; Steward et al. 2013). In the Korean bay study, 73% of the viral reads were from bacteriophage, and 26% were from algal viruses, with only 1% involving other viruses (Hwang et al. 2017).

Marine microbial communities are often comprised of a few dominant species and many rare ones, and results from the Tara Oceans Expedition found approximately 37 k bacterial and archaeal species, 100 k protist groups and 5.5 k double-stranded bacterial and archaeal virus populations (Moran 2015). The present study found a high degree of variation in gene expression profiles among samples. Variation could also be expected in the community data as well as the functional expression data. Some apparent randomness could be expected, but overall the communities are expected to be predictable at given times, depths and composition of organic matter (Moran 2015). As is suggested from the

community composition and transcript activity of the present study, previous metatranscriptome studies have also found a substantial increase in viral transcripts after a phytoplankton bloom, presumably due to infected cells in lytic stage (Gilbert et al. 2008).

The effect of $pCO_2$ was also investigated in the differential expression analysis, but this had a lesser effect than that the temporal effect observed. Interestingly, in the present study, overexpression of protein chaperones was observed at low $pCO_2$, which is the opposite to previous observations in high $pCO_2$ mesocosms, where metatranscriptomic responses of overexpressed chaperonin transcripts was observed (Gilbert et al. 2008).

Overall, taxonomic and transcriptomic profiling complemented each other and when combined provided a comprehensive view of the changes observed in this study. This matches the observation by Salazar et al. (2019) that the correlation between taxonomic composition and functional composition is variable. The community composition analysis provided more information on taxa associated with abiotic factors, and the functional analysis highlighted the large effect of viral activity that differed over time during the summer months.

## Conclusions

Increased knowledge on the potential for biotic influences into mortality events in oyster aquaculture, associated with abiotic factors, is an important objective to facilitate monitoring or mitigation of losses. Changing environmental conditions may occur over short time scales through upwelling or changing of currents, or may have large, structured changes that occur temporally. In the present study, microbial communities were derived from metatranscriptome data, which avoided primer bias that occurs in amplicon-based approaches, and captures data from all kingdoms of life. The observed variability among microbial communities was found to be associated with both temperature and $pCO_2$, as well as other changes that occurred between early and late summer. These temporal and abiotic factors appeared to be disconnected and were two different trends occurring in the marine environment. The functional gene expression analysis pointed to a strong difference in viral activity moving into the late season, and a much lesser effect of abiotic factors such as $pCO_2$, temperature and salinity. Together, these analyses provide community composition and functional gene differences associated with abiotic factors and time, and likely captured viral termination of bacterial, cyanobacterial and algal blooms later in the season. Metatranscriptomics allowed the characterization of both community changes as well as gene expression activity changes within these communities simultaneously, providing a comprehensive view of the changes occurring in these water bodies.

## Acknowledgements

This project would not have been possible without the involvement, contributions, and sample provision from the industry partner, Island Scallops (ISL; CEO Robert Saunders). J. Finke and B. Sutherland were supported during this work by a grant from the Gordon and Betty Moore Foundation (GBMF, Grant number 5600) awarded to C. Suttle and K. Miller. This project was supported by an Aquaculture Collaborative Research and Development Program (ACRDP) grant from Fisheries and Oceans, Canada, awarded to K. Miller (Grant number P-14-02-001).

## Data Accessibility

Supplemental files including the reference metatranscriptome (fasta), quantified transcript expression levels (.csv), and annotation for the functional and taxonomic analyses (.txt) are available on FigShare: https://doi.org/10.6084/m9.figshare.18128966.v1

Pipeline for analyzing metatranscriptome data: https://github.com/bensutherland/eRNA_taxo

Pipeline used for annotating transcripts: https://github.com/enormandeau/go_enrichment

Script to extract taxonomic lineages:

https://github.com/janfelix/bioinformatic_tools/blob/master/taxid2lineage.py

# 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.

Arzul, Isabelle, Serge Corbeil, Benjamin Morga, and Tristan Renault. 2017. "Viruses Infecting Marine Molluscs." *Journal of Invertebrate Pathology* 147: 118–35. https://doi.org/10.1016/j.jip.2017.01.009.

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.

Burge, Colleen A, Linda R Judah, Loveday L Conquest, Frederick J Griffin, Daniel P Cheney, Andrew Suhrbier, Brent Vadopalas, Paul G Olin, Tristan Renault, and Carolyn S Friedman. 2007. "Summer Seed Mortality of the Pacific Oyster, Crassostrea Gigas Thunberg Grown in Tomales Bay, California, USA: The Influence of Oyster Stock, Planting Time, Pathogens, and Einvironmental Stressors." *Journal of Shellfish Research* 26 (1): 163–72. http://dx.doi.org/10.2983/0730-8000(2007)26[163:SSMOTP]2.0.CO;2

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

Cassis, David, Christopher M. Pearce, and Maria T. Maldonado. 2011. "Effects of the Environment and Culture Depth on Growth and Mortality in Juvenile Pacific Oysters in the Strait of Georgia, British Columbia." *Aquaculture Environment Interactions* 1 (3): 259–74. https://doi.org/10.3354/aei00025.

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.

Davison, Andrew J, Richard Eberle, Bernhard Ehlers, Gary S Hayward, Duncan J Mcgeoch, Anthony C Minson, Philip E Pellett, et al. 2009. "The Order Herpesvirales." *Virology Division News* 154: 171–77. https://doi.org/10.1007/s00705-008-0278-4.

De Schryver, Peter, and Olav Vadstein. 2014. "Ecological Theory as a Foundation to Control Pathogenic Invasion in Aquaculture." *ISME Journal* 8 (12): 2360–68. https://doi.org/10.1038/ismej.2014.84.

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.

Garcia, Céline, Maeva Robert, Isabelle Arzul, Bruno Chollet, Jean-pierre Joly, Laurence Miossec, Thierry Comtet, and Franck Berthe. 2006. "Viral Gametocytic Hypertrophy of Crassostrea Gigas in France : From Occasional Records to Disease Emergence?" *Disease of Aquatic Organisms* 70 (Farley 1976):

193–99. https://doi.org/10.3354/dao070193

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.

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.

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.

Li, Yan, Jian G Qin, Catherine A Abbott, Xiaoxu Li, and Kirsten Benkendorff. 2007. "Synergistic Impacts of Heat Shock and Spawning on the Physiology and Immune Health of Crassostrea Gigas: An Explanation for Summer Mortality in Pacific Oysters." *American Journal of Physiology*, no. 293: 2353–62. https://doi:10.1152/ajpregu.00463.2007.

669 Lokmer, Ana, M Anouk Goedknegt, David W Thieltges, Dario Fiorentino, Sven Kuenzel, John F Baines,
670 and K Mathias Wegner. 2016. "Spatial and Temporal Dynamics of Pacific Oyster Hemolymph
671 Microbiota across Multiple Scales" *Front. Microbiol.* 7 (August): 1–18.
672 https://doi.org/10.3389/fmicb.2016.01367.
673 Lokmer, Ana, and Karl Mathias Wegner. 2015. "Hemolymph Microbiome of Pacific Oysters in Response
674 to Temperature, Temperature Stress and Infection." *ISME Journal* 9 (3): 670–82.
675 https://doi.org/10.1038/ismej.2014.160.
676 Louca, S, L Wegener Parfrey, and M Doebeli. 2016. "Decoupling Function and Taxonomy in the Global
677 Ocean Microbiome." *Science (New York, N.Y.)* 353 (6305): 1272–77.
678 https://doi.org/10.1126/science.aaf4507
679 Lv, Xiaofei, Bin Ma, Junbao Yu, Scott X. Chang, Jianming Xu, Yunzhao Li, Guangmei Wang,
680 Guangxuan Han, Guan Bo, and Xiaojing Chu. 2016. "Bacterial Community Structure and Function
681 Shift along a Successional Series of Tidal Flats in the Yellow River Delta." *Scientific Reports* 6
682 (November): 1–10. https://doi.org/10.1038/srep36550.
683 Maloy, Aaron P, Susan E Ford, Richard C Karney, and Katherine J Boettcher. 2007. "Roseovarius
684 Crassostreae , the Etiological Agent of Juvenile Oyster Disease (Now to Be Known as Roseovarius
685 Oyster Disease) in Crassostrea Virginica." *Aquaculture* 269: 71–83.
686 https://doi.org/10.1016/j.aquaculture.2007.04.008.
687 Marcelino, Vanesa R., Laszlo Irinyi, John Sebastian Eden, Wieland Meyer, Edward C. Holmes, and Tania
688 C. Sorrell. 2019. "Metatranscriptomics as a Tool to Identify Fungal Species and Subspecies in
689 Mixed Communities." *IMA Fungus* 10. https://doi.org/10.1186/s43008-019-0012-8.
690 Martenot, Claire, Ophélie Gervais, Bruno Chollet, Maryline Houssin, and Tristan Renault. 2017.
691 "Haemocytes Collected from Experimentally Infected Pacific Oysters, Crassostrea Gigas: Detection
692 of Ostreid Herpesvirus 1 DNA, RNA, and Proteins in Relation with Inhibition of Apoptosis." *PLoS*
693 *ONE* 12 (5): 1–19. https://doi.org/10.1371/journal.pone.0177448.
694 Martin, M. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads."
695 *EMBnet* 7 (10): 2803–9. https://doi.org/10.14806/ej.17.1.200
696 Merou, Nicolas, Cyrielle Lecadet, Stephane Pouvreau and Isabelle Arzul. 2019. "An eDNA/eRNA-based
697 approach to investigate the life cycle of non-cultivable shellfish micro-parasites: the case of
698 Bonamia ostreae, a parasite of the European flat oyster Crassostrea edulis." *Microbial*
699 *Biotechnology* 0 (0): 1-12. doi:10.1111/1751-7915.13617.
700 Moran, Mary Ann. 2015. "The Global Ocean Microbiome." *Science* 350 (6266).
701 https://doi.org/10.1126/science.aac8455.
702 Moran, Mary Ann, Brandon Satinsky, Scott M. Gifford, Haiwei Luo, Adam Rivers, Leong Keat Chan,
703 Jun Meng, et al. 2013. "Sizing up Metatranscriptomics." *ISME Journal* 7 (2): 237–43.
704 https://doi.org/10.1038/ismej.2012.94.
705 Neves, Andre L.A., Fuyong Li, Bibaswan Ghoshal, Tim McAllister, and Le L. Guan. 2017. "Enhancing
706 the Resolution of Rumen Microbial Classification from Metatranscriptomic Data Using Kraken and
707 Mothur." *Frontiers in Microbiology* 8 (DEC): 1–13. https://doi.org/10.3389/fmicb.2017.02445.
708 Oksanen, Jari, Guilleome F Blanchet, Roeland Kindt, Pierre Legendre, Peter R Minchin, R B O'Hara,
709 Gavin L Simpson, Peter Solymos, M Henry H Stevens, and Helene Wagner. 2016. "Vegan:
710 Community Ecology Package. R Package Version 2.3-3." 2016.
711 http://cran.r-project.org/package=vegan.
712 Pachauri, Rajendra K. 2014. "Climate Change 2014: Synthesis Report. Contribution of Working Groups
713 I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change."
714 *Contribution of Working Groups I, II and III to the Fifth Assessment Report of the*
715 *Intergovernmental Panel on Climate Change*.
716 Paillard, Christine, Kjetil Korsnes, Patrick Le Chevalier, Christine Le Boulay, Lisbeth Harkestad, Anne
717 Grete Eriksen, Endre Willassen, et al. 2008. "Vibrio Tapetis-like Strain Isolated from Introduced
718 Manila Clams Ruditapes Philippinarum Showing Symptoms of Brown Ring Disease in Norway."
719 *Disease of Aquatic Organisms* 81: 153–61. https://doi.org/10.3354/dao01950.

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.

Renault, Tristan, Cécile Lipart, and Isabelle Arzul. 2001. "A Herpes-like Virus Infects a Non-Ostreid Bivalve Species : Virus Replication in Ruditapes Philippinarum Larvae." *Disease of Aquatic Organisms* 45 (June 1997): 1–7. https://doi.org/10.3354/dao045001

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.

Soletchnik, Patrick, Christophe Lambert, and Katherine Costil. 2005. "Summer Mortality of Crassotsrea Gigas (Thunberg) in Relation to Environmental Rearing Conditions." *Journal of Shellfish Research*

771 24 (1): 197–207. http://dx.doi.org/10.2983/0730-8000(2005)24[197:SMOCGT]2.0.CO;2
772 Solomieu Barbosa, Valerie, Tristan Renault and Marie-Agnes Traves. 2015. "Mass mortality in bivalves
773 and the intricate case of the Pacific Oyster, *Crassostrea gigas*." *Journal of Invertebrate Pathology*
774 131: 2-10. https://doi.org/10.1016/j.jip.2015.07.011.
775 Stevick, Rebecca J., Saebom Sohn, Tejashree H. Modak, David R. Nelson, David C. Rowley, Karin
776 Tammi, Roxanna Smolowitz, Kathryn Markey Lundgren, Anton F. Post, and Marta Gómez-Chiarri.
777 2019. "Bacterial Community Dynamics in an Oyster Hatchery in Response to Probiotic Treatment."
778 *Frontiers in Microbiology* 10 (MAY): 1–13. https://doi.org/10.3389/fmicb.2019.01060.
779 Steward, Grieg F, Alexander I Culley, Jaclyn A Mueller, Elisha M Wood-Charlson, Mahdi Belcaid, and
780 Guylaine Poisson. 2013. "Are We Missing Half of the Viruses in the Ocean ?" *The ISME Journal* 7:
781 672–79. https://doi.org/10.1038/ismej.2012.121.
782 Sun, D, and L Hallin. 2018. "British Columbia's Fisheries and Aquaculture Sector, 2016 Edition."
783 Sunagawa, Shinichi, Luis Pedro Coelho, Samuel Chaffron, Jens Roat Kultima, Karine Labadie, Guillem
784 Salazar, Bardya Djahanschiri, et al. 2015. "Structure and Function of the Global Ocean
785 Microbiome." *Science* 348 (6237): 1–10. https://doi.org/10.1126/science.1261359.
786 Sutherland, B. J. G., Rycroft, C., Ferchaud, A.-L., Saunders, R., Li, L., Liu, S., Chan, A. M., Otto, S. P.,
787 Suttle, C. A., & Miller, K. M. (2020). Relative genomic impacts of translocation history, hatchery
788 practices, and farm selection in Pacific oyster *Crassostrea gigas* throughout the Northern
789 Hemisphere. *Evolutionary Applications, 13*(6), 1380-1399. doi:10.1111/eva.12965
790 Tai, V, B Palenik. 2009. "Temporal variation of Synechococcus clades at a coastal Pacific Ocean
791 monitoring site." ISME Journal. 3 (8): 903-915. https://doi.org/10.1038/ismej.2009.35.
792 Thomson, Paul G, Andrew T Davidson, and Lynsey Maher. 2016. "Increasing CO 2 Changes Community
793 Composition of Pico- and Nano-Sized Protists and Prokaryotes at a Coastal Antarctic Site" *MEPS*
794 554: 51–69. https://doi.org/10.3354/meps11803.
795 Trabal, Natalia, José M Mazón-suástegui, Ricardo Vázquez-juárez, Felipe Asencio-valle, Enrique
796 Morales-bojórquez, and Jaime Romero. 2012. "Molecular Analysis of Bacterial Microbiota
797 Associated with Oysters (Crassostrea Gigas and Crassostrea Corteziensis) in Different Growth
798 Phases at Two Cultivation Sites." *Microb Ecol* 64: 555–69. https://doi.org/10.1007/s00248-012-
799 0039-5.
800 Tsai, A. Y., G. C. Gong, and J. Hung. 2013. "Seasonal Variations of Virus-and Nanoflagellate-Mediated
801 Mortality of Heterotrophic Bacteria in the Coastal Ecosystem of Subtropical Western Pacific."
802 *Biogeosciences* 10 (5): 3055–65. https://doi.org/10.5194/bg-10-3055-2013.
803 VanEtten, J L, M V Graves, W Boland, and N Delaroque. 2002. "Phycodnaviridae – Large DNA Algal
804 Viruses." *Archives of Virology* 147: 1479–1516. https://doi.org/10.1007/s00705-002-0822-6.
805 VanEtten, J L, L C Lane, and D D Dunigan. 2010. "DNA Viruses: The Really Big Ones (Giruses)." *Annu*
806 *Rev Microbiol* 64: 83–99. https://doi.org/10.1146/annurev.micro.112408.134338.DNA.
807 VanEtten, J L, R H Meints, D Kuczmarski, D E Burbank, and K Lee. 1982. "Viruses of Symbiotic
808 Chlorella-like Algae Isolated from Paramecium Bursaria and Hydra Viridis." *Proceedings of the*
809 *National Academy of Sciences* 79 (12): 3867–71. https://doi.org/10.1073/pnas.79.12.3867.
810 Wang, Qing, Ruiwen Cao, Xuanxuan Ning, Liping You, Changkao Mu, Chunlin Wang, Lei Wei, Ming
811 Cong, Huifeng Wu, and Jianmin Zhao. 2016. "Effects of Ocean Acidification on Immune Responses
812 of the Pacific Oyster Crassostrea Gigas." *Fish and Shellfish Immunology* 49: 24–33.
813 https://doi.org/10.1016/j.fsi.2015.12.025.
814 Wegner, Karl Mathias, Nils Volkenborn, Hannes Peter, and Alexander Eiler. 2013. "Disturbance Induced
815 Decoupling between Host Genetics and Composition of the Associated Microbiome." *BMC*
816 *Microbiology* 13 (1). https://doi.org/10.1186/1471-2180-13-252.
817 Weinbauer, Markus G., and Fereidoun Rassoulzadegan. 2004. "Are Viruses Driving Microbial
818 Diversification and Diversity?" *Environmental Microbiology* 6 (1): 1–11.
819 https://doi.org/10.1046/j.1462-2920.2003.00539.x.
820 Wigington, Charles H, Derek Sonderegger, Corina P D Brussaard, Alison Buchan, Jan F Finke, Jed A
821 Fuhrman, Jay T Lennon, et al. 2016. "Marine Virus and Microbial Cell Abundances." *Nature*

822 *Microbiology*, no. January: 4–11. https://doi.org/10.1038/nmicrobiol.2015.24.
823 Worden, Alexandra Z, Jessica K Nolan, and Brian Palenik. 2004. "Assessing the Dynamics and Ecology
824 of Marine Picophytoplankton : The Importance of the Eukaryotic Component." *Limnology and*
825 *Oceanography* 49 (1): 168–79. https://doi.org/10.4319/lo.2004.49.1.0168
826 Yu, Shu Xian, Yun Long Pang, Yin Chu Wang, Jia Lin Li, and Song Qin. 2018. "Distribution of Bacterial
827 Communities along the Spatial and Environmental Gradients from Bohai Sea to Northern Yellow
828 Sea." *PeerJ* 2018 (1): 1–21. https://doi.org/10.7717/peerj.4272.
829 Yung, Cheuk Man, Christopher S. Ward, Katherine M. Davis, Zackary I. Johnson, and Dana E. Hunt.
830 2016. "Insensitivity of Diverse and Temporally Variable Particle-Associated Microbial
831 Communities to Bulk Seawater Environmental Parameters." *Applied and Environmental*
832 *Microbiology* 82 (11): 3431–37. https://doi.org/10.1128/AEM.00395-16.
833 Zhang, Linlin, Li Li, Ximing Guo, Gary W. Litman, Larry J. Dishaw, and Guofan Zhang. 2015. "Massive
834 Expansion and Functional Divergence of Innate Immune Genes in a Protostome." *Scientific Reports*
835 5: 1–11. https://doi.org/10.1038/srep08693.
836 Zhao, Xuelin, Hong Yu, Lingfeng Kong, and Qi Li. 2012. "Transcriptomic Responses to Salinity Stress in
837 the Pacific Oyster Crassostrea Gigas." *PLoS ONE* 7 (9).
838 https://doi.org/10.1371/journal.pone.0046244.
839

# 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 |

846 **Table 2. Top ten significant genera of linear models against low, medium, and high $pCO_2$**
847 **concentration and T-test of early vs. late summer.** P-values express the model significance, slope and
848 statistic values indicate positive and negative correlations.

**Linear model genus vs. $pCO_2$**

| R2 | P-value | Intercept | Slope | Genus | Order | Phylum | Superkingdom |
|---|---|---|---|---|---|---|---|
| 0.718 | <0.001 | 0.582 | 0.001 | *Methyloceanibacter* | *Rhizobiales* | *Proteobacteria* | *Bacteria* |
| 0.597 | <0.001 | 2.157 | -0.002 | *Deferribacter* | *Deferribacterales* | *Deferribacteres* | *Bacteria* |
| 0.552 | <0.001 | 2.229 | -0.001 | *Gossypium* | *Malvales* | *Streptophyta* | *Eukaryota* |
| 0.614 | 0.001 | 1.956 | -0.002 | *Thioalkalivibrio* | *Chromatiales* | *Proteobacteria* | *Bacteria* |
| 0.694 | 0.001 | 2.902 | -0.002 | *Mitrocomella* | *Leptothecata* | *Cnidaria* | *Eukaryota* |
| 0.466 | 0.001 | 2.234 | -0.002 | *Sogarnavirus* | *Picornavirales* | *Pisuviricota* | *Viruses* |
| 0.460 | 0.001 | 2.320 | -0.001 | *Methylovorus* | *Nitrosomonadales* | *Proteobacteria* | *Bacteria* |
| 0.447 | 0.001 | 2.644 | -0.001 | *Methylophilus* | *Nitrosomonadales* | *Proteobacteria* | *Bacteria* |
| 0.462 | 0.001 | 1.712 | -0.001 | *Aeromonas* | *Aeromonadales* | *Proteobacteria* | *Bacteria* |
| 0.470 | 0.002 | 0.525 | 0.001 | *Steinhofvirus* | *Caudovirales* | *Uroviricota* | *Viruses* |

**T-test genus vs. early and late summer**

| Statistic | P-value | Parameter | Stderr | Genus | Order | Phylum | Superkingdom |
|---|---|---|---|---|---|---|---|
| 7.053 | 0.000 | 13.653 | 0.121 | *Coregonus* | *Salmoniformes* | *Chordata* | *Eukaryota* |
| 6.884 | 0.000 | 11.777 | 0.118 | *Biomphalaria* | | *Mollusca* | *Eukaryota* |
| 4.129 | 0.001 | 15.288 | 0.169 | *Leucotheavirus* | *Caudovirales* | *Uroviricota* | *Viruses* |
| 3.759 | 0.004 | 9.647 | 0.207 | *Nesterenkonia* | *Micrococcales* | *Actinobacteria* | *Bacteria* |
| 3.713 | 0.002 | 16.922 | 0.217 | *Glaciecola* | *Alteromonadales* | *Proteobacteria* | *Bacteria* |
| 3.693 | 0.002 | 13.866 | 0.164 | *Emcibacter* | *Emcibacterales* | *Proteobacteria* | *Bacteria* |
| 3.614 | 0.002 | 15.826 | 0.125 | *Glaesserella* | *Pasteurellales* | *Proteobacteria* | *Bacteria* |
| 3.566 | 0.004 | 10.998 | 0.217 | *Nephromyces* | *Nephromycida* | *Apicomplexa* | *Eukaryota* |
| 3.560 | 0.006 | 9.414 | 0.194 | *Tannerella* | *Bacteroidales* | *Bacteroidetes* | *Bacteria* |
| 3.221 | 0.006 | 14.866 | 0.112 | *Borrelia* | *Spirochaetales* | *Spirochaetes* | *Bacteria* |

849

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

# Figures

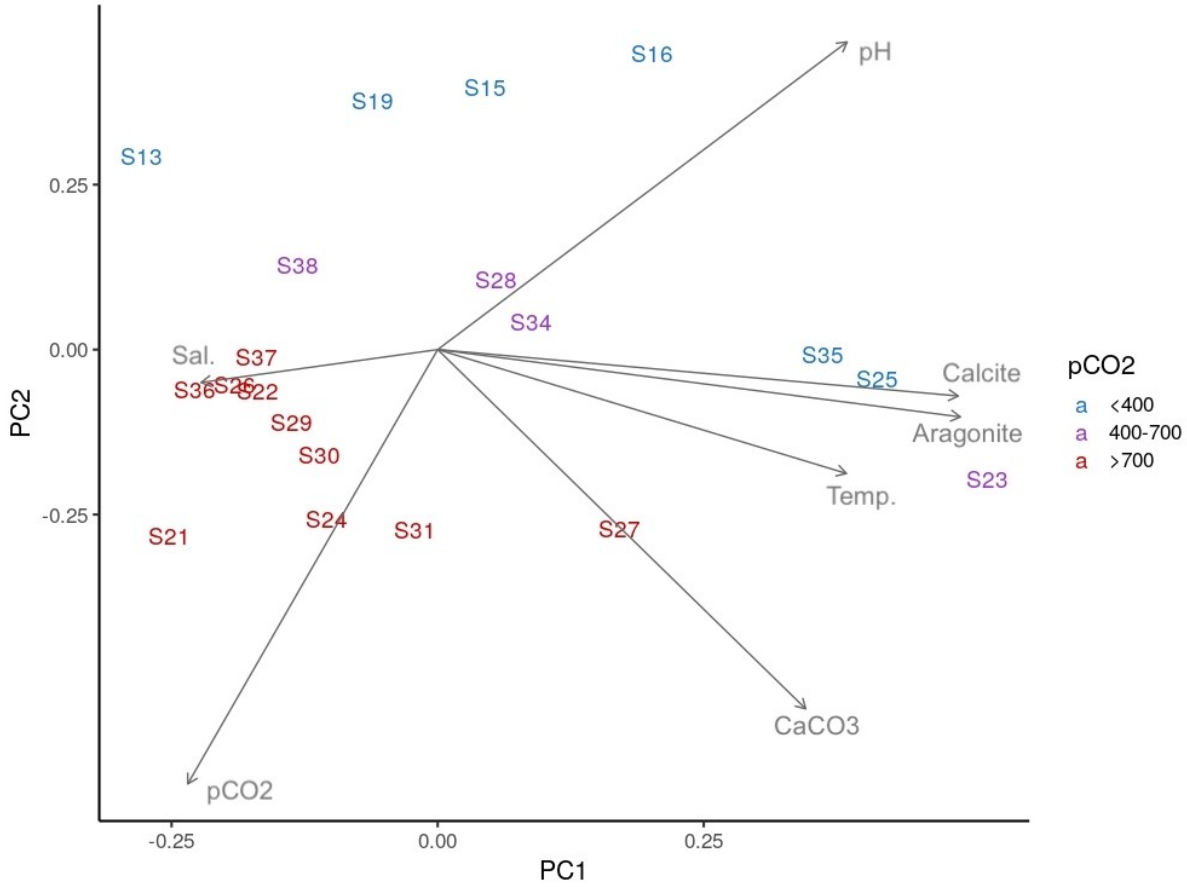



**Figure 1. Principal Components Analysis (PCA) of samples based on environmental conditions.** Samples are indicated as labels, their corresponding $pCO_2$ classification is indicated by colour (blue=low, violet=medium, red=high). Arrow direction and length indicate the relative effect and strength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

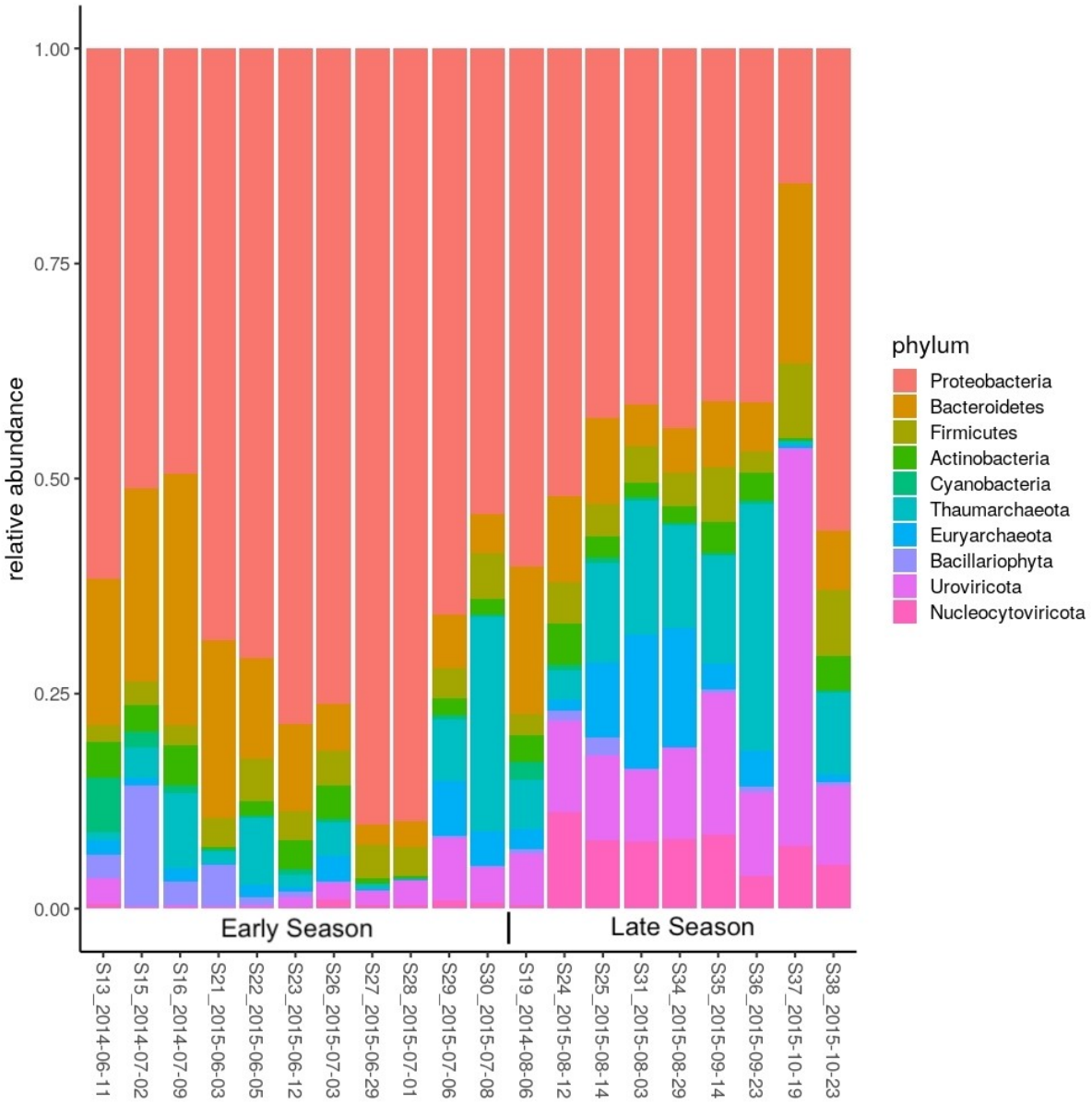



**Figure 2. Community composition of the top 10 phyla across samples.** Stacked bar-plot of the relative abundance (total number of alignments to taxa) across samples for microbial phyla above 0.5% total abundance, phyla are indicated by color coding in legend, samples are arranged by early season vs.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

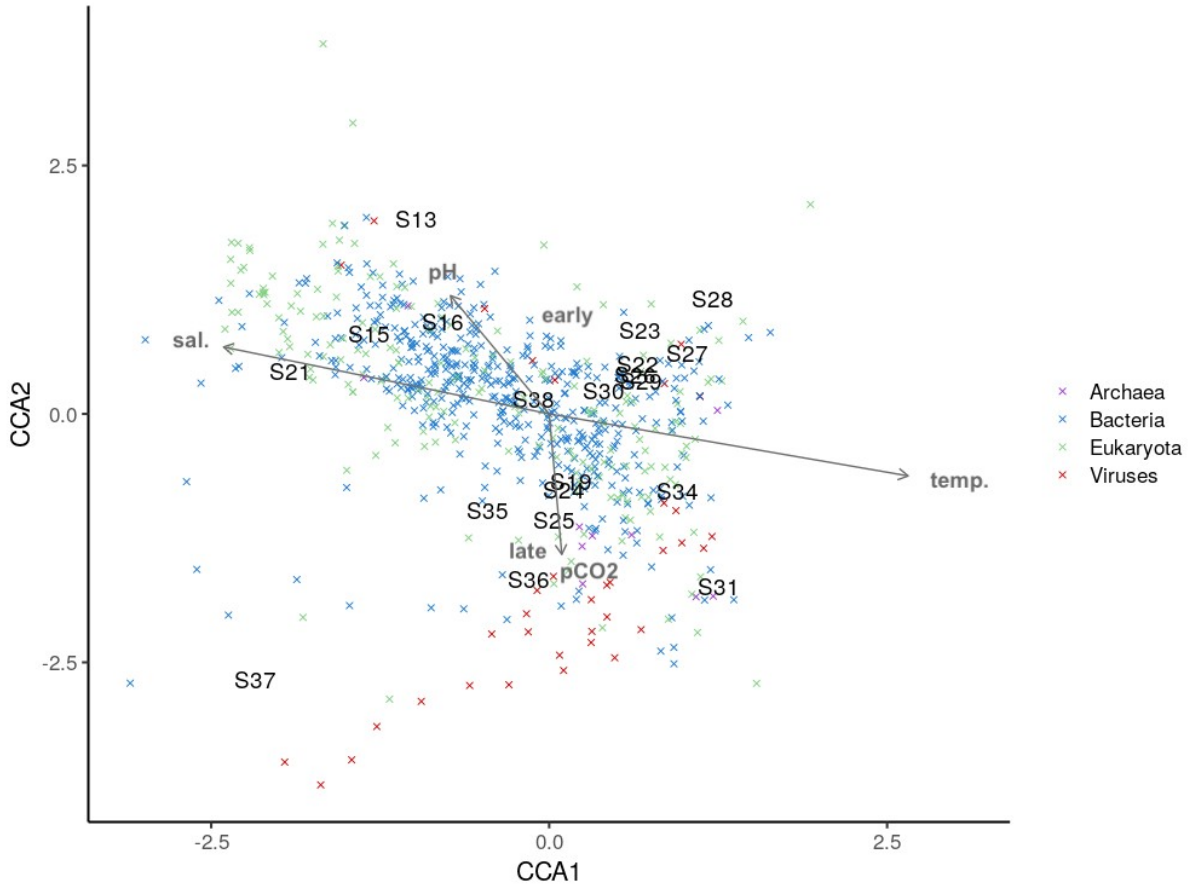
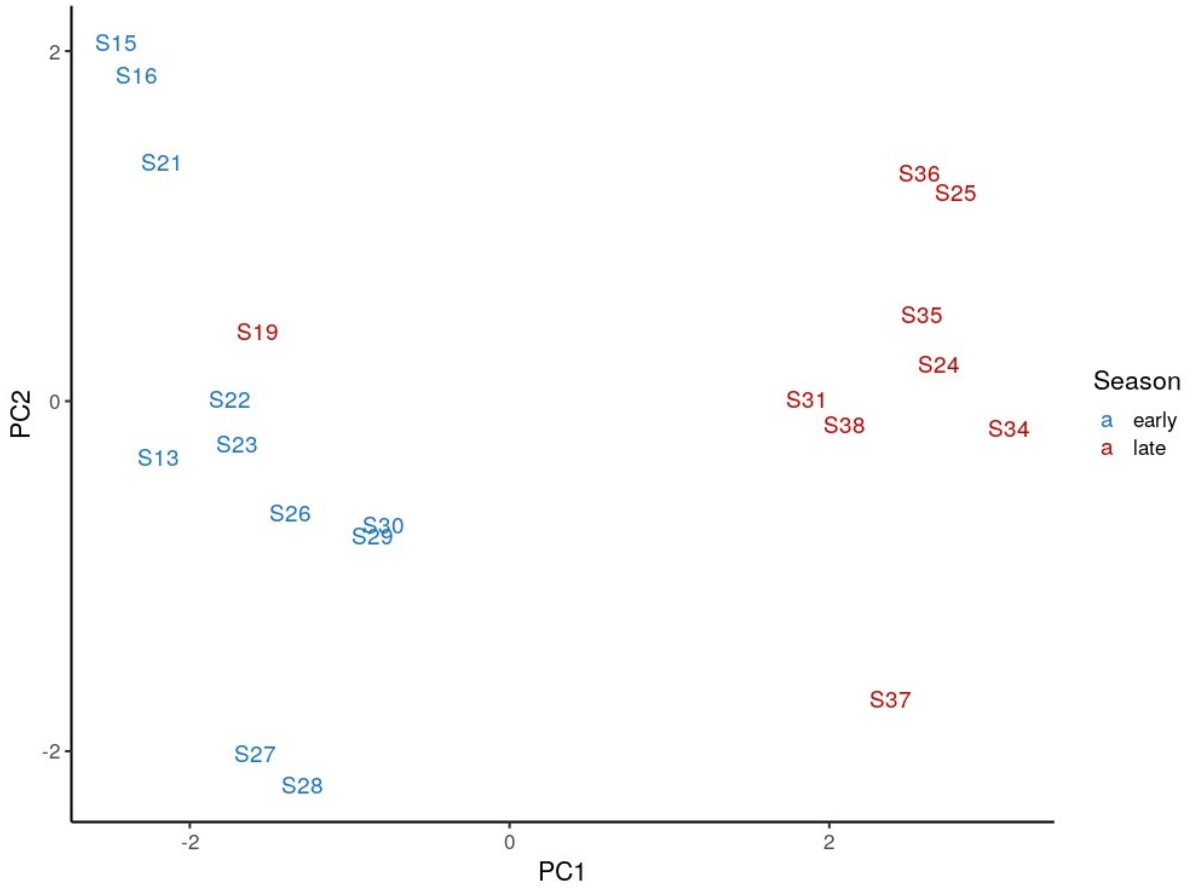



**Figure 3. Canonical correspondence analysis (CCA) showing the effect of environmental variables on the variation in relative abundance of genera among samples.** Samples are shown as labels and the relative association of genera is shown as crosses. Crosses for genera are color coded by their corresponding 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.

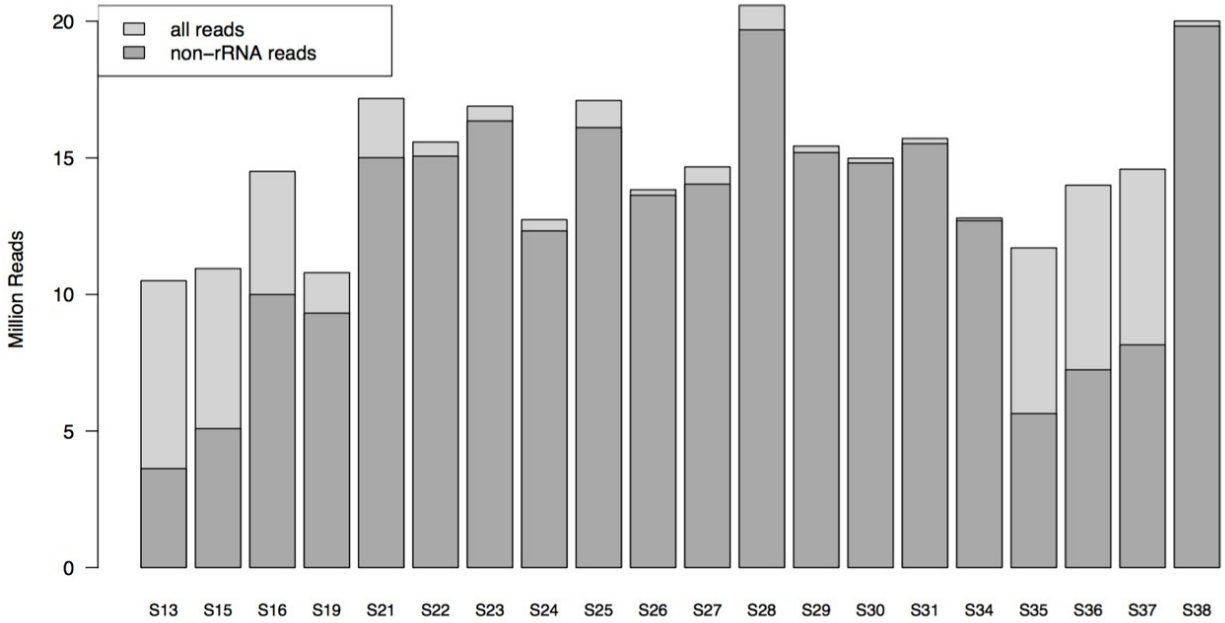
876



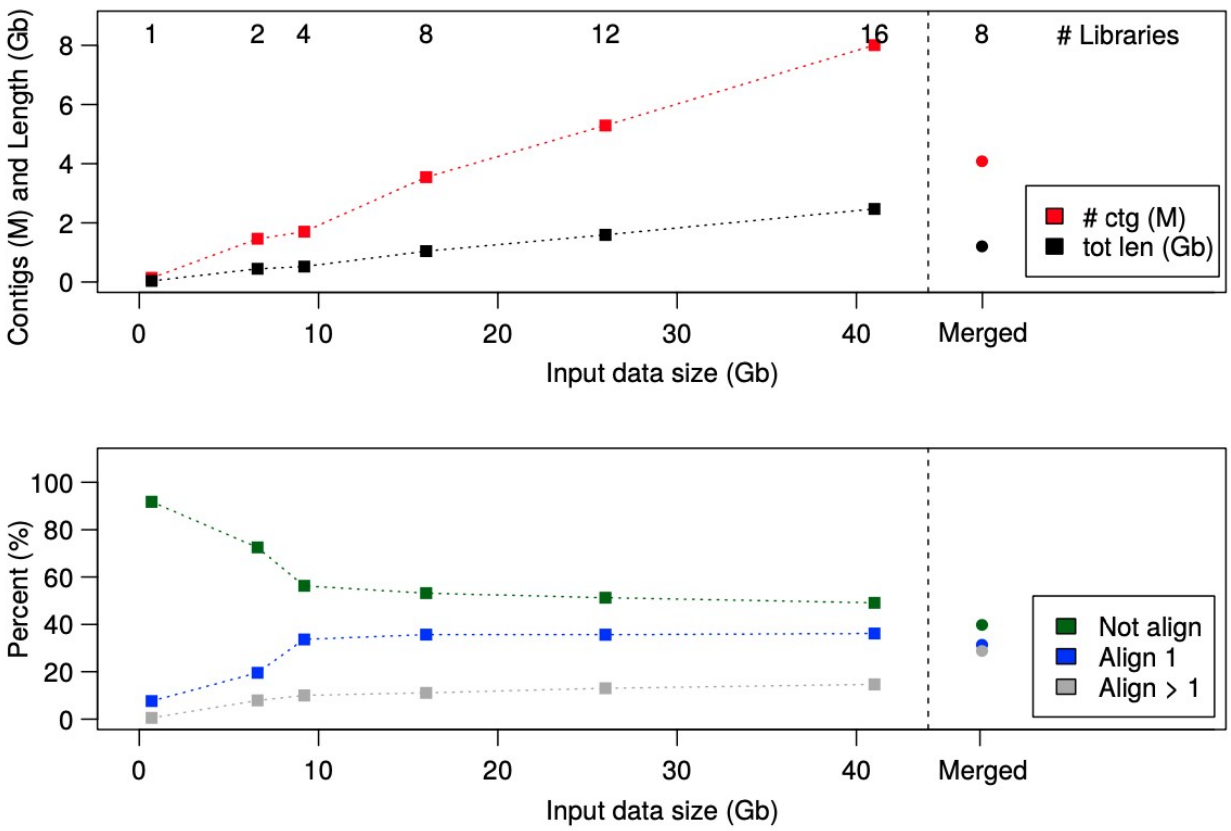

877

878 **Figure 4. Unsupervised multidimensional scaling (MDS) plot on samples based on gene expression.**
879 Dimension 1 explains the most variation, separating the late and early season samples. Samples are
880 labeled by sample number and the $pCO_2$ level measured during the sampling. Samples S13, S15, S16, and
881 S19 are from 2014, and the rest are from 2015. Full details on samples can be viewed in Supplemental
882 File S1.

# Supplementary Materials




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