AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology

Abundance and Biogeography of Picoprasinophyte Ecotypes and Other Phytoplankton in the Eastern North Pacific Ocean

Melinda P. Simmons,[a,b] Sebastian Sudek,[a] Adam Monier,[a,c] Alexander J. Limardo,[a,b] Valeria Jimenez,[a,b] Christopher R. Perle,[d] Virginia A. Elrod,[a] J. Timothy Pennington,[a] Alexandra Z. Worden[a,b]

Monterey Bay Aquarium Research Institute, Moss Landing, California, USA[a]; Department of Ocean Sciences, University of California Santa Cruz, Santa Cruz, California, USA[b]; Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom[c]; Department of Math and Sciences, Florida State College at Jacksonville, Jacksonville, Florida, USA[d]

Eukaryotic algae within the picoplankton size class (≤2 μm in diameter) are important marine primary producers, but their spatial and ecological distributions are not well characterized. Here, we studied three picoeukaryotic prasinophyte genera and their cyanobacterial counterparts, *Prochlorococcus* and *Synechococcus*, during two cruises along a North Pacific transect characterized by different ecological regimes. Picoeukaryotes and *Synechococcus* reached maximum abundances of 1.44×10^5 and 3.37×10^5 cells · ml^{-1}, respectively, in mesotrophic waters, while *Prochlorococcus* reached 1.95×10^5 cells · ml^{-1} in the oligotrophic ocean. Of the picoeukaryotes, *Bathycoccus* was present at all stations in both cruises, reaching 21,368 ± 327 18S rRNA gene copies · ml^{-1}. *Micromonas* and *Ostreococcus* clade OI were detected only in mesotrophic and coastal waters and *Ostreococcus* clade OII only in the oligotrophic ocean. To resolve proposed *Bathycoccus* ecotypes, we established genetic distances for 1,104 marker genes using targeted metagenomes and the *Bathycoccus prasinos* genome. The analysis was anchored in comparative genome analysis of three *Ostreococcus* species for which physiological and environmental data are available to facilitate data interpretation. We established that two *Bathycoccus* ecotypes exist, named here BI (represented by coastal isolate *Bathycoccus prasinos*) and BII. These share 82% ± 6% nucleotide identity across homologs, while the *Ostreococcus* spp. share 75% ± 8%. We developed and applied an analysis of ecomarkers to metatranscriptomes sequenced here and published -omics data from the same region. The results indicated that the *Bathycoccus* ecotypes cooccur more often than *Ostreococcus* clades OI and OII do. Exploratory analyses of relative transcript abundances suggest that *Bathycoccus* NRT2.1 and AMT2.2 are high-affinity NO_3^- and low-affinity NH_4^+ transporters, respectively, with close homologs in multiple picoprasinophytes. Additionally, in the open ocean, where dissolved iron concentrations were low (0.08 nM), there appeared to be a shift to the use of nickel superoxide dismutases (SODs) from Mn/Fe/Cu SODs closer inshore. Our study documents the distribution of picophytoplankton along a North Pacific ecological gradient and offers new concepts and techniques for investigating their biogeography.

Marine phytoplankton are responsible for roughly half of global net primary production (1), and photosynthetic picoeukaryotes (cells ≤2 μm in diameter) have been shown to play important roles in primary production (2–6). Although these small cells come from diverse lineages, their distinguishing morphological features are limited and difficult to visualize (7, 8). Therefore, documentation of basic ecological parameters, such as abundance and distribution of different taxa, is still limited.

Green algae that belong to the prasinophytes include many marine picoeukaryotes. Class II prasinophytes (i.e., *Mamiellophyceae*) are the most widespread (7, 8). For example, a circumglobal study on diversity of marine eukaryotes found class II prasinophytes in all 45 samples (9). Molecular phylogenies and field studies have revealed multiple clades or ecotypes within three major class II genera, *Bathycoccus*, *Micromonas*, and *Ostreococcus* (10–12). Genome analyses demonstrate extensive diversity among *Micromonas* clades that reflect species level divisions (12, 13) and lesser but still notable divergence between *Ostreococcus* species (14). The *Bathycoccus* genus is thought to consist of a single species (*Bathycoccus prasinos*), and only one cultured strain (Bban7, also known as RCC1105) has a sequenced genome (15). Cultured *Bathycoccus* strains and environmental clones have 100% 18S rRNA gene identity, unlike the *Micromonas* and *Ostreococcus* clades (16–18). However, targeted metagenomic data from the tropical Atlantic suggest that different *Bathycoccus* ecotypes may exist, based largely on an intron/intein presence/absence polymorphism in the gene *PRP8* (18, 19). Notably, the internal transcribed spacer (ITS) from the tropical Atlantic wild *Bathycoccus* metagenome also branches separately from cultured strains (18), while the ITS from cultured strains and two coastal Chilean *Bathycoccus* metagenomes are similar (20). Sequence variations have led to the proposal that two (18) or more (20) ecotypes exist and that the ecotypes may partition between mesotrophic and oligotrophic environments (18). However, little is known about overall genetic distances and whether the observed sequence variants do indeed correspond to ecotype-level differences.

Although much is known about other picophytoplankton, especially the cyanobacteria *Prochlorococcus* and *Synechococcus*, less data are available on the abundance of different class II prasino-

Received 21 August 2015 **Accepted** 28 December 2015

Accepted manuscript posted online 4 January 2016

Citation Simmons MP, Sudek S, Monier A, Limardo AJ, Jimenez V, Perle CR, Elrod VA, Pennington JT, Worden AZ. 2016. Abundance and biogeography of picoprasinophyte ecotypes and other phytoplankton in the eastern North Pacific Ocean. Appl Environ Microbiol 82:1693–1705. doi:10.1128/AEM.02730-15.

Editor: P. D. Schloss, University of Michigan

Address correspondence to Alexandra Z. Worden, azworden@mbari.org.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.02730-15.

Copyright © 2016, American Society for Microbiology. All Rights Reserved.

Downloaded from http://aem.asm.org/ on March 5, 2019 by guest

phytes in nature (10, 11, 21–23). Here, we enumerate picoeukaryotes as well as their cyanobacterial counterparts along an ecological gradient in the eastern North Pacific. Using quantitative PCR (qPCR), we enumerated two *Ostreococcus* clades (OI and OII), *Micromonas*, and *Bathycoccus* at coastal, mesotrophic, and oligotrophic ocean sites. The ecotypes were further resolved by establishing genetic distances among homologs identified in three genome-sequenced *Ostreococcus* species, the *B. prasinos* genome (15), and three *Bathycoccus* targeted metagenomes (18, 20). This information was used to investigate the oceanographic distributions of *Ostreococcus* and newly defined *Bathycoccus* ecotypes in metatranscriptomes generated here and in prior metatranscriptomic and metagenomic data from the same region. The metatranscriptomes were also used to explore gene expression in relation to the environmental gradient under study. Collectively, the results document the distributions and genetic divergence of widespread picoeukaryotic phytoplankton.

MATERIALS AND METHODS

Field sampling. The Transect survey was carried out in 2009 (cruise WFAD09), and the Drift study was performed in 2010 (cruise CANON10, also referred to as S410). Conductivity-temperature-depth (CTD) casts sampled 16 and 13 stations during the Transect and Drift surveys, respectively. During the Transect survey, the same stations were sampled on both the outward-bound (westward direction) and inbound legs of the cruise; these were separated by 8 days, during which other cruise activities and sampling occurred. Water for nutrients, flow cytometry, and DNA were collected using Niskin bottles mounted on a rosette with a CTD instrument and fluorometer. Five-hundred or 1,000 ml of seawater was filtered onto a 0.2-μm-pore-sized Supor filter for DNA (Pall Scientific, Port Washington, NY, USA) and frozen at −80°C. RNA samples from WFAD09 (50 liters) were collected using Niskin bottles, gravity filtered through 20-μm Nitex mesh, and subsequently gravity filtered onto an 0.8-μm-pore-size 142-mm-diameter Supor filter using a peristaltic pump. Bisected filters were transferred to 50-ml tubes and frozen at −80°C approximately 1 h after CTD instrument retrieval. Flow cytometry samples were preserved at a final concentration of 0.25% using an EM-grade glutaraldehyde (Electron Microscopy Sciences, Hatfield, PA, USA) (24). After 20 min of incubation at room temperature in the dark, the samples were frozen in liquid N_2. Dissolved iron profiles were generated from 3- to 10-ml samples collected using a trace metal clean rosette during CN207, a cruise 2 years prior to WFAD09 for which metagenomics data were available (analyzed here) and that had water characteristics similar to those of WFAD09 (19, 25, 26). Samples for iron were gently N_2 pressure filtered directly from the Teflon-lined Niskin bottles through a 0.2-μm acid-clean AcroPak Supor filter in a class 100 clean hood.

Nutrients, iron, chlorophyll, and flow cytometry. NH_4^+ was analyzed as in reference 27, and NO_3^-, NO_2^-, PO_4^{3-}, $Si(OH)_4$, and chlorophyll *a* (Chl *a*) were analyzed as in reference 28. Dissolved Fe samples were preconcentrated in line and analyzed using modifications of the sequential injection analysis (SIA) method described in reference 29. The following parameters have previously been reported from these cruises: temperature, salinity, NO_3^-, NH_4^+, PO_4^{3-}, and Chl *a* at stations H3, 67-70, and 67-155 on cruise CN207 (19); temperature, salinity, and Chl *a* for the Drift study (30); temperature, salinity, NO_3^-, NO_2^-, NH_4^+, and Chl *a* for 15 stations on cruise WFAD09 (31); and PO_4^{3-} for stations H3, 67-70, and 67-155 on cruise WFAD09 (32). Flow cytometry samples were run on an Influx (Becton Dickinson, San Jose, CA, USA) at approximately 25 μl · min^{-1}, as described in reference 5. Winlist 6.0 and 7.1 (Verity Software House) were used to analyze listmodes and characterize *Prochlorococcus*, *Synechococcus*, and eukaryote populations as in references 33 and 34.

Quantitative PCR. DNA was extracted using a modification of the DNeasy plant kit (Qiagen, Valencia CA, USA), including the addition of a bead-beating step (10). *Bathycoccus*, *Micromonas*, and *Ostreococcus* clades OI and OII were then enumerated in triplicate reactions, along with inhibition tests and no-template control reactions. These assays were performed in 25-μl volumes using a TaqMan master mix, as described in reference 10. Inhibition tests were performed using 2 μl of DNA template and an additional 2 μl of 18S rRNA gene plasmid (0.5×10^4 or 0.5×10^5 copies · μl^{-1}) per reaction. Based on these tests, environmental template solutions were diluted between 1:4 and 1:40 to prevent inhibition. Thermal cycling consisted of 10 min at 95°C (initial denaturation), followed by 45 cycles of 15 s at 95°C and 1 min at 60°C, using an AB7500 (Applied Biosystems, Foster City, CA, USA). Data were collected during the annealing phase. Ten-fold plasmid serial dilutions were used to generate standard curves. Threshold and baseline values were calculated using AB7500 software. Copy numbers per reaction were obtained by regression of the cycle threshold (C_T) against log scale copy numbers of standards and converted to 18S rRNA copies · ml^{-1} of seawater based on the volume of seawater filtered (and extracted) and the amount of template used. *Bathycoccus prasinos* RCC1105, *Ostreococcus lucimarinus*, *Ostreococcus* sp. strain RCC809, *Micromonas pusilla* CCMP1545, and *Micromonas* sp. strain RCC299 each have two complete rRNA operons in current genome assemblies (13–15), while *Ostreococcus tauri* appears to have one (35). Therefore, the 18S rRNA copies · ml^{-1} herein can be considered equivalent to twice the number of cells · ml^{-1} (depending on the cell cycle stage).

Statistical analysis. After testing for normality (using the Kolmogorov-Smirnov test), *t* tests or Mann-Whitney rank sum tests implemented in SigmaPlot (version 13.5) were used to compare water mass characteristics where, e.g., each prasinophyte taxon was detected versus not detected. Only photic zone samples (defined using a Secchi disk) were included. For the principal-component analysi (PCA), a correlation matrix was used to standardize photic zone data; observations with missing data were omitted. For the analysis using NH_4^+ data, the PCA contained 16 variables and 27 observations. Omitting NH_4^+ allowed for 39 observations with 15 variables. Analysis was performed using Matlab (OriginLab, Northampton, MA, USA).

Metatranscriptome library construction and sequencing. cDNA libraries were constructed from samples collected during the Transect survey (Fig. 1) from the transition zone (TZ; C45) and oligotrophic ocean deep chlorophyll maximum (OO_{DCM}; C28) at 13:02 and 14:08, respectively, at 19:37 for the coastal ocean (CO; C5), and at 09:11 for the oligotrophic ocean surface (OO_{SURF}; C26). Extraction supplies were obtained from Life Technologies (Grand Island, NY) unless otherwise noted, and the procedure allowed DNA (not used) and RNA to be extracted from the same material. Buffers were prepared with nuclease-free H_2O and filter sterilized; plasticware was also nuclease free/sterile. Filters were placed in petri dishes, covered with ~2 ml of lysis buffer (5 ml of RNAlater, 25% sucrose, 2.5 mg · ml^{-1} lysozyme, 5 mM Tris [pH 8], and 27.5 mM each EDTA and ethylene glycol tetraacetic acid; Sigma, St. Louis, MO, USA), sliced into ~1-cm^2 pieces, and transferred with the buffer into polypropylene tubes. The lysis buffer was brought to 20 ml, and samples were incubated for 1 h at 37°C. Four milligrams of proteinase K (Qiagen, Valencia, CA, USA) was added, and samples were frozen (liquid N_2) and thawed (55°C) three times. Samples were incubated at 55°C for 2 h under agitation with 4 mg of proteinase K and sodium dodecyl sulfate (Sigma, St. Louis, MO, USA) to 1% (vol/vol) and centrifuged 2 min at 4,500 × *g*. Supernatants were transferred to fresh tubes, and filter pieces were stored at −80°C for later reextraction. Extractions used pH 8 phenol, and nucleic acids were recovered by isopropanol precipitation, as in reference 36. Pellets were resuspended in 1 ml of RNA extraction buffer (4 M guanidine thiocyanate, 25 mM sodium citrate, 0.5% Sarkosyl; Sigma) and acidified with 50 μl of 2 M sodium acetate, pH 4 (Sigma). The resulting solution was extracted with pH 4 phenol, chloroform, and isoamyl alcohol (ratios of 125:24:1, respectively; Sigma), and RNA was precipitated with isopropanol (36). Pellets were washed twice with 75% ethanol and resuspended in water. RNA was purified using the RNeasy kit and the RNase-free DNase kit (Qiagen). To maximize recovery, the procedure was repeated on the refrozen filter pieces. The resulting two RNA aliquots per sample

Downloaded from http://aem.asm.org/ on March 5, 2019 by guest

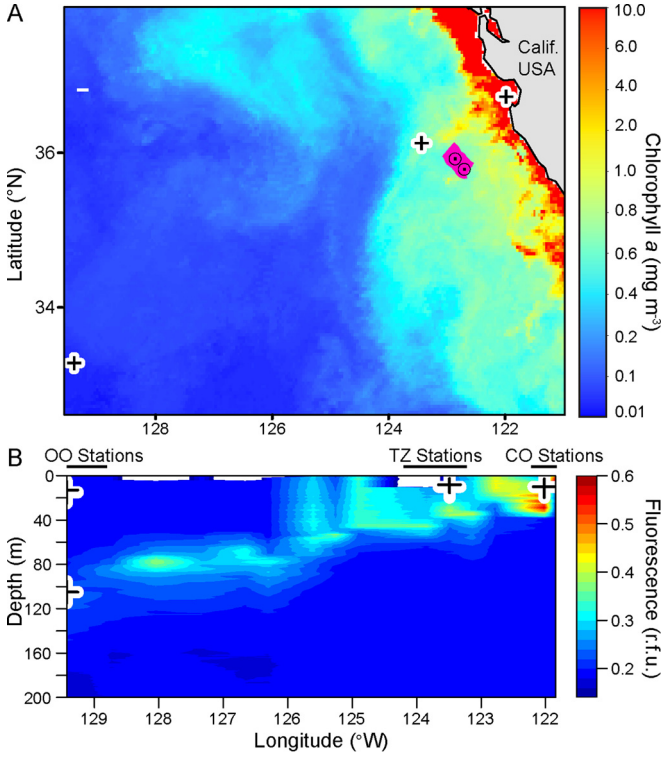

FIG 1 Locations sampled during the North Pacific Transect and Drift cruises. (A) Remotely sensed high-resolution Chl *a* concentrations during the month of the Drift study, October 2010 (MODIS/Aqua). Locations sampled for metatranscriptomes during the line 67 Transect survey (black crosses) and the overall trajectory of the drifter (pink) during the Drift study are depicted. Drift profiles for which qPCR was performed are also indicated (⊙; C44 and C51). (B) *In vivo* fluorescence (Chl *a* derived) along the Transect survey (2009). Horizontal lines above the plot show locations sampled in the three regime types (oligotrophic ocean, OO; transition zone, TZ; and coastal ocean, CO). Metatranscriptome sample sites (black crosses) were 25, 172, and 785 km from shore, which corresponded to casts C2, C42, and C20, respectively. r.f.u., relative fluorescence units.

were combined, yielding 1.8 to 12.3 μg of total RNA from ~25 liters of seawater. Then, 500-ng aliquots of RNA were amplified in two rounds of *in vitro* transcription using the Message AMP II antisense RNA (aRNA) amplification kit and manufacturer's protocols, except for the use of primer T7-BpmI $_{16}$VN (5′-GCCAGTGAATTGTAATACGACTCACTATA GGGGCGACTGGAGTTTTTTTTTTTTTTTTVN-3′) instead of primer T7 Oligo (37), yielding between 112 and 237 μg of aRNA per sample. A total of 60 μg of aRNA per sample was converted to single-stranded cDNA using the Superscript III first-strand synthesis system (Invitrogen, Carlsbad, CA, USA) with random hexamer primers. Double-stranded cDNA was produced by incubating DNA for 2 h at 16°C with 40 U of DNA polymerase I, 10 U of DNA ligase, and 200 μM deoxynucleoside triphosphates (dNTPs) in 5× second-strand buffer. RNA was digested with 2 U of RNase H. cDNAs were blunt ended by adding 5 U of T4 DNA polymerase and incubating them at 16°C for 5 min. Purification used Agencourt AMPure XP magnetic beads (Beckman Coulter, Indianapolis, IN, USA). The resulting material was size selected (300 to 3,000 bp) on a low-melting-point 1% agarose gel and extracted using the QIAquick gel extraction kit (Qiagen). Poly(A) tails were digested using BpmI (NEB, Ipswich, MA, USA) according to the manufacturer's protocol, and cDNA was purified with AMPure XP beads. cDNA quantity (1.2 to 1.5 μg) and quality were assessed with Qubit and Bioanalyzer (Agilent, Santa Clara, CA, USA), respectively. Sequencing was performed on a 454 FLX+ instrument using titanium chemistry (Roche/454, Branford, CT, USA).

Genetic distances among putative *Bathycoccus* and *Ostreococcus* ecotypes. To identify genes for characterizing *Bathycoccus* and *Ostreococcus* genetic diversity and for use as ecomarkers in environmental samples, groups of homologous sequences were first identified in the predicted proteomes from *B. prasinos* (strain Bban7) and three *Ostreococcus* genomes (taxa listed above). *Bathycoccus* sequences from three targeted metagenomes (19, 20) were also used after identifying all possible open reading frames (ORFs; i.e., sequence between stop codons with a minimal length of 60 amino acid residues). The analysis was not performed for *Micromonas* because, of the seven established clades (12), isolates from only two have complete genome sequences. Clusters of homologs were computed using OrthoMCL (38) based on all versus all BLASTP (39) searches of *Bathycoccus* and *Ostreococcus* proteomes, using a cutoff E value of 10^{-100} E. Clusters containing one sequence from each input source were kept to ensure analysis of exclusively single-copy orthologs (i.e., clusters composed of sequences for the four and three *Bathycoccus* and *Ostreococcus* predicted proteomes, respectively). Clusters producing alignments shorter than 100 amino acid residues were discarded. Final data sets were composed of 1,104 *Bathycoccus* clusters and 3,212 *Ostreococcus* clusters. Protein sequences from each cluster were aligned with T-Coffee v9 (40) and back translated to codon alignments to improve gene nucleotide alignments. Columns with gaps were removed from the nucleotide alignments, and identities of each codon alignment were used to estimate percent identity at the genus level. To determine if the genetic distances were significant among *Bathycoccus* and *Ostreococcus* ecotypes, *t* tests were conducted on nucleotide identity differences between pairs of ecotypes in R (41).

To identify ecotype/species ecomarkers, the prior all versus all BLASTP results were run again in OrthoMCL with the *Bathycoccus* and *Ostreococcus* data together. Only groups composed of single-copy genes and populated by all seven sequence sources (i.e., homolog groups containing seven ORFs, one from each distinct *Bathycoccus* and *Ostreococcus* sequence source) were retained. To avoid potential ORF length biases in sequence recruitment (mainly due to large indels or to Bban7 having longer ORFs likely related to gene model differences), the sequences within a cluster were corrected by discarding indels. Sequences from the same cluster were aligned using T-Coffee, and positions with more than three gaps were removed. Only groups of corrected sequences for which the longest sequence was ≤30% longer than the smallest were retained. Sequences with mean identities <50% were discarded, because their alignments were error prone, inflating the distance among taxa. Sequences with >95% identity were also discarded, because they were too similar to be accurately classified for the purposes of this study. A total of 112 homolog groups were identified that could be used as ecomarkers to classify meta-omic reads into different *Bathycoccus* and *Ostreococcus* types. To test if the *Bathycoccus* and *Ostreococcus* ecotype distributions differed significantly from each other across the sampling locations, pairwise *t* tests were conducted on their ecomarker relative abundances, and *P* values were adjusted for multiple comparisons using Bonferroni correction.

Metatranscriptome analyses. Twelve published metatranscriptomes (30) from samples collected during the Drift study were reanalyzed; specifically, samples were obtained at 4-h intervals starting at 14:00 on 16 September 2010 (36°2.712, −123°1.302) and ending 44 h later (35°47.412, −122°42.54). A 13th metatranscriptome from the prior publication on the Drift study (30) was excluded, because it led to overrepresentation of one time point. Additionally, unlike the Transect survey metatranscriptomes, the Drift study sequences were generated using methods tuned for prokaryotic transcript recovery, and fewer reads were generated per sample. Redundant sequences were removed from Transect and Drift metatranscriptomes using CD-Hit v4.6 (42), because they were considered to be 454 artifacts (43), and rRNAs were removed using riboPicker (44) (see Table S1A in the supplemental material). In both cases, default settings were used. The 454 reads were assigned to taxonomic groups using BLASTX (39) against predicted proteomes from sequenced genomes and

Downloaded from http://aem.asm.org/ on March 5, 2019 by guest

transcriptomes from cultures of the following: *Aureococcus anophagefferens*, *B. prasinos*, *Bigelowiella natans*, *Chlamydomonas reinhardtii*, *Ectocarpus siliculosus*, *Emiliania huxleyi*, *Guillardia theta*, *Micromonas* sp. RCC299, *Micromonas pusilla* CCMP1545, *Micromonas* sp. strain CCMP2099, *O. tauri*, *O. lucimarinus*, *Ostreococcus* sp. RCC809, *Pyramimonas parkeae*, *Phaeodactylum tricornutum*, *Thalassiosira pseudonana*, algal viruses, and vascular plants. Predicted proteins from genome projects were retrieved from GenBank or the Joint Genome Institute genome portal (45). The results were filtered with cutoffs of >60% identity, a bit score >50, and an E value of $\leq 10^{-20}$ with the reference protein sequence. Sequences that passed this filter were used as BLASTX queries against the GenBank nonredundant (nr) database (46). This filtering step was performed only for the subset of algae with the most reads, i.e., all those analyzed further, because of its computational intensity (Table S1B). Those that matched a different taxon ID in nr with a better overall score were removed from the set (4.9% ± 3.6% of those initially assigned to genome sequenced picoprasinophytes) (Table S1B). To count reads per reference protein, assigned reads were divided into species-specific sets, and the proteins of the corresponding species were used as TBLASTN queries against each of the parsed read sets. Hits were only counted if they matched the protein uniquely or if the protein was present in duplicate. Reads mapped to a taxon were summed, and gene counts in that sample were expressed as a percentage of the taxon total sum (in that sample). For Drift study metatranscriptomes, this percentage was computed from the two 14:00 (combined) and two 22:00 (combined) data sets to compare to the Transect survey samples collected at the same time of day. The 12 Drift study metatranscriptomes were also summed to maximize *B. prasinos* transcript counts, and the summed values are used for comparisons here.

The minimum cutoff for comparative analyses was ≥2 reads per predicted protein (on average for a taxon) in two or more samples. *B. prasinos* and *O. lucimarinus* met these criteria, but the 67-155 15-m sample did not for the eukaryotic phytoplankton taxon analyzed and was therefore excluded from expression comparisons. Final comparisons were analyzed between the genes in the sample with the fewest total reads to a taxon (but ≥2 per gene on average, i.e., *B. prasinos* in the 105-m OO_{DCM} sample), with a focus on those with higher relative percent expression than in samples that had more overall reads assigned to that taxon. Because a greater number of overall *Bathycoccus* reads were available from the CO and the Drift study, gene expression detection sensitivity was expected to be superior at those sites compared with reads from the OO_{DCM}. Thus, higher relative percentage of reads from a particular gene in the OO_{DCM} (compared to the CO and the Drift study) should be reflective of OO_{DCM} responses without additional rarefraction steps.

Super oxide dismutases (SODs) and nitrogen transporters were identified in metatranscriptomic data after first performing reciprocal BLASTP (39) searches using previously annotated proteins (47) against nr and the Joint Genome Institute (JGI) prasinophyte browsers to recruit previously unreported orthologs. For the ammonium transporter (AMT) phylogeny, *Bathycoccus* and *Ostreococcus* RCC809 sequences were added to alignments from reference 47, the alignment was manually adjusted, and ambiguously aligned or nonhomologous positions were masked. Phylip was used to construct neighbor joining (NJ) distance trees (48), and maximum likelihood (ML) methods were performed in PhyML (49) with 100 bootstrap replicates.

Ecomarker analysis was also used to identify *Bathycoccus* and *Ostreococcus* reads in the WFAD09 and CANON2010 metatranscriptomes and in line 67 metagenomes available from 2007 (19). The 112 *Bathycoccus*/*Ostreococcus* ecomarkers were used as TBLASTN (39) queries against environmental reads. TBLASTN hits (bit-score, >50; E value, $<10^{-5}$) were then used as BLASTX queries against nr. Only reads having their best BLASTX hit to one of the *Bathycoccus*/*Ostreococcus* ecomarkers (and bit-score, >50; E value, $<10^{-5}$) were retained. These candidate reads were then used as BLASTN queries against the four *Bathycoccus* and three *Ostreococcus* ORF data sets. ORF data sets from *Micromonas* CCMP1545 and RCC299 (13) were used to remove false positives (i.e., reads with best BLASTN hits to *Micromonas* were removed). Reads were classified as being *Bathycoccus* or *Ostreococcus* if they produced a BLASTN hit with a bit-score of ≥250 and nucleotide identity ≥95% and binned to one of the four *Bathycoccus* or three *Ostreococcus* types based on their best BLASTN hit. Relative contributions were then computed for each sample. Those with the same statistics for two or more distinct *Bathycoccus* or *Ostreococcus* sequences were discarded (7.6% of WFAD09 candidate reads, 4.7% of CANON2010, and 4.9% of CN207).

Nucleotide sequence accession number. Metatranscriptomes generated in this study are available in the NCBI Sequence Read Archive under project number PRJNA300413.

RESULTS AND DISCUSSION

Environmental conditions. Two cruises sampled picophytoplankton in the northeast Pacific Ocean off central California during boreal autumn (Fig. 1A). The Transect study sampled three ecological regimes along California Cooperative Oceanic Fisheries Investigations (CalCOFI) line 67: the coastal ocean (CO), a mesotrophic/transition zone (TZ), and the oligotrophic ocean (OO) (Fig. 1B) (cruise WFAD09) (32). Along line 67, surface waters at coastal stations are typically colder and more nutrient rich, reflecting coastal upwelling (19, 50), while sea surface temperatures increase and nutrient concentrations decrease offshore (see Data Set S1 in the supplemental material). The Drift study followed a semi-Lagrangian drifting environmental sample processor (ESP) to resample the same mesotrophic water mass near line 67 (Fig. 1A) (cruise CANON10).

Nutrient and chlorophyll concentrations as well as other physicochemical parameters illuminated broad differences across the Transect survey (Fig. 1 and 2; see also Fig. S1 in the supplemental material). NO_3^- and NH_4^+ declined from a euphotic zone maximum of 20 μM and 0.51 μM (31), respectively, to below detection limits (0.020 μM and ~0.010 μM, respectively; lower NH_4^+ values were detected but may not have been quantitative), while PO_4^{3-} was always detectable (Table 1; see also Data Set S1 in the supplemental material). Chl *a* decreased and the chlorophyll maximum deepened with increasing distance from shore (Fig. 1B). At the most offshore stations, the deep chlorophyll maximum (DCM) ranged from 85 to 105 m, and DCM NO_3^- and NH_4^+ ranged from 0.001 to 0.048 μM and 0.003 to 0.009 μM, respectively (Fig. 2A; see also Data Set S1). During the 8 days between sampling the TZ on the westward leg of the Transect survey cruise (casts C8 and C10) and on the inbound leg (C42), the water column structure became less stratified and Chl *a* became more homogeneous from 0 to 35 m (see Fig. S1 and Data Set S1). Drift survey NO_3^- concentrations were between those of the Transect survey CO and TZ. Chl *a* extended deeper in the water column during the Drift study than in the CO, but it did not exhibit the more pronounced subsurface maxima observed at TZ sites (Fig. 2A and Table 1; see also Data Set S1).

Phytoplankton abundance. Phytoplankton groups varied dramatically across the regimes investigated during the Transect survey. Small eukaryote abundance increased 2- to 3-fold between the OO surface and DCM, while the abundance of *Synechococcus* declined beneath 60 m and *Prochlorococcus* dominated numerically throughout the water columns (maximum of 195,118 cells · ml^{-1}) (Fig. 2B; see also Data Set S1 in the supplemental material). During TZ sampling on the westward leg of the cruise, abundance of small eukaryotes in the upper 40 m (e.g., C10, ranging from 15,336 to 19,646 cells · ml^{-1}) was an order of magnitude higher than in the OO (Table 1; Fig. 2B). *Prochlorococcus*, *Synechococcus*,

Downloaded from http://aem.asm.org/ on March 5, 2019 by guest

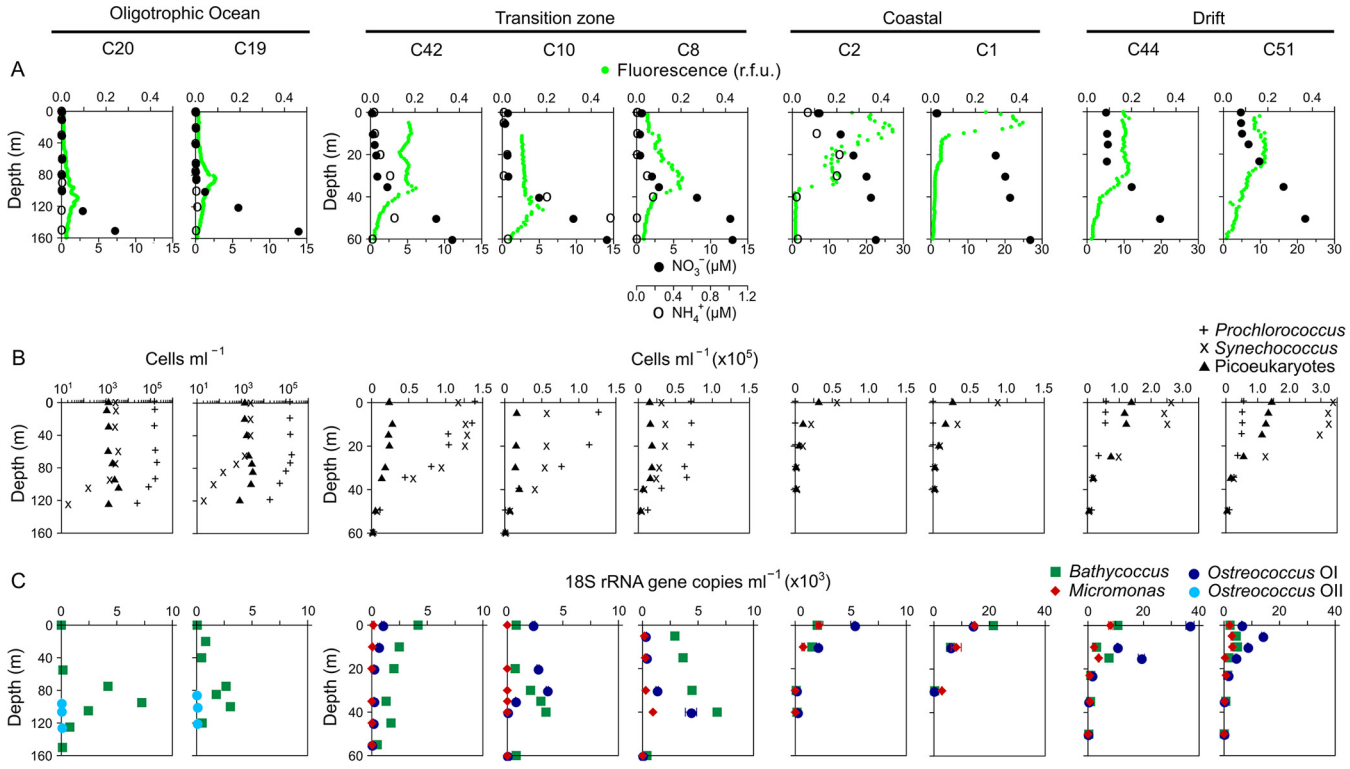

FIG 2 Profiles from the OO, TZ, and CO locations sampled during the Transect survey and Drift study. Note different *y*-axis scales for OO casts (C20 and C19) from other plots and different *x*-axis scales for various organismal abundance plots. (A) NO_3^-, NH_4^+, and *in vivo* fluorescence (Chl *a* derived). (B) *Prochlorococcus*, *Synechococcus*, and photosynthetic picoeukaryote abundance by flow cytometry. (C) *Ostreococcus* clade OI, clade OII, *Micromonas*, and *Bathycoccus* 18S rRNA gene copies · ml^{-1} by qPCR with error bars representing the standard deviation of technical triplicates. *Ostreococcus* clade OII was assayed but rarely detected. Note that Transect survey profiles for metatranscriptome sample sites are represented by C2, C42, and C20; outward-bound or westward-leg TZ casts are C10 and C8, while inward-bound sampling was C42, 8 days later.

and picoeukaryotes were more abundant in the TZ during the inward-bound leg, when metatranscriptome sampling was performed, than during the westward-bound leg. Small eukaryotes had a maximum of 28,137 cells · ml^{-1} at 0 m. *Prochlorococcus* increased to 139,435 cells · ml^{-1} (0 m) and was roughly equal to *Synechococcus* (117,037 cells · ml^{-1}, 0 m). In the CO, small eukaryotes and *Synechococcus* were abundant only in the upper 10 m, while few *Prochlorococcus* cells were present (Fig. 2B; see also Data Set S1). The *Prochlorococcus* cells that could be distinguish using cytometry amounted to <2,000 cells · ml^{-1}. Eukaryotes and *Syn-*

TABLE 1 Environmental parameters in representative Transect survey and Drift study samples[a]

Parameter	Transect survey (2009)[b]				Drift study (2010) (start to end of drift)
	OO_{DCM}	OO_{SURF}	TZ	CO	
Latitude	33.286	33.286	36.126	36.740	36.045 to 35.790
Longitude	−129.429	−129.429	−123.490	−121.02	−123.022 to −122.709
Depth (m)	105	15	10	10	25
Time of day	14:08	09:11	13:02	19:37	Diel
Date (mo/day)	10/7	10/6	10/11	10/2	9/16 to 9/18
NO_3^- (μM)	0.001	0.000	0.257	12.909	8.51 ± 1.87
NH_4^+ (μM)	0.003	0.000	0.047	0.263	Not measured
$Si(OH)_4$ (μM)	2.32	1.35	2.92	9.912	7.76 ± 2.00
PO_4^{3-} (μM)	0.34	0.47	0.49	1.153	1.49 ± 0.19
Total Chl *a* (μg · $liter^{-1}$)	0.75	0.10	1.79	2.12	0.85 ± 0.08
Synechococcus (cells · ml^{-1})[c]	159	2,797	126,858	21,512	111,300 ± 18,357
Prochlorococcus (cells · ml^{-1})[c]	82,502	156,236	135,649	1,938	35,969 ± 1,272
Total eukaryotes (cells · ml^{-1})[c]	3,674	1,086	28,137	10,574	64,824 ± 13,216

[a] Transect data come from the depth/locations used for metatranscriptome sequencing. For the Drift study, ranges or averages are provided for samples corresponding to the 12 metatranscriptomes (30) reanalyzed here. See Data Set S1 in the supplemental material for all data.
[b] Transect nutrients and Chl *a* were measured at approximately the same depth on a cast adjacent to that of metatranscriptome water collection.
[c] Flow cytometry counts for the Drift study are the averages and standard deviations for casts C42 and C48, the 06:00 cast for which qPCR was performed.

Downloaded from http://aem.asm.org/ on March 5, 2019 by guest

echococcus were more abundant in the Drift study than in the Transect survey. For example, at the surface of Drift study profile C44, *Synechococcus* and eukaryotes were 241,697 cells · ml^{-1} and 117,062 cells · ml^{-1}, respectively. All three phytoplankton groups extended deeper in the water column in the Drift study than in the CO (Fig. 2B), including *Prochlorococcus* populations which were easily definable in the former but not in the latter.

During the Transect survey, *Prochlorococcus* HLI and LLI ecotypes dominated the OO and TZ samples based on a previous report (32). Clade IV dominated *Synechococcus* in the OO, and clades I, V/VI/VII, EPC1, and EPC2 were present at 3% to 10% of TZ cyanobacterial 16S amplicons (32). However, knowledge on the abundance of small eukaryotic taxa is limited to *Ostreococcus* (10) in this region. We quantified the picoeukaryotic prasinophytes *Bathycoccus*, *Micromonas*, and *Ostreococcus* in the profiles using qPCR (Fig. 2C). *Ostreococcus* clade OII was detected only in the OO, with a maximum of 72 18S rRNA copies · ml^{-1} at the DCM. *Ostreococcus* clade OI was not detected in the OO but was present in CO and TZ waters. Laboratory studies suggest that isolates from clades OI and OII correspond to differently light-adapted ecotypes (51–53). However, our data support prior field-based results which indicate that spatial factors may govern *Ostreococcus* ecotype distributions, with clade OI present in mesotrophic waters and clade OII in oligotrophic waters (10). Here, *Ostreococcus* clade OI abundances also varied between the westward leg (up to 4,375 ± 511 18S rRNA copies · ml^{-1}) and inward-bound (<1,000 18S rRNA copies · ml^{-1}) TZ samplings. Moreover, OI abundance increased with depth in C8 and C10 (westward leg; Fig. 2C). Maximum *Ostreococcus* clade OI abundance was observed at the surface during the Drift study (36,713 ± 1,485 18S rRNA copies · ml^{-1}).

The maximum *Micromonas* 18S rRNA copies · ml^{-1} was observed at the CO (14,613 ± 196), with lower abundances in the Transect survey TZ (949 ± 170) and Drift study (8,080 ± 641). These values likely represent contributions from multiple *Micromonas* clades, since seven have been characterized to date (12) and several of these cooccur in the southern California bight (17). *Micromonas* and *Ostreococcus* clade OI were detected in the same 29 qPCR samples (41 were analyzed), and both were undetectable in the OO. The average water temperature for these 29 samples (which include both Transect survey and Drift study) was 14°C ± 2°C.

Unlike *Micromonas* or *Ostreococcus* clade OI and clade OII, *Bathycoccus* was detected in all 41 qPCR samples. *Bathycoccus* dominated the prasinophyte groups in the majority of Transect survey profiles and had the highest number of 18S rRNA copies · ml^{-1} observed (21,368 ± 327, surface CO) (Fig. 2C). In the OO and TZ, maximum *Bathycoccus* abundance was near the DCM (e.g., 7,251 ± 289 18S rRNA copies · ml^{-1}, OO_{DCM}; and 6,722 ± 328 18S rRNA copies · ml^{-1}, westward-leg TZ, 40 m), whereas, in the CO and inward-bound TZ, as well as the Drift study, the maximum occurred at the surface (Fig. 2C). These distribution patterns are generally in accord with the deepening of the DCM offshore (Fig. 1B). During the Drift study, *Bathycoccus* reached 10,759 ± 436 18S rRNA copies · ml^{-1} (0 m, C44), but *Ostreococcus* clade OI was the dominant prasinophyte (Fig. 2C).

Nine other quantitative studies on picoprasinophytes have enumerated more than one genus. In the coastal Mediterranean Sea, *Bathycoccus* was undetectable in autumn but contributed 12% to 20% of the total picoeukaryotic population in winter (absolute copies · ml^{-1} were not provided) (54). Few (≤5%) *Ostreococcus* 18S rRNA copies · ml^{-1} were recovered in either season in this study, which used different qPCR primers (54) than ours (10). Using tyramide signal amplification-fluorescent *in situ* hybridization (TSA-FISH) probes, *Micromonas* and *Bathycoccus* together accounted for 60% of the green algal cells in the Norwegian and Barents Seas in late summer, while *Ostreococcus* was not detected (21). Similarly, a TSA-FISH study demonstrated that prasinophytes dominate the autotrophic picoeukaryotic community in the English Channel, specifically *Micromonas*, *Bathycoccus*, and, to a lesser extent, *Ostreococcus* (averages of 45%, 8%, and 1% of picoeukaryotes, respectively) (55). At the Bermuda biological time series station (BATS), *Bathycoccus*, *Ostreococcus* clade OII, and *Micromonas* were present at 96, 2,273, and 1,344 18S rRNA copies · ml^{-1}, respectively, at 40 m in a profile collected during spring mixing (56). However, during summertime, only *Bathycoccus* remained detectable (~100 18S rRNA copies · ml^{-1} in the DCM). One difference between the Pacific OO region studied here and BATS is the PO_4^{3-} availability, which is below detection (<0.010 μM) at BATS in summer but is ~0.36 μM in the Pacific OO surface waters (see Data Set S1 in the supplemental material). Perhaps the most similar study in terms of general environs is one at an upwelling influenced coastal site off Chile, which used TSA-FISH to quantify picoprasinophytes (23). This study found that *Ostreococcus* was more abundant than *Micromonas* and *Bathycoccus* in 17 of 23 samples analyzed over the year and, overall, was the most abundant member of the eukaryotic picophytoplankton.

Phytoplankton abundance in relation to environmental data. The clearest patterns in our data were the lack of *Ostreococcus* clade OI and *Micromonas* in the OO and the absence of *Ostreococcus* clade OII at study sites inshore of the OO. In the CO, which had the shallowest photic zone of all the stations, maximum 18S rRNA copies · ml^{-1} occurred at the surface for *Bathycoccus*, *Ostreococcus* clade OI, and *Micromonas* (Fig. 2A and C; see also Data Set S1 in the supplemental material). During westward-leg sampling of the TZ (C8, C10), the *Bathycoccus*, *Ostreococcus* clade OI, and *Micromonas* maxima were associated with increased NO_3^- and NH_4^+ deeper in the water column (Fig. 2; see also Data Set S1). The inward-bound TZ (C42) had more homogenous Chl *a* between the surface and 40 m than the outward-bound samples and exhibited an increase in *Bathycoccus* and *Ostreococcus* clade OI 18S rRNA copies · ml^{-1} at the surface compared to other depths, as seen in the CO and the Drift study (C44). Notably, NH_4^+ was also higher (0.04 μM) at the surface in this inward-bound TZ cast than for the westward-leg TZ casts (e.g., 0.01 μM).

Patterns were investigated further by partitioning photic zone samples according to whether a taxon was detected by qPCR or not. It should be noted that our sample set was biased for mesotrophic and coastal waters more than for oligotrophic waters. We found that PO_4^{3-}, NO_3^-, and NH_4^+ were significantly higher ($P < 0.003$) in samples containing *Micromonas* and *Ostreococcus* clade OI (which always cooccurred) than in samples in which these taxa were not detected. Median nutrient values for photic zone samples with *Micromonas* and *Ostreococcus* clade OI were 0.802 μM PO_4^{3-} and 4.903 μM NO_3^- ($n = 27$). For the photic zone samples without these two picoprasinophytes ($n = 11$), the median PO_4^{3-} was 0.433 μM and NO_3^- was below detection. Likewise, NH_4^+ was higher for samples with these taxa (0.080 μM, $n = 14$) than for those without (0.002 μM, $n = 12$). Samples that contained *Ostreococcus* clade OII had significantly ($P < 0.01$)

Downloaded from http://aem.asm.org/ on March 5, 2019 by guest

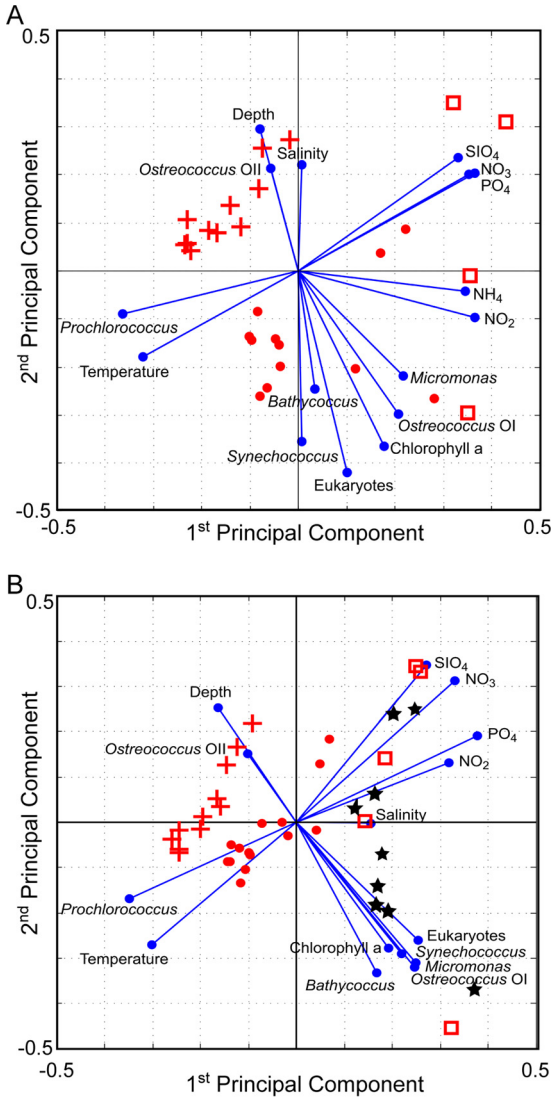

FIG 3 Principal-component analysis (PCA) using available data from all cruises and stations. The points represent scaled PCA scores for observations, and the vectors indicate loadings. PCA, including NH_4^+ ($n = 27$, 16 dimensions) (A) and PCA excluding NH_4^+ ($n = 39$, 15 dimensions) (B). Samples are in red for the Transect survey OO (+), TZ (●), and CO (□) and black for the Drift study (★). In both analyses, the first component is driven primarily by the concentrations of macronutrients and their inverse relationships to *Prochlorococcus* and temperature, and the second component reflects differences in biology and chemistry linked to depth (or covarious characteristics) in the water column.

lower NH_4^+ (median, 0.005 μM) than those with *Micromonas* and *Ostreococcus* clade OI (median, 0.080 μM), but other measured parameters (PO_4^{3-}, NO_3^-, temperature, and salinity) were not significantly different. Significant differences were not observed for any of these parameters for the samples containing *Bathycoccus* (all samples) and the subset that also contained *Micromonas* and *Ostreococcus* clade OI.

Two principal-component analyses (PCA) were performed on photic zone data. One was limited to samples that had NH_4^+ data ($n = 27$; Fig. 3A), which was not measured during the Drift study. The other included all photic zone samples with nutrient measurements ($n = 39$) (Fig. 3B) apart from NH_4^+. The first, second, and third components accounted for 34.1%, 29.2%, and 11.3% of the variance, respectively (74.6% cumulative variance), for samples with NH_4^+ measurements and 32.3%, 26.6%, and 13.2% of the variance, respectively (72.1% cumulative variance), for the broader set. PCA scores for each station clustered loosely according to geographic location, with TZ samples falling along a gradient between OO and CO stations (Fig. 3A). Notably, Drift study samples showed the greatest variability but, overall, were closest to Transect survey CO samples. *Ostreococcus* clade OII and *Prochlorococcus* had distinct patterns from the other groups studied (i.e., *Synechococcus*, *Ostreococcus* OI, *Bathycoccus*, *Micromonas*, and small eukaryotes). The abundance of *Ostreococcus* OII showed a positive relationship with OO depth, while *Prochlorococcus* abundance was positively associated with higher temperatures in surface waters. The other phytoplankton groups were positively related to each other (Fig. 3B). NH_4^+ and NO_2^- concentrations were also positively associated, likely reflecting ammonia oxidation by bacteria and/or archaea (31). Higher NH_4^+ and NO_2^- standing stocks were associated with increased abundance of *Synechococcus*, *Ostreococcus* OI, *Bathycoccus*, *Micromonas*, and picoeukaryotes, while other inorganic nutrients (e.g., NO_3^-) were less so. Overall, the PCA results highlight the association of cold, nutrient-rich, upwelling-influenced surface water near the coast with high abundances of *Ostreococcus* clade OI, *Bathycoccus*, *Micromonas*, and *Synechococcus* and highlight the transition to warmer, nutrient-poor California current waters containing *Prochlorococcus*, *Bathycoccus*, and *Ostreococcus* clade OII.

Among the picoprasinophytes we quantified, only *Bathycoccus* was present at all stations. This lack of biogeographical pattern might reflect the presence of multiple *Bathycoccus* ecotypes (which would both be amplified by our qPCR primer-probe set). Moreover, at the inward-bound TZ station, a *Bathycoccus* abundance peak was observed at the surface and a second slight peak was present below 40 m (Fig. 2C), as might be expected if the ecotypes were adapted for different depth-associated parameters, such as light or nutrients.

***Bathycoccus* ecotype genetic distances.** We sought to test whether the omnipresence of *Bathycoccus* in the euphotic zone reflected success of a single cosmopolitan species, *B. prasinos*, or if it reflected differential contributions from more than one ecotype. First, we needed to establish genetic distances between putative *Bathycoccus* ecotypes (18, 20) to enable their identification in -omics field data. Because a definition of picoeukaryote ecotype-level divergence does not exist, we first analyzed three *Ostreococcus* genome sequences to establish a benchmark for interpreting *Bathycoccus* genetic distances. The available *Ostreococcus* genomes come from species that belong to *Ostreococcus* clades for which distributions have been characterized in parts of the Pacific and Atlantic Oceans (8, 10). The *Ostreococcus* clade OI qPCR primer-probe set used here was designed to amplify both *O. lucimarinus* and *O. tauri* (10) to distinguish them from *Ostreococcus* clade OII. However, *O. lucimarinus* and *O. tauri* share only ~90% of their protein-encoding genes (14) and form separate clades using 18S rRNA gene sequences or other markers (12); hence, we have termed *O. tauri* clade OI/C according to reference 16. A comparative genome analysis has not been published that includes *Ostreococcus* sp. RCC809 (clade OII), but, as shown here (Fig. 2) and previously (10), OI and OII have distinct biogeographies that support clear niche differentiation. Thus far, cooccurrence of clades OI and OII has only been observed where Atlantic continental

Downloaded from http://aem.asm.org/ on March 5, 2019 by guest

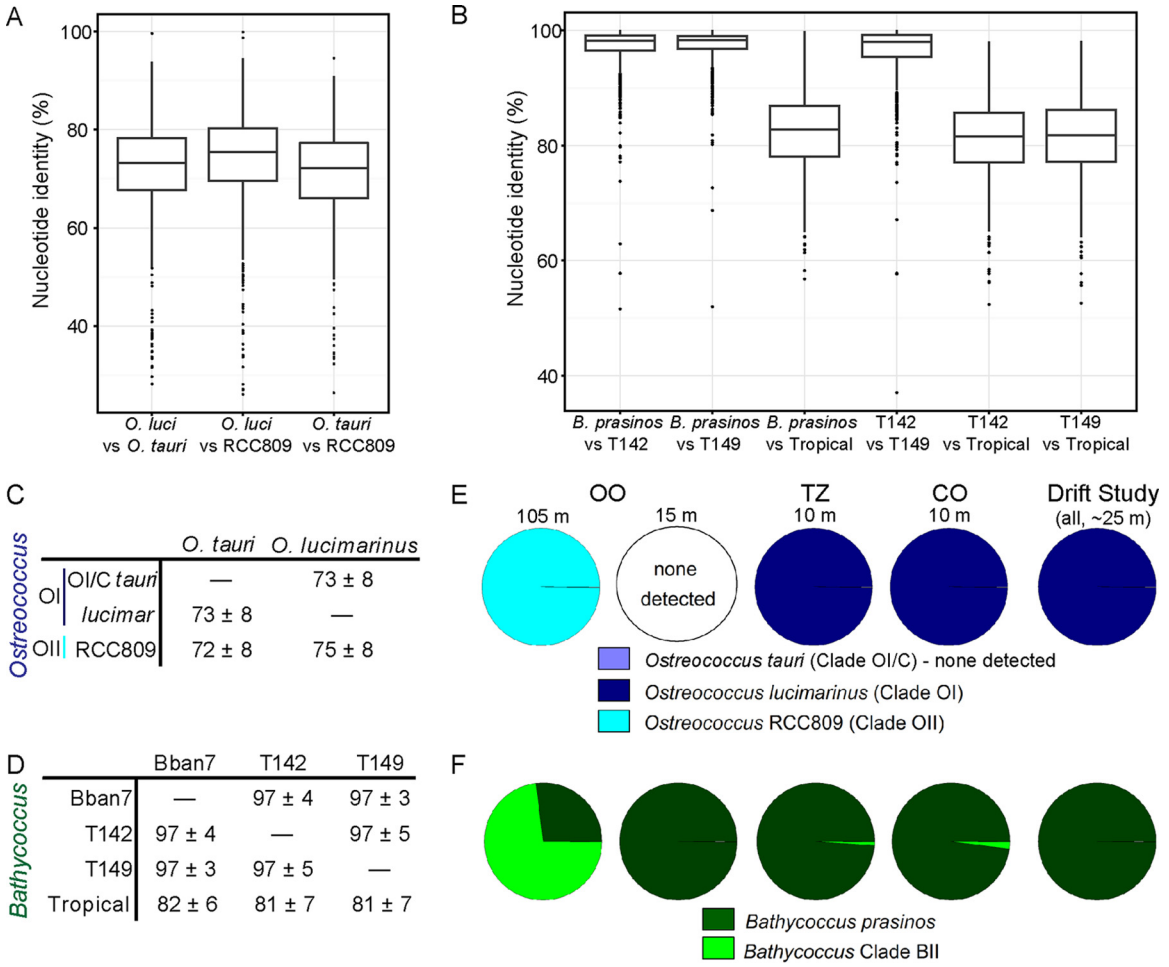

FIG 4 *Bathycoccus* and *Ostreococcus* ecotypes defined by homolog cluster analysis and natural distributions. (A and B) Percent nucleotide identity distributions for 3,212 single copy homologs in the three genome-sequenced *Ostreococcus* species (A) and 1,104 single copy homologs in the *B. prasinos* genome and *Bathycoccus* targeted metagenomes (B) (15, 18, 20). The black dots represent distribution outliers. Note the difference in *y*-axis scales. The divergence across homologous genes observed here allows differentiation of the *Ostreococcus* and the *Bathycoccus* ecotypes. (C and D) *Ostreococcus* (C) and *Bathycoccus* (D) average nucleotide identities based on the homolog groups identified in genomes and prior *Bathycoccus* targeted metagenomes (15, 19, 20). (E and F) Ecomarker analysis for *Ostreococcus* (E) and *Bathycoccus* (F) ecotypes in the Transect survey and Drift study metatranscriptomes.

shelf waters intersect with the Gulf Stream current (10). We identified 3,212 single-copy homologs present in *O. lucimarinus*, *O. tauri*, and *Ostreococcus* sp. RCC809 that could be used to define nucleotide distances between these three species. Our analyses showed that these homologs had, on average, 72% to 75% nucleotide identity (Fig. 4A and C; see also Data Set S2 in the supplemental material) and were significantly different ($P < 0.001$).

The same type of analysis was performed for available *Bathycoccus* genome-level data. In this case, 1,104 homologous proteins were identified in the *B. prasinos* Bban7 genome (15) and the *Bathycoccus* targeted metagenomes from the tropical Atlantic (19) and coastal Chile (20) (see Data Set S2 in the supplemental material). The genes encoding these proteins in the coastal Chilean metagenomes (T142 and T149; sorted from a well-mixed coastal water column, 2 days and 15 km apart, at 5 and 30 m, respectively) have on average 97% ± 4% nucleotide identity, and their identities are not statistically different ($P > 0.05$) from *B. prasinos*, which was isolated in the coastal Mediterranean (Fig. 4B and D). Thus, we concluded that the coastal Chilean metagenomes and *B. prasinos* were reflective of a single ecotype, rather than multiple species or ecotypes; here, these three are considered representatives of the clade BI ecotype. The tropical Atlantic *Bathycoccus* metagenome came from warm oligotrophic waters and had lower identity to *B. prasinos* (82% ± 6%; $P < 0.001$) and to the coastal Chilean metagenomes (Fig. 4B and D). Thus, while the tropical Atlantic *Bathycoccus* metagenome and *B. prasinos* shared higher identity than the three *Ostreococcus* species, their divergence was significant and notable (Fig. 4); hence, we termed it the clade BII ecotype. The conclusion that two *Bathycoccus* ecotypes exist is supported by phylogenetic analysis of their internal transcribed spacer (18), a marker for which divergence has been related to speciation (57).

Taxon distributions in -omics data. We analyzed the relative contributions of clades BI and BII (which could not be delineated by our qPCR primer-probe set) in samples from the Transect survey and the Drift study, as well as clades OI and OII (which were delineated using qPCR), to gain insight into their distributions. Metatranscriptomes were generated from surface samples for the

Downloaded from http://aem.asm.org/ on March 5, 2019 by guest

TABLE 2 Numbers of metatranscriptomic reads assigned to genome-sequenced eukaryotes by sample location and depth based on a BLASTX search against a database of genome-level information from eukaryotic algae and viruses[a]

Eukaryotic species	Taxon (genome sequenced)	Predicted proteome[b]	Transect survey finding				
			OO_{DCM}	OO_{SURF}	TZ	CO	Drift
Class II prasinophytes (*Mamiellophyceae*)	*Bathycoccus prasinos* Bban7	7,921	16,958	602	12,079	30,613	75,887
	Ostreococcus lucimarinus	7,605	451	473	1,604	12,437	180,370
	Ostreococcus tauri	7,987	403	534	528	968	15,392
	Ostreococcus RCC809	7,492	1,095	802	935	1,634	32,956
	Micromonas RCC299	10,109	982	1,223	1,413	11,388	30,984
	Micromonas CCMP2099[c]	19,316	2,648	2,955	1,831	4,495	12,923
	Micromonas CCMP1545	9,702	591	735	702	3,365	7,059
Class I prasinophytes	*Pyramimonas parkeae*[c]	20,299	3,493	3,871	3,391	2,666	15,939
Chlorophytes	*Chlamydomonas reinhardtii*	17,113	3402	3504	3908	3367	16,392
Unassigned green	*Arabidopsis thaliana*	35,386	3,503	4,862	4,126	3,951	25,117
Stramenopiles	*Aureococcus anophagefferens*	11,501	10,283	5,751	12,684	5,913	48,783
	Ectocarpus siliculosus	16,269	3,181	4,110	3,758	3,138	14,123
	Phaeodactylum tricornutum	10,417	1,271	1,851	1,521	6,417	8,220
	Thalassiosira pseudonana	11,673	1,431	2,819	2,984	9,392	16,886
Cryptophytes	*Guillardia theta*	24,840	2,433	2,935	2,260	4,711	11,185
Rhizarians	*Bigelowiella natans*	21,708	2,059	2,661	2,116	2,408	10,961
Haptophytes	*Emiliania huxleyi*	33,340	6,681	6,778	6,425	6,893	22,894

[a] See Materials and Methods; Drift study data from 12 samples in reference 30 were summed and are reported here as a single sample.
[b] The number of proteins predicted in genome projects for each species is shown as the predicted proteome.
[c] The predicted proteome for this taxon is based on assembled transcriptomes available at iMicrobe under project number CAM_P_0001089. Predicted protein counts are likely inflated due to assembly methods.

CO and TZ as well as the OO surface and DCM (OO_{SURF} and OO_{DCM}) samples, with 1.4×10^6 reads on average per sample (Table 2; see also Table S1 in the supplemental material). For the Drift study, 12 published metatranscriptomes were reanalyzed, because the initial publication (30) did not include reference genomes from *Bathycoccus* and several other phytoplankton. The samples were collected between 23 m to 25 m, which was either in the mixed layer or below (Fig. S1). Metatranscriptome reads from the Transect survey and Drift study were assigned to predicted proteomes available for the nuclear genomes of eukaryotic phytoplankton, including two nonmarine species which were used to recruit reads from green algae that are not represented in current genome databases. *Bathycoccus* recruited the most reads in the Transect survey (Table 2). In the Drift study, *Ostreococcus* clade OI recruited the most reads. Multiple *Micromonas* clades were present, and, when their reads were summed, they became the second and third most-represented genera in the CO and Drift metatranscriptome data, respectively (Table 2). Thus, the higher relative abundance of *Ostreococcus* clade OI in the Drift study and of clade OII in the OO, and the generally high abundance of *Bathycoccus* in all samples but the OO_{SURF} (Table 2), matched the patterns observed by qPCR at the corresponding depths (Fig. 2C).

Other taxa were detected in the metatranscriptomic data as well (Table 2; see also Table S1B in the supplemental material). In the OO_{SURF}, haptophyte and pelagophyte sequences dominated (represented by *Emiliania huxleyi* and *Aureococcus anophagefferens*, respectively) (Table 2). Reads assigned to *A. anophagefferens* were present at several sites and likely came from the pelagophyte *Pelagomonas calceolata*, which can be abundant along line 67 (58). Finally, reads assigned to the diatom *Thalassiosira pseudonana* were notable in the CO (Table 2).

In a separate analysis of the metatranscriptomes, distributions of the *Ostreococcus* and *Bathycoccus* ecotypes were analyzed using information from the homolog analysis (Fig. 4A to D). First, we identified 112 genes in the homolog groups that were present in all the picoprasinophyte genomes and targeted metagenomes analyzed (Table S2; see also Data Set S2 in the supplemental material). These 112 genes are here termed ecomarkers. As seen in qPCR data, reads from *O. lucimarinus* (clade OI) ecomarkers dominated the Drift study and in the CO and TZ samples (Fig. 4E). Likewise, *Ostreococcus* clade OII was observed exclusively in the OO. Although some reads were assigned to *O. tauri* clade OI/C in the BLASTX analysis (Table 2; see also Table S1 in the supplemental material), none were assigned to *O. tauri* ecomarkers, reflecting the higher stringency of the nucleotide-based approach. The results provide deep-sequencing support for prior inferences from clone libraries that *O. tauri* is likely restricted to a few bays and lagoons, while *O. lucimarinus* is widespread in mesotrophic/coastal environments (8, 10).

Unlike results for *Ostreococcus*, ecomarkers from *Bathycoccus* clade BI (*B. prasinos*) and clade BII cooccurred in the CO, TZ, and OO_{DCM} samples (Fig. 4F). BII dominated (73%) the OO_{DCM} *Bathycoccus* signal, and BI reads were 27% of total. In contrast, only BI ecomarkers were detected in the Drift study. It should be noted that sample rarefication was not performed (59), and ecotype distributions across samples should be compared with that in mind, especially for samples in which few ecomarkers were recovered (see Table S2 in the supplemental material). We examined the observed patterns further by applying the ecomarkers to 2007 line 67 (cruise CN207) metagenomic data generated at the same time of year (19). In 2007, *Ostreococcus* ecotypes followed the same regional distributions observed in the ecomarker analysis of the metatranscriptomes (and qPCR), where clades OI and OII did not overlap and *O. tauri* was undetected. For *Bathycoccus*, the

Downloaded from http://aem.asm.org/ on March 5, 2019 by guest

FIG 5 Highly expressed *Bathycoccus* proteins of known function that recruited ≥0.1% of OO_{DCM} *Bathycoccus* reads and comparison to the percentage of reads recruited at other sites. Protein identification numbers (ORCAE database) (ID) and InterPro Scan assigned functions (analyzed here) are provided. Only reads recruited to the *Bathycoccus prasinos* protein set (from the sequenced genome) were analyzed. The five shades of blue represent read recruitment percentages in categories (dark to light) of ≥1.00%, <1.00% to ≥0.100%, <0.100% to ≥0.010%, <0.010% to ≥0.001%, and <0.001% to 0.000%.

ecomarker analyses of metagenomic and metatranscriptomic results also agreed, although *Bathycoccus* was not detected in the 2007 OO_{SURF}. Remarkably, proportions of *Bathycoccus* ecotypes were stable between the two line 67 cruises (and the two data types), resulting in 76.2% ± 4.6% BII reads and 23.8% ± 4.5% in the OO_{DCM}, versus 99.5% ± 0.7% and 99.7% ± 0.5% BI reads in the TZ and CO, respectively. Thus, the high *Bathycoccus* abundances observed in our study do not appear to be from a single cosmopolitan type. Rather, the two *Bathycoccus* ecotypes appear to have differential distributions but with greater niche overlap than for the *Ostreococcus* ecotypes.

Exploratory gene expression analyses. Exploratory gene expression analyses were performed for *Bathycoccus* and *Ostreococcus* in our unreplicated metatranscriptomes. These taxa were selected because ≥2 reads (on average) were mapped per predicted protein in the genome, at two or more sites (Table 2; see also Table S1 in supplemental material). In fitting with the cell cycle of phytoplankton and prior metatranscriptome results (60), 12 and 10 photosynthesis-related genes were in the top 100 *Bathycoccus* genes expressed in the TZ and OO_{DCM} (midday) samples, respectively. In the CO and Drift 22:00 (evening) samples, ribosomal proteins were prominent (see also Data Set S3 in the supplemental material). In the OO_{DCM}, 26 of the 100 genes recruiting the most *Bathycoccus* reads were of unknown function (Data Set S3 and Table S4). In the TZ and CO, 19 of the top 100 expressed were of unknown function; for the summed Drift metatranscriptomes, this number was 11. *Ostreococcus* genes recruiting the highest percentage of reads in a sample had functions similar to those from *Bathycoccus* in CO and Drift data (Fig. 5; see also Data Set S4), but *Ostreococcus* reads were too few to be analyzed in the TZ and OO_{DCM}.

We assessed *Bathycoccus* genes that might help to understand acclimation to habitat-specific factors, especially in the OO_{DCM} (see Materials and Methods). Fourteen known-function and 14 unknown-function *Bathycoccus* genes with ≥0.01% read levels in the OO_{DCM} were not detected elsewhere (see Table S3 in the supplemental material). Further, 147 *Bathycoccus* genes each received ≥0.1% OO_{DCM} reads, of which 112 had InterPro-assigned (61) functions (Fig. 5; see also Data Set S3 in the supplemental material). Proteins recruiting ≥0.1% reads in the OO_{DCM} appeared to often have higher relative transcript abundance than in other samples (Fig. 5), including proteins involved in folding and translation (e.g., EF1B, HSP70, HSP20, and a family 10 glycosyl transferase) as well as a nickel superoxide dismutase (NiSOD). Four SOD isozymes are known, each binding different metals (Mn, Fe, Ni, and Cu/Zn) but catalyzing the same dismutation reaction. SODs provide protection from reactive oxygen species (ROS) that are formed during photosynthesis, under stress and during other cellular processes (62). To interpret this finding, we first annotated SODs in the *B. prasinos* and *Ostreococcus* sp. RCC809 genomes and improved previous *O. lucimarinus*, *O. tauri*, *M. pusilla* CCMP1545, and *Micromonas* sp. RCC299 annotations (13, 14). We identified all four SOD metalloforms in the class II prasino-

Downloaded from http://aem.asm.org/ on March 5, 2019 by guest

phyte genomes except *O. tauri*. This includes previously overlooked NiSODs and different Cu/ZnSODs which formed three homolog groups Cu/ZnSOD1, Cu/ZnSOD2, and Cu/ZnSOD3. In the metatranscriptome data, the *Bathycoccus* Mn/FeSOD was represented at equivalent levels (≥0.01% category) (Fig. S2) at all Transect sites but appeared to be less abundant in the Drift study. Results in the CO and the Drift study for *Ostreococcus* mirrored those of *Bathycoccus* (Fig. S2). In contrast, *Bathycoccus* Cu/ZnSOD2 and Cu/ZnSOD3 were found in the CO and the Drift study samples but were not detected in the OO_{DCM} and the TZ (Fig. S2). Although these results must be interpreted with caution, given that the OO_{DCM} and TZ samples had lower *Bathycoccus* read counts than the CO and summed Drift samples, similar field results have been reported for diatom Cu/ZnSODs (63). Moreover, the *Bathycoccus* NiSOD was close to 10-fold higher at the OO_{DCM} than at other sites and had higher relative read counts (again ~10-fold) than the other SOD metalloforms. We quantified Fe in the 2007 transect cruise (CN207) and found that, above 100 m, it declined from ~200 pM in TZ stations to between 50 and 90 pM in the OO (Data Set S1 and Fig. S3). These results point to Ni acquisition and NiSOD usage as a strategy for avoiding Fe dependence (and potentially Mn and Cu dependence) under limiting conditions.

The study region encompassed a strong gradient in photic zone NO_3^- and NH_4^+ concentrations (Table 1; Fig. 2A). We annotated ammonium and nitrate transporters in *B. prasinos* and *Ostreococcus* sp. RCC809 (clade OII), because they had not been characterized previously (see Table S5 in the supplemental material). The six and four NH_4^+ transporters (AMTs) in *B. prasinos* and *O. lucimarinus*, respectively, had read percentages >0.01% in most samples, and two showed expression in the ≥0.1% category (see Fig. S4 in the supplemental material). We then performed a phylogenetic analysis of the AMTs. This showed that four *Bathycoccus* and three *Ostreococcus* sp. RCC809 AMT genes are present in single copies and branch in clades containing known AMTs from *M. pusilla* CCMP1545, *Micromonas* sp. RCC299, *O. lucimarinus*, and *O. tauri* (47) (Fig. S5). The two other *Bathycoccus* AMT1 proteins (AMT1.3a, GenBank accession number CCO19599; AMT1.3b, GenBank accession number CCO17111) form a sister group to *Micromonas* sp. RCC299 AMT1.3, for which close homologs were unknown (47). The overall AMT count in the *B. prasinos* genome is higher than in other genome-sequenced class II prasinophytes (Fig. S5). Expression of AMT1.2, AMT2.1 (plastid targeted), and AMT2.2 was low or undetectable in the OO_{DCM}, while AMT1.1, AMT1.3a, and AMT1.3b generally contributed >0.02% of total *Bathycoccus* reads across the regimes (Fig. S4) (note, again, there were lower overall *Bathycoccus* reads in the OO_{DCM} and TZ than in the CO and summed Drift, which could influence reliability of results from genes with low read counts in the OO_{DCM} and TZ). The latter two transporters showed read percentages of 0.026 ± 0.015 across the sites, indicating use under a variety of nitrogen concentrations (Table 1; see also Data Set S1 in the supplemental material). AMT2.2 relative read percentages were higher in the Drift study (<0.100% to ≥0.010% category) and the CO (≥0.100% category), whereas NH_4^+ concentrations were higher than the OO_{DCM} and TZ, in which *Bathycoccus* AMT2.2 percentages were lower (0.006% and <0.001%, respectively) (Fig. S4; see also Data Set S1). Thus, AMT2.2 is likely a low-affinity transporter that could serve as an indicator of organismal thresholds with respect to high NH_4^+ availability.

We identified a putative low-affinity NO_3^- permease (NRT1) and a high-affinity NO_3^- transporter (NRT2) in the *B. prasinos* and *Ostreococcus* sp. RCC809 genomes by homology (see Table S5 in the supplemental material). Affinity assignments for class II prasinophyte transporters have thus far been based only on homology to verified transporters in, e.g., *Chlamydomonas reinhardtii* and land plants. Here, *Bathycoccus* NRT1 and NRT2.1 expression was observed in all samples except the CO (see Fig. S4 in the supplemental material). However, relative read percentages for these genes were low (i.e., in the <0.010% to ≥0.001% category) (Fig. S4), except at the OO_{DCM}, where the NRT2.1 read percentage was 0.029% and the concentration of NO_3^- was two orders of magnitude lower (<0.010 μM) than at other stations (Table 1). *Ostreococcus* NRT2.1 showed no expression in the CO and the Drift study. Thus, we surmise that NRT2.1 is a high-affinity NO_3^- transporter and provides a useful indicator of organism thresholds for low nitrate availability.

Conclusions. Picoprasinophytes are often abundant in coastal zones, and we show that *Bathycoccus*, *Micromonas*, and *Ostreococcus* clade OI can be abundant in the eastern North Pacific mesotrophic transition zone. Overall, *Ostreococcus* clade OI showed high 18S gene counts in the upwelling-influenced and coastal waters, but not elsewhere. *Micromonas* and *Ostreococcus* clade OI cooccurred. *Bathycoccus* was present at all sites and abundant even at the DCM in the oligotrophic ocean. However, the *Bathycoccus* cells enumerated by qPCR do not come from *B. prasinos* alone but, rather, are from two ecotypes, named here BI and BII, based on the new ecomarker methodology we established for analyzing these groups. Prior reported sequence variations in Chilean *Bathycoccus* targeted metagenomes were initially thought to represent other ecotypes but likely represent single nucleotide polymorphisms (frequency of <3%) present in the BI ecotype population. Furthermore, ecomarker analysis demonstrated that *O. lucimarinus*-like strains dominate the clade OI signal and that clade OI/C (*O. tauri*-like) is not present. Similar to previous studies, the *Ostreococcus* clade OI and clade OII ecotypes showed distinct spatial distributions related to system nutrient levels (10). The *Bathycoccus* ecotypes overlap more frequently and are present over larger gradients in NH_4^+ and NO_3^- concentrations, extending to the oligotrophic ocean. Our studies indicate that NRT2.1 is a high-affinity nitrate transporter and that AMT2.2 is a low-affinity ammonium transporter in the class II prasinophytes. Two additional AMT transporters in *Bathycoccus* that are absent from its closest relative (*Ostreococcus*) appear to be persistently expressed in the samples we studied that extended to the open ocean. Collectively, these results provide targets for future field analyses and testable hypotheses for the taxa that are in culture.

ACKNOWLEDGMENTS

We thank the captain and crew of the R/V *Western Flyer*, the shore-side marine operations support staff, as well as F. P. Chavez, E. Demir-Hilton, L. Deng, D. McRose, and A. M. Gehman. We also thank K. S. Johnson for making iron measurements possible. We thank A. E. Allen for providing the Global Ocean Survey extraction procedures. We thank M. J. van Baren for assistance in parsing metatranscriptomic data. We are grateful to M. B. Kogut and to the anonymous reviewers for thoughtful comments on the manuscript.

This research was supported by the David and Lucile Packard Foundation as well as by DOE grants DE-SC0004765, NSF-IOS0843119, and GBMF3788 to A.Z.W.

Downloaded from http://aem.asm.org/ on March 5, 2019 by guest

REFERENCES

1. **Field CB, Behrenfeld MJ, Randerson JT, Falkowski P.** 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. Science **281:**237–240. http://dx.doi.org/10.1126/science.281.5374.237.
2. **Li WKW.** 1994. Primary production of prochlorophytes, cyanobacteria, and eukaryotic ultraplankton: measurements from flow cytometric sorting. Limnol Oceanogr **39:**169–175. http://dx.doi.org/10.4319/lo.1994.39.1.0169.
3. **Worden AZ, Nolan JK, Palenik B.** 2004. Assessing the dynamics and ecology of marine picophytoplankton: the importance of the eukaryotic component. Limnol Oceanogr **49:**168–179. http://dx.doi.org/10.4319/lo.2004.49.1.0168.
4. **Grob C, Ulloa O, Claustre H, Huot Y, Alarcon G, Marie D.** 2007. Contribution of picoplankton to the total particulate organic carbon concentration in the eastern South Pacific. Biogeosciences **4:**837–852. http://dx.doi.org/10.5194/bg-4-837-2007.
5. **Cuvelier ML, Allen AE, Monier A, McCrow JP, Messié M, Tringe SG, Woyke T, Welsh R, Ishoey T, Lee JH, Binder BJ, Dupont CL, Latasa M, Guigand C, Buck KR, Hilton J, Thiagarajan M, Caler E, Read B, Lasken RS, Chavez FP, Worden AZ.** 2010. Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. Proc Natl Acad Sci U S A **107:**14679–14684. http://dx.doi.org/10.1073/pnas.1001665107.
6. **Jardillier L, Zubkov MV, Pearman J, Scanlan DJ.** 2010. Significant CO_2 fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. ISME J **4:**1180–1192. http://dx.doi.org/10.1038/ismej.2010.36.
7. **Vaulot D, Eikrem W, Viprey M, Moreau H.** 2008. The diversity of small eukaryotic phytoplankton (≤3 μm) in marine ecosystems. FEMS Microbiol Rev **32:**795–820. http://dx.doi.org/10.1111/j.1574-6976.2008.00121.x.
8. **Worden AZ, Not F.** 2008. Ecology and diversity of picoeukaryotes, p 594. *In* Kirchman DL (ed), Microbial ecology of the oceans, 2nd ed. Wiley, Hoboken, NJ.
9. **de Vargas C, Audic S, Henry N, Decelle J, Mahe F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, Carmichael M, Poulain J, Romac S, Colin S, Aury JM, Bittner L, Chaffron S, Dunthorn M, Engelen S, Flegontova O, Guidi L, Horak A, Jaillon O, Lima-Mendez G, Lukes J, Malviya S, Morard R, Mulot M, Scalco E, Siano R, Vincent F, Zingone A, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Tara Oceans C, Acinas SG, Bork P, Bowler C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Raes J, Sieracki ME, Speich S, et al.** 2015. Ocean plankton: eukaryotic plankton diversity in the sunlit ocean. Science **348:**1261605.
10. **Demir-Hilton E, Sudek S, Cuvelier ML, Gentemann C, Zehr JP, Worden AZ.** 2011. Global distribution patterns of distinct clades of the photosynthetic picoeukaryote *Ostreococcus*. ISME J **5:**1095–1107. http://dx.doi.org/10.1038/ismej.2010.209.
11. **Foulon E, Not F, Jalabert F, Cariou T, Massana R, Simon N.** 2008. Ecological niche partitioning in the picoplanktonic green alga *Micromonas pusilla*: evidence from environmental surveys using phylogenetic probes. Environ Microbiol **10:**2433–2443. http://dx.doi.org/10.1111/j.1462-2920.2008.01673.x.
12. **Simmons MP, Bachy C, Sudek S, van Baren MJ, Sudek L, Ares M, Jr, Worden AZ.** 2015. Intron invasions trace algal speciation and reveal nearly identical Arctic and Antarctic *Micromonas* populations. Mol Biol Evol **32:**2219–2235. http://dx.doi.org/10.1093/molbev/msv122.
13. **Worden AZ, Lee JH, Mock T, Rouze P, Simmons MP, Aerts AL, Allen AE, Cuvelier ML, Derelle E, Everett MV, Foulon E, Grimwood J, Gundlach H, Henrissat B, Napoli C, McDonald SM, Parker MS, Rombauts S, Salamov A, Von Dassow P, Badger JH, Coutinho PM, Demir E, Dubchak I, Gentemann C, Eikrem W, Gready JE, John U, Lanier W, Lindquist EA, Lucas S, Mayer KF, Moreau H, Not F, Otillar R, Panaud O, Pangilinan J, Paulsen I, Piegu B, Poliakov A, Robbens S, Schmutz J, Toulza E, Wyss T, Zelensky A, Zhou K, Armbrust EV, Bhattacharya D, Goodenough UW, Van de Peer Y, et al.** 2009. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. Science **324:**268–272. http://dx.doi.org/10.1126/science.1167222.
14. **Palenik B, Grimwood J, Aerts A, Rouze P, Salamov A, Putnam N, Dupont C, Jorgensen R, Derelle E, Rombauts S, Zhou K, Otillar R, Merchant SS, Podell S, Gaasterland T, Napoli C, Gendler K, Manuell A, Tai V, Vallon O, Piganeau G, Jancek S, Heijde M, Jabbari K, Bowler C, Lohr M, Robbens S, Werner G, Dubchak I, Pazour GJ, Ren Q, Paulsen I, Delwiche C, Schmutz J, Rokhsar D, Van de Peer Y, Moreau H, Grigoriev IV.** 2007. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. Proc Natl Acad Sci U S A **104:**7705–7710. http://dx.doi.org/10.1073/pnas.0611046104.
15. **Moreau H, Verhelst B, Couloux A, Derelle E, Rombauts S, Grimsley N, Van Bel M, Poulain J, Katinka M, Hohmann-Marriott M, Piganeau G, Rouze P, Da Silva C, Wincker P, Van de Peer Y, Vandapoele K.** 2012. Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. Genome Biol **13:**R74. http://dx.doi.org/10.1186/gb-2012-13-8-r74.
16. **Guillou L, Eikrem W, Chretiennot-Dinet M, Le Gall F, Massana R, Romari K, Pedros-Alio C, Vaulot D.** 2004. Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. Protist **155:**193–214. http://dx.doi.org/10.1078/143446104774199592.
17. **Worden AZ.** 2006. Picoeukaryote diversity in coastal waters of the Pacific Ocean. Aquat Microb Ecol **43:**165–175. http://dx.doi.org/10.3354/ame043165.
18. **Monier A, Sudek S, Fast NM, Worden AZ.** 2013. Gene invasion in distant eukaryotic lineages: discovery of mutually exclusive genetic elements reveals marine biodiversity. ISME J **7:**1764–1774. http://dx.doi.org/10.1038/ismej.2013.70.
19. **Monier A, Welsh RM, Gentemann C, Weinstock G, Sodergren E, Armbrust EV, Eisen JA, Worden AZ.** 2012. Phosphate transporters in marine phytoplankton and their viruses: cross-domain commonalities in viral-host gene exchanges. Environ Microbiol **14:**162–176. http://dx.doi.org/10.1111/j.1462-2920.2011.02576.x.
20. **Vaulot D, Lepere C, Toulza E, De la Iglesia R, Poulain J, Gaboyer F, Moreau H, Vandepoele K, Ulloa O, Gavory F, Piganeau G.** 2012. Metagenomes of the picoalga *Bathycoccus* from the Chile Coastal Upwelling. PLoS One **7:**e39648. http://dx.doi.org/10.1371/journal.pone.0039648.
21. **Not F, Massana R, Latasa M, Marie D, Colson C, Eikrem W, Pedros-Alio C, Vaulot D, Simon N.** 2005. Late summer community composition and abundance of photosynthetic picoeukaryotes in Norwegian and Barents Seas. Limnol Oceanogr **50:**1677–1686. http://dx.doi.org/10.4319/lo.2005.50.5.1677.
22. **Countway PD, Caron DA.** 2006. Abundance and distribution of *Ostreococcus* sp. in the San Pedro Channel, California, as revealed by quantitative PCR. Appl Environ Microbiol **72:**2496–2506. http://dx.doi.org/10.1128/AEM.72.4.2496-2506.2006.
23. **Collado-Fabbri S, Vaulot D, Ulloa O.** 2011. Structure and seasonal dynamics of the eukaryotic picophytoplankton community in a wind-driven coastal upwelling ecosystem. Limnol Oceanogr **56:**2334–2346. http://dx.doi.org/10.4319/lo.2011.56.6.2334.
24. **Olson RL, Zettler ER, DuRand MD.** 1993. Phytoplankton analysis using flow cytometry, p 175–186. *In* Kemp PF, Sherr BF, Sherr EB, Cole JJ (ed), Handbook of methods in aquatic ecology. Lewis Publishers, Boca Raton, FL.
25. **Paerl RW, Johnson KS, Welsh RM, Worden AZ, Chavez FP, Zehr JP.** 2011. Differential distributions of synechococcus subgroups across the California current system. Front Microbiol **2:**59.
26. **Orsi WD, Smith JM, Wilcox HM, Swalwell JE, Carini P, Worden AZ, Santoro AE.** 2015. Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. ISME J **9:**1747–1763. http://dx.doi.org/10.1038/ismej.2014.260.
27. **Plant JN, Johnson KS, Needoba JA, Coletti LJ.** 2009. NH_4-Digiscan: an *in situ* and laboratory ammonium analyzer for estuarine, coastal, and shelf waters. Limnol Oceanogr Methods **7:**144–156. http://dx.doi.org/10.4319/lom.2009.7.144.
28. **Pennington JT, Chavez FP.** 2000. Seasonal fluctuations of temperature, salinity, nitrate, chlorophyll and primary production at station H3/M1 over 1989 to 1996 in Monterey Bay, California. Deep Sea Res Part II Top Stud Oceanogr **47:**947–973. http://dx.doi.org/10.1016/S0967-0645(99)00132-0.
29. **Johnson KS, Elrod VA, Fitzwater SE, Plant JN, Chavez FP, Tanner SJ, Gordon RM, Westphal DL, Perry KD, Wu JF, Karl DM.** 2003. Surface ocean-lower atmosphere interactions in the northeast Pacific Ocean gyre: aerosols, iron, and the ecosystem response. Global Biogeochem Cycles **17:**1063. http://dx.doi.org/10.1029/2002GB002004.

Downloaded from http://aem.asm.org/ on March 5, 2019 by guest

30. **Ottesen EA, Young CR, Eppley JM, Ryan JP, Chavez FP, Scholin CA, DeLong EF.** 2013. Pattern and synchrony of gene expression among sympatric marine microbial populations. Proc Natl Acad Sci U S A **110:**E488–E497. http://dx.doi.org/10.1073/pnas.1222099110.
31. **Santoro AE, Sakamoto CM, Smith JM, Plant JN, Gehman AL, Worden AZ, Johnson KS, Francis CA, Casciotti KL.** 2013. Measurements of nitrite production and nitrite-producing organisms in and around the primary nitrite maximum in the central California current. Biogeosciences **10:**7395–7410. http://dx.doi.org/10.5194/bg-10-7395-2013.
32. **Sudek S, Everroad RC, Gehman AM, Smith JM, Poirier CL, Chavez FP, Worden AZ.** 2015. Cyanobacterial distributions along a physicochemical gradient in the northeastern Pacific Ocean. Environ Microbiol **17:**3692–3707. http://dx.doi.org/10.1111/1462-2920.12742.
33. **Chisholm SW, Armbrust EV, Olson RJ.** 1986. The individual cell in phytoplankton ecology: cell cycles and flow cytometry. Can Bull Fish Aquat Sci **214:**343–369.
34. **Olson RJ, Zettler ER, Chisholm SW, Dusenberry JA.** 1991. Advances in oceanography through flow cytometry, p 351–399. *In* Demers S (ed), Particle analysis in oceanography. Springer-Verlag, Berlin, Germany.
35. **Derelle E, Ferraz C, Rombauts S, Rouze P, Worden AZ, Robbens S, Partensky F, Degroeve S, Echeynie S, Cooke R, Saeys Y, Wuyts J, Jabbari K, Bowler C, Panaud O, Piegu B, Ball SG, Ral JP, Bouget FY, Piganeau G, De Baets B, Picard A, Delseny M, Demaille J, Van de Peer Y, Moreau H.** 2006. From the cover: genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. Proc Natl Acad Sci U S A **103:**11647–11652. http://dx.doi.org/10.1073/pnas.0604795103.
36. **Sambrook J, Russell DW.** 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
37. **Stewart FJ, Ottesen EA, DeLong EF.** 2010. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. ISME J **4:**896–907. http://dx.doi.org/10.1038/ismej.2010.18.
38. **Li L, Stoeckert CJ, Jr, Roos DS.** 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res **13:**2178–2189. http://dx.doi.org/10.1101/gr.1224503.
39. **Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.
40. **Notredame C, Higgins DG, Heringa J.** 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol **302:**205–217. http://dx.doi.org/10.1006/jmbi.2000.4042.
41. **R Development Core Team.** 2015. R: a language and environment for statistical computing, v3. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
42. **Huang Y, Niu B, Gao Y, Fu L, Li W.** 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics **26:**680–682. http://dx.doi.org/10.1093/bioinformatics/btq003.
43. **Gomez-Alvarez V, Teal TK, Schmidt TM.** 2009. Systematic artifacts in metagenomes from complex microbial communities. ISME J **3:**1314–1317. http://dx.doi.org/10.1038/ismej.2009.72.
44. **Schmieder R, Lim YW, Edwards R.** 2012. Identification and removal of ribosomal RNA sequences from metatranscriptomes. Bioinformatics **28:**433–435. http://dx.doi.org/10.1093/bioinformatics/btr669.
45. **Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, Goodstein D, Kuo A, Minovitsky S, Nikitin R, Ohm RA, Otillar R, Poliakov A, Ratnere I, Riley R, Smirnova T, Rokhsar D, Dubchak I.** 2012. The genome portal of the Department of Energy Joint Genome Institute. Nucleic Acids Res **40:**D26-32. http://dx.doi.org/10.1093/nar/gkr947.
46. **Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Mizrachi I, Ostell J, Panchenko A, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Slotta D, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Wang Y, John Wilbur W, Yaschenko E, Ye J.** 2010. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res **38:**D5–D16. http://dx.doi.org/10.1093/nar/gkp967.
47. **McDonald SM, Plant JN, Worden AZ.** 2010. The mixed lineage nature of nitrogen transport and assimilation in marine eukaryotic phytoplankton: a case study of Micromonas. Mol Biol Evol **27:**2268–2283. http://dx.doi.org/10.1093/molbev/msq113.
48. **Felsenstein J.** 2005. PHYLIP (phylogeny inference package) version 3.6. Department of Genome Sciences, University of Washington, Seattle, WA.
49. **Guindon S, Gascuel O.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **52:**696–704. http://dx.doi.org/10.1080/10635150390235520.
50. **Collins CA, Pennington JT, Castro CG, Rago TA, Chavez FP.** 2003. The California Current system off Monterey, California: physical and biological coupling. Deep Sea Res Part II Top Stud Oceanogr **50:**2389–2404. http://dx.doi.org/10.1016/S0967-0645(03)00134-6.
51. **Rodríguez F, Derelle E, Guillou L, Le Gall F, Vaulot D, Moreau H.** 2005. Ecotype diversity in the marine picoeukaryote *Ostreococcus* (Chlorophyta, Prasinophyceae). Environ Microbiol **7:**853–859. http://dx.doi.org/10.1111/j.1462-2920.2005.00758.x.
52. **Six C, Finkel ZV, Rodriguez F, Marie D, Partensky F, Campbell DA.** 2008. Contrasting photoacclimation costs in ecotypes of the marine eukaryotic picoplankter *Ostreococcus*. Limnol Oceanogr **53:**255–265. http://dx.doi.org/10.4319/lo.2008.53.1.0255.
53. **Cardol P, Bailleul B, Rappaport F, Derelle E, Beal D, Breyton C, Bailey S, Wollman FA, Grossman A, Moreau H, Finazzi G.** 2008. An original adaptation of photosynthesis in the marine green alga *Ostreococcus*. Proc Natl Acad Sci U S A **105:**7881–7886. http://dx.doi.org/10.1073/pnas.0802762105.
54. **Zhu F, Massana R, Not F, Marie D, Vaulot D.** 2005. Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. FEMS Microbiol Ecol **52:**79–92. http://dx.doi.org/10.1016/j.femsec.2004.10.006.
55. **Not F, Latasa M, Marie D, Cariou T, Vaulot D, Simon N.** 2004. A single species, *Micromonas pusilla* (*Prasinophyceae*), dominates the eukaryotic picoplankton in the western English Channel. Appl Environ Microbiol **70:**4064–4072. http://dx.doi.org/10.1128/AEM.70.7.4064-4072.2004.
56. **Treusch AH, Demir-Hilton E, Vergin KL, Worden AZ, Carlson CA, Donatz MG, Burton RM, Giovannoni SJ.** 2012. Phytoplankton distribution patterns in the northwestern Sargasso Sea revealed by small subunit rRNA genes from plastids. ISME J **6:**481–492. http://dx.doi.org/10.1038/ismej.2011.117.
57. **Coleman AW.** 2009. Is there a molecular key to the level of biological species in eukaryotes? Mol Phylogenet Evol **50:**197–203. http://dx.doi.org/10.1016/j.ympev.2008.10.008.
58. **Worden AZ, Janouskovec J, McRose D, Engman A, Welsh RM, Malfatti S, Tringe SG, Keeling PJ.** 2012. Global distribution of a wild alga revealed by targeted metagenomics. Curr Biol **22:**R675-677. http://dx.doi.org/10.1016/j.cub.2012.07.054.
59. **McMurdie PJ, Holmes S.** 2014. Waste not, want not: why rarefying microbiome data are inadmissible. PLoS Comput Biol **10:**e1003531. http://dx.doi.org/10.1371/journal.pcbi.1003531.
60. **Marchetti A, Varela DE, Lance VP, Johnson Z, Palmucci M, Giordano M, Armbrust EV.** Iron and silicic acid effects on phytoplankton productivity, diversity, and chemical composition in the central equatorial Pacific Ocean. Limnol Oceanogr **55:**11–29.
61. **Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S.** 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics **30:**1236–1240. http://dx.doi.org/10.1093/bioinformatics/btu031.
62. **Wolfe-Simon F, Grzebyk D, Schofield O, Falkowski PG.** 2005. The role and evolution of superoxide dismutases in algae. J Phycol **41:**453–465. http://dx.doi.org/10.1111/j.1529-8817.2005.00086.x.
63. **Marchetti A, Schruth DM, Durkin CA, Parker MS, Kodner RB, Berthiaume CT, Morales R, Allen AE, Armbrust EV.** 2012. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to various iron availability. Proc Natl Acad Sci U S A **109:**E317–E325. http://dx.doi.org/10.1073/pnas.1118408109.

Downloaded from http://aem.asm.org/ on March 5, 2019 by guest