

## ORIGINAL ARTICLE

# Expression patterns of elemental cycling genes in the Amazon River Plume

Brandon M Satinsky<sup>1,5</sup>, Christa B Smith<sup>2</sup>, Shalabh Sharma<sup>2</sup>, Marine Landa<sup>2</sup>, Patricia M Medeiros<sup>2</sup>, Victoria J Coles<sup>3</sup>, Patricia L Yager<sup>2</sup>, Byron C Crump<sup>4</sup> and Mary Ann Moran<sup>2</sup>

<sup>1</sup>Department of Microbiology, University of Georgia, Athens, GA, USA; <sup>2</sup>Department of Marine Sciences, University of Georgia, Athens, GA, USA; <sup>3</sup>Horn Point Laboratory, University of Maryland Center for Environmental Science, Cambridge, MD, USA and <sup>4</sup>College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA

**Metatranscriptomics and metagenomics data sets benchmarked with internal standards were used to characterize the expression patterns for biogeochemically relevant bacterial and archaeal genes mediating carbon, nitrogen, phosphorus and sulfur uptake and metabolism through the salinity gradient of the Amazon River Plume. The genes were identified in 48 metatranscriptomic and metagenomic data sets summing to >500 million quality-controlled reads from six locations in the plume ecosystem. The ratio of transcripts per gene copy (a direct measure of expression made possible by internal standard additions) showed that the free-living bacteria and archaea exhibited only small changes in the expression levels of biogeochemically relevant genes through the salinity and nutrient zones of the plume. In contrast, the expression levels of genes in particle-associated cells varied over orders of magnitude among the stations, with the largest differences measured for genes mediating aspects of nitrogen cycling (*nifH*, *amtB* and *amoA*) and phosphorus acquisition (*pstC*, *phoX* and *phoU*). Taxa varied in their baseline gene expression levels and extent of regulation, and most of the spatial variation in the expression level could be attributed to changes in gene regulation after removing the effect of shifting taxonomic composition. We hypothesize that changes in microbial element cycling along the Amazon River Plume are largely driven by shifting activities of particle-associated cells, with most activities peaking in the mesohaline regions where N<sub>2</sub> fixation rates are elevated.**

*The ISME Journal* (2017) 11, 1852–1864; doi:10.1038/ismej.2017.46; published online 7 April 2017

## Introduction

The Amazon River is the world's largest in terms of discharge, releasing more freshwater into the ocean than the next six largest rivers combined. Terrestrially derived nutrients are carried offshore in the plume and mix into the western tropical North Atlantic Ocean in amounts sufficient to impact global marine primary productivity and carbon sequestration (Richey *et al.*, 1989; Subramaniam *et al.*, 2008; Coles *et al.*, 2013).

Previous studies of the Amazon Plume microbiome have shown that primary production in low salinity waters is dominated by coastal diatoms whose activity is supported by river-supplied nutrients once riverine turbidity declines (Smith and

Demaster, 1996; Simon *et al.*, 2009; Goes *et al.*, 2014). As salinity increases and nitrogen becomes more limited along the plume, diatoms *Rhizosolenia* and *Hemiaulus* become dominant (Subramaniam *et al.*, 2008; Goes *et al.*, 2014), in part because they carry a symbiotic N<sub>2</sub>-fixing cyanobacterium *Richelia intracellularis* in an arrangement referred to as a diatom–diazotroph association (DDA) (Goebel *et al.*, 2010; Foster *et al.*, 2011). In oceanic regions where river-supplied phosphorus and silicate are both depleted, N<sub>2</sub>-fixing *Trichodesmium* become important (Subramaniam *et al.*, 2008; Goes *et al.*, 2014). The high productivity of these autotrophs generates a region of the western tropical North Atlantic that takes up an excess of ~30 Tg C per year (Cooley *et al.*, 2007; Subramaniam *et al.*, 2008).

Heterotrophic members of river plumes similarly have a major influence over biogeochemistry through roles in respiration and sequestration of fixed carbon (Amon and Benner, 1998; Dagg *et al.*, 2004) and regeneration and uptake of inorganic nitrogen and phosphorus (Gardner *et al.*, 1994; Dagg and Breed, 2003). These important processes

Correspondence: MA Moran, Department of Marine Sciences, University of Georgia, Athens, GA 30605, USA.

E-mail: mmoran@uga.edu

<sup>5</sup>Current address: Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

Received 23 January 2017; accepted 16 February 2017; published online 7 April 2017

**Table 1** Metagenome and metatranscriptome data summaries for six Amazon River Plume Stations in June 2010 sampled at a depth of 2 m

	Station 10	Station 23	Station 3	Station 2	Station 25	Station 27
Location (°lat, long)	4.88, -51.36	10.68, -54.42	7.29, -53.00	10.29, -54.51	11.31, -56.43	12.41, -52.22
Salinity	22.6	26.5	30.8	31.8	31.9	36.0
<i>Raw reads</i>						
Metagenomic ( $\times 10^7$ )	4.44	5.69	7.39	5.08	7.87	6.01
Metatranscriptomic ( $\times 10^8$ )	1.25	1.38	1.00	1.27	1.27	1.95
<i>Joined reads after QC</i>						
Metagenomic ( $\times 10^7$ )	1.15	1.24	1.39	1.72	1.88	1.64
Metatranscriptomic ( $\times 10^7$ )	1.95	3.07	1.77	2.49	3.48	3.49
Mean read length (bp)	192	200	197	191	205	196
<i>Protein encoding reads</i>						
Metagenomic ( $\times 10^6$ )	4.98	2.31	5.55	6.38	5.86	6.77
Metatranscriptomic ( $\times 10^6$ )	4.08	3.21	2.36	3.53	2.41	6.52
<i>Bacteria and Archaea gene copies <math>l^{-1}</math> (<math>\times 10^{12}</math>)</i>						
% FL	68.3%	98.2%	97.3%	93.0%	99.2%	92.1%
% PA	31.8%	1.8%	2.7%	7.0%	0.79%	7.9%
<i>Bacteria and Archaea transcripts <math>l^{-1}</math> (<math>\times 10^{11}</math>)</i>						
% FL	49.3%	93.8%	39.7%	21.5%	64.4%	84.3%
% PA	50.8%	6.2%	60.3%	78.5%	35.6%	15.7%

Abbreviations: FL, free-living; PA, particle-associated. QC, quality control; Sequence data descriptions represent the sum of the FL and PA fractions, each averaged across duplicate samples of each. Per liter calculations are based on recovery of internal standards (Supplementary Table S1).

have been more difficult to investigate compared with the activities of autotrophs, however, because of the hundreds of largely uncharacterized bacterial and archaeal species involved (Pedrós-Alió, 2006) and the diversity of links in the elemental cycles they mediate (Moran *et al.*, 2016). Further, while heterotrophic taxa can harbor the capability for multiple biogeochemical roles, whether these are expressed and at what level is often controlled by the environmental conditions at a given time and place.

Direct measures of transcription of biogeochemically relevant genes can fill key gaps in our understanding of ongoing microbial functions and their regulation in the Amazon Plume microbiome, particularly when data sets are benchmarked with internal standards to generate quantitative gene and transcript inventories (genes or transcripts per liter; transcripts per gene copy) rather than proportional abundances (percentage of metagenome or metatranscriptome) (Gifford *et al.*, 2011; Satinsky *et al.*, 2013; Hilton *et al.*, 2014; Smets *et al.*, 2016). This approach is analogous to carrying out both quantitative PCR and reverse-transcription quantitative PCR on all functional genes of interest in a microbial system. In the few cases where PCR-based gene copy-normalized transcription data have been reported for natural communities, they provided unique information on shifts in gene regulation (Varaljay *et al.*, 2015) and spatial and temporal patterns of microbial activity (Church *et al.*, 2010).

For this study, six locations along the Amazon Plume varying in salinity from 23 to 36 were

inventoried in May–June 2010, and replicated, quantitative and taxon-specific gene and transcript abundances in the metagenomes and metatranscriptomes were calculated for 40 biogeochemically relevant functional genes. This data set provides measures of microbial gene expression levels relevant to carbon (C), nitrogen (N), phosphorus (P) and sulfur (S) cycling and examines these patterns in the context of environmental conditions in three salinity zones of the Earth's largest river plume.

## Materials and methods

Surface samples were obtained in duplicate for two discrete size fractions at six stations (Table 1). Collection and processing were carried out as previously described (Satinsky *et al.*, 2014a, b). In brief, microbial cells from two size fractions (2–156  $\mu\text{m}$ , particle-associated; and 0.2–2  $\mu\text{m}$ , free-living) were collected by filtration from a depth of 2 m and preserved in RNAlater (Applied Biosystems, Austin, TX, USA). RNA for metatranscriptomic analysis was extracted, residual DNA was removed and ribosomal RNA (rRNA) was depleted. The mRNA-enriched RNA was then linearly amplified and converted to cDNA. DNA for metagenomic analysis was extracted and residual proteins and RNA were removed. For both nucleic acid types, internal standards in the form of artificial mRNAs (metatranscriptomic samples) or genomic DNA from *Thermus thermophilus* HB8 (metagenomic samples)

were added during extraction, immediately prior to cell lysis (Satinsky *et al.*, 2013). Methods for internal standard additions can be found at protocols.io (dx.doi.org/10.17504/protocols.io.ftgbnjw and dx.doi.org/10.17504/protocols.io.ffwbjpe).

Following processing, nucleic acid preparations were sheared ultrasonically to ~200–250 bp fragments and libraries were constructed for Illumina sequencing (150 PE; Illumina Inc., San Diego, CA, USA). A bioinformatic pipeline was established for quality control of the Illumina reads, including joining and trimming paired-end sequences and quantifying and removing internal standard sequences (Satinsky *et al.*, 2014a, b). rRNA reads were identified in the metatranscriptomes by a blastn search against a database containing representative rRNA sequences. Protein-encoding reads were identified by RAPSearch2 (Zhao *et al.*, 2012) against the RefSeq protein database, and best hits were used to assign reads to the closest reference genome. A blastx against a custom database consisting of biogeochemically relevant gene sequences was used to identify reads for selected steps in element cycling (Satinsky *et al.*, 2014b; Zielinski *et al.*, 2016). Per liter numbers of each gene and transcript type were calculated based on internal standard recovery (Supplementary Table S1) by comparing the abundance of internal standard reads in the Illumina libraries to the number of standards added at the initiation of nucleic acid extraction (Satinsky *et al.*, 2014a, b). Biological replicates were averaged and the expression ratios and s.d. were calculated as previously described (Satinsky *et al.*, 2014a, b). For cases in which transcripts were observed but corresponding gene counts were zero (4.7% of the 480 expression ratio calculations), minimum estimates of expression ratios were made by setting the gene count equal to the limit of detection; the detection limit differed among samples based on sequencing depth (Supplementary Table S1).

To analyze population structure in selected reference genome bins, ribosomal protein-encoding reads from five taxa abundant in the metatranscriptomes were identified through a similarity search using RAPsearch2 against the full NCBI RefSeq protein database. For the two most highly expressed ribosomal proteins within a genome bin, reads were pooled from all stations and size fractions and clustered at 95% nucleotide identity using CD-HIT (Huang *et al.*, 2010). Reads from each individual station and size fraction were then separately classified into the resultant clusters using blastn against the representative sequence with a minimum alignment length of 50 nt. Even a completely homogenous population will produce multiple read clusters within each gene because the 225 nt reads cover only a part of the gene. However, the cluster pattern will be consistent across samples provided the population structure is consistent. Singletons (clusters with only a single hit across all samples) were removed from subsequent analyses, as were any stations or size fractions with

<20 reads across all clusters. Distributions of remaining read clusters at each station/fraction were compared using a Bray–Curtis dissimilarity analysis to identify homogeneous genome bins defined here as those with dissimilarity scores of  $\leq 30\%$ . The final dissimilarity score was based on the average of the two ribosomal proteins analyzed for each bin.

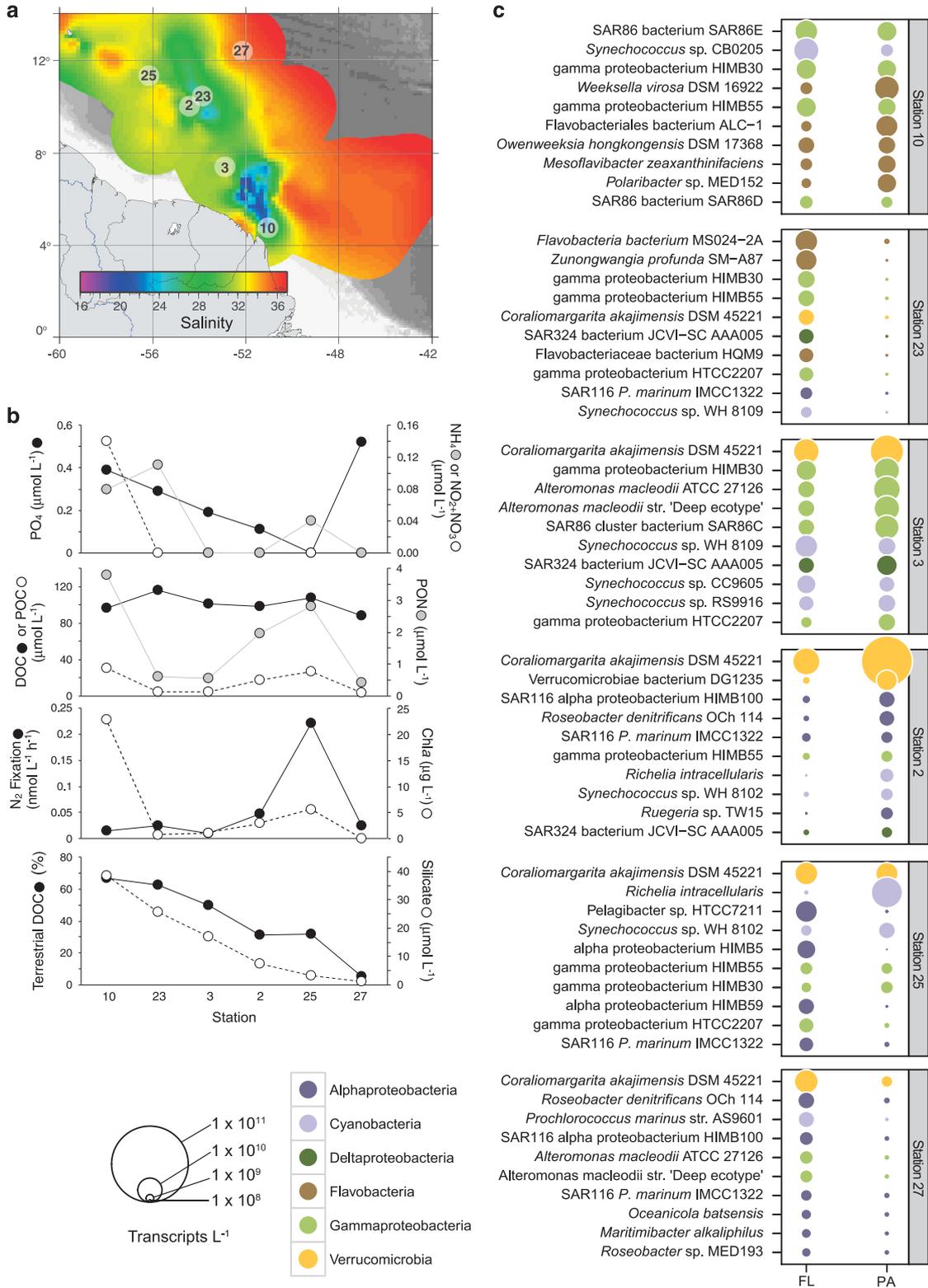
To estimate the number of cells containing a gene of interest, genome equivalents were calculated for a sample by obtaining counts for three housekeeping genes (*recA*, *rpoB* and *dnaK*), normalizing the counts to the gene length of *recA* and then averaging the normalized counts across the three genes (Howard *et al.*, 2006). As each bacterial or archaeal genome is expected to contain one copy of each housekeeping gene, this value estimates the number of genomes sampled. Biogeochemically relevant gene counts were then expressed as a percentage of the average single-copy gene count (Supplementary Table S2). This calculation overestimated the number of cells containing a gene if the gene was present in more than one copy per genome, although this is unlikely for most of the genes analyzed.

Expression and transcript inventories were compared for the two size fractions by calculating free-living:particle-associated ratios for each gene at each station (average in the free-living community/average in the particle-associated community for both transcripts gene copy<sup>-1</sup> and transcripts l<sup>-1</sup>).

The maximal information-based non-parametric data exploration program (MINE) (Reshef *et al.*, 2011) was used to identify gene expression ratios having strong associations with environmental parameters. The program analyzes scatterplots of pair-wise data sets (in this case, each gene expression ratio data set against each environmental parameter data set) to find the grid with the most mutual information (that is, when knowing one variable provides the most information about the other). MINE finds both linear and non-linear associations.

Metagenomic and metatranscriptomic sequences from the May–June 2010 Amazon Continuum study are available from NCBI under accession numbers SRP039390 (metagenomes) and SRP037995 (metatranscriptomes) (Satinsky *et al.*, 2014b). The NCBI sequences are fastq files with internal standard sequences and rRNA sequences (metatranscriptomes only) removed. Sequences are also available at the iMicrobe database under project number CAM\_P\_0001194 (<http://data.imicrobe.us>). The iMicrobe sequences are QC'd fasta files of joined paired-end reads, also with internal standards and rRNA sequences (metatranscriptomes only) removed. Metadata accompanying the omics data sets are available at the BCO-DMO data repository (<http://www.bco-dmo.org/project/2097>).

Measurements of environmental parameters were made as previously described (Goes *et al.*, 2014; Medeiros *et al.*, 2015).



**Figure 1** (a) Amazon Plume stations sampled May–June 2010 superimposed on salinity contours interpolated from the shipboard underway system. (b) Environmental data. (c) Transcript inventories (mRNAs  $\text{L}^{-1}$ ) for the 10 bacterial or archaeal taxonomic bins contributing the most transcripts at each station.

## Results and discussion

Genes and transcripts were collected in May–June 2010 from free-living (operationally defined as 0.2–2.0  $\mu\text{m}$ ) and particle-associated (2.0–156  $\mu\text{m}$ ) microbial communities in surface water at six stations in the Amazon River Plume (Table 1). The stations were grouped into three salinity-based classes following a scheme suggested previously (Subramaniam *et al.*, 2008): low-salinity Stations 10 and 23 (salinity <30), mesohaline Stations 3, 2 and 25 (salinity 30–35), and oceanic Station 27 (salinity >35) (Figure 1, Table 1).

### Taxonomic composition

Patterns of bacterial and archaeal genes and transcripts were determined from duplicate metatranscriptomic and metagenomic data sets for each size fraction and station (24 of each type). A previous study analyzing only data from Station 10 showed

that free-living cells made a greater contribution to the community transcriptome because of their higher cell numbers (resulting in more transcripts per liter), but particle-associated cells had higher rates of gene expression (more transcripts per gene copy) (Satinsky *et al.*, 2014a). In this study, data from five additional plume locations in addition to Station 10 allowed a comparison of microbial expression patterns from the outer continental shelf to oceanic waters.

The most transcriptionally dominant genome bins of microbial heterotrophs at the low salinity Stations 10 and 23 included three marine Gammaproteobacteria (HIMB30, HIMB55 and SAR86), along with SAR116 (*Candidatus puniceispirillum* IMCC1322), SAR324 and several Flavobacteria (Figure 1). At mesohaline Stations 3, 2 and 25, the Verrucomicrobium *Coralimargarita akajimensis* DSM45221 reference genome bin contributed the most transcripts, while at oceanic Station 27 the most transcriptionally dominant heterotrophic gen-

**Table 2** Transcript gene copy<sup>-1</sup> ratios in bacterial and archaeal cells

Community or Organism	Transcripts gene copy <sup>-1</sup>	Method	Reference
Amazon Plume FL: all genes, all stations <sup>a</sup>	0.040	Internal standard	This paper
Amazon Plume PA: all genes, all stations <sup>a</sup>	1.331	Internal standard	This paper
Amazon Plume PA: all genes, low salinity <sup>a</sup>	0.058	Internal standard	This paper
Amazon Plume PA: all genes, mesohaline <sup>a</sup>	2.584	Internal standard	This paper
Amazon Plume PA: all genes, oligotrophic <sup>a</sup>	0.117	Internal standard	This paper
Amazon Plume FL: 40 genes, all stations <sup>a</sup>	0.086	Internal standard	This paper
Amazon Plume PA: 40 genes, all stations <sup>a</sup>	0.884	Internal standard	This paper
Amazon Plume PA: <i>nifH</i> , Station 27	32.3 ± 5.7 <sup>b</sup>	Internal standard	This paper
Amazon Plume PA: <i>pstA</i> , Station 2	5.12 ± 0.42 <sup>b</sup>	Internal standard	This paper
<i>Coralimargarita akajimensis</i> DSM4522 <sup>c</sup>	0.167	Internal standard	This paper
Gammaproteobacteria HIMB30 <sup>c</sup>	0.051	Internal standard	This paper
SAR324 JCVI-SC AAA005 <sup>c</sup>	0.045	Internal Standard	This paper
Alteromonadales <sup>d</sup>	0.073	Internal standard	This paper
Roseobacter clade <sup>d</sup>	0.027	Internal standard	This paper
SAR11 clade <sup>d</sup>	0.002	Internal standard	This paper
SAR86 clade <sup>d</sup>	0.006	Internal standard	This paper
<i>Escherichia coli</i>	0.309	Biochemical analysis	<sup>e</sup>
<i>Escherichia coli</i>	0.403	Single-cell microscopy	<sup>f</sup>
<i>Vibrio alginolyticus</i>	0.485	RNA recovery	<sup>g</sup>
<i>Mycoplasma pneumonia</i>	0.04	Microarray with internal standard	<sup>h</sup>
OM43 strain NB0046	0.3–1.4	Internal standard	<sup>i</sup>

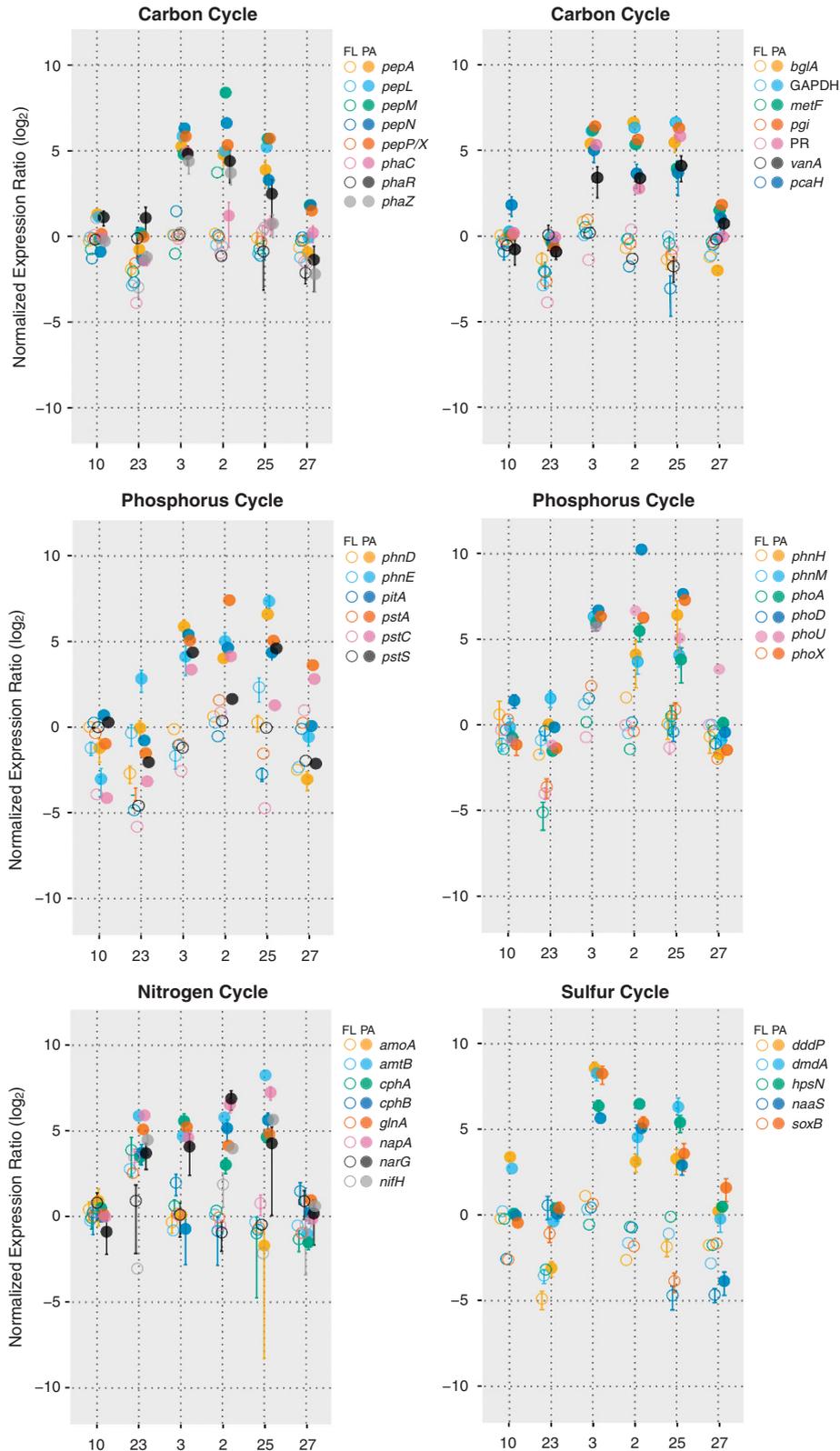
<sup>a</sup>Bacteria and archaea. <sup>b</sup>S.d. in the duplicate metatranscriptomes can be calculated for individual genes. <sup>c</sup>Homogeneous populations at Station 23 and 3. <sup>d</sup>All stations and size fractions. <sup>e</sup>Neidhardt and Umbarger (1996); calculated by dividing mRNAs per cell by the 4466 genes in the *E. coli* K12 genome. <sup>f</sup>Taniguchi *et al.* (2010); calculated by extrapolation from 137 genes as in Moran *et al.* (2013). <sup>g</sup>Taniguchi *et al.* (2010); based on a calculation of 2288 mRNA molecules per cell as in Moran *et al.* (2013) and 4722 coding sequences in the genome. <sup>h</sup>Maier *et al.* (2011). <sup>i</sup>Gifford *et al.* (2016); calculated by addition of 14 internal mRNA standards.

**Figure 2** Expression levels of 40 biogeochemically relevant genes plotted by station. Expression ratios were normalized within each gene to the median across all stations and both size fractions (transcripts gene copy<sup>-1</sup>/median transcripts gene copy<sup>-1</sup>) and plotted as log<sub>2</sub> values. Carbon and phosphorus genes are split into two panels for clarity. Gene abbreviations: Carbon genes—*pepALMN*, *P/X*, various aminopeptidases; *phaCRZ*, polyhydroxyalkanoate synthesis, degradation and regulation; *bglA*, beta-glucosidase; *metF*, methylene-tetrahydrofolate reductase; *pgi*, glucose-6-phosphate isomerase; PR, proteorhodopsin; *pcaH*, protocatechuate 3,4-dioxygenase; *vanA*, vanillate demethylase. Phosphorus genes—*phnDEHM*, phosphonate transport and assimilation; *pitA*, low affinity phosphate transporter; *pstACS*, high-affinity phosphate transporter components; *phoADUX*, various alkaline phosphatases. Nitrogen genes—*amoA*, ammonium monooxygenase; *amtB*, ammonium transporter; *cphAB*, cyanophycin synthesis and degradation; *glnA*, glutamine synthetase; *napA*, nitrate reductase; *narG*, nitrate reductase; *nifH*, nitrogenase. Sulfur genes—*dddP*, DMSP lyase; *dmdA*, DMSP demethylase; *hpsN*, DHPS dehydrogenase; *naaS*, *N*-acetyltaurine amidohydrolase; *soxB*, sulfate thiohydrolase. Detailed description of gene functions are provided in Supplementary Table S6.

ome bins were *C. akajimensis*, SAR116 clade, Roseobacter clade members and *Alteromonas*.

Autotrophic bacteria from the cyanobacterium genus *Synechococcus* were important in the transcriptomes of most stations (Figure 1) except for

oceanic Station 27 where *Prochlorococcus* and *Trichodesmium* were more abundant. At Stations 2 and 25, the cyanobacterium *Richelia intracellularis*, known to form a symbiotic ('DDA') association with diatom *Hemiaulus hauckii*, was the dominant cya-



nobacterial transcript bin. These data agree with microscopic counts and pigment analysis that placed Stations 10 and 23 into an 'estuarine type' with mixed population of diatoms, cryptophytes and *Synechococcus*; Stations 2, 3 and 25 into a 'mesohaline type' dominated by DDAs and Station 27 into an 'oceanic type' characterized by *Trichodesmium* (Goes et al., 2014).

#### Transcripts per gene copy

There were fewer transcripts in each liter of plume seawater than there were genes by ~30-fold (Table 2). Exceptions (when transcript gene<sup>-1</sup> ratios were >1) were found only at the mesohaline stations for certain genes (*nifH*, *amtB*, *pstA,C,S* and *pepM*). This low transcript-to-gene ratio is similar to previous findings in marine microbial communities (Beier et al., 2015; Cottrell and Kirchman, 2016). Even for model bacteria growing exponentially, cells contain fewer transcripts than genes (~0.4 transcripts gene copy<sup>-1</sup> for exponential phase *Escherichia coli* and ~0.5 transcripts gene copy<sup>-1</sup> for *Vibrio alginolyticus*; Kramer and Singleton, 1992; Taniguchi et al., 2010) (Table 2). Typical mRNA half-lives are on the order of minutes in prokaryotic cells (Maaloe et al., 1983; Steglich et al., 2010; Taniguchi et al., 2010) and the Amazon Plume data are consistent with a low inventory but high turnover rate of mRNA (Moran et al., 2013).

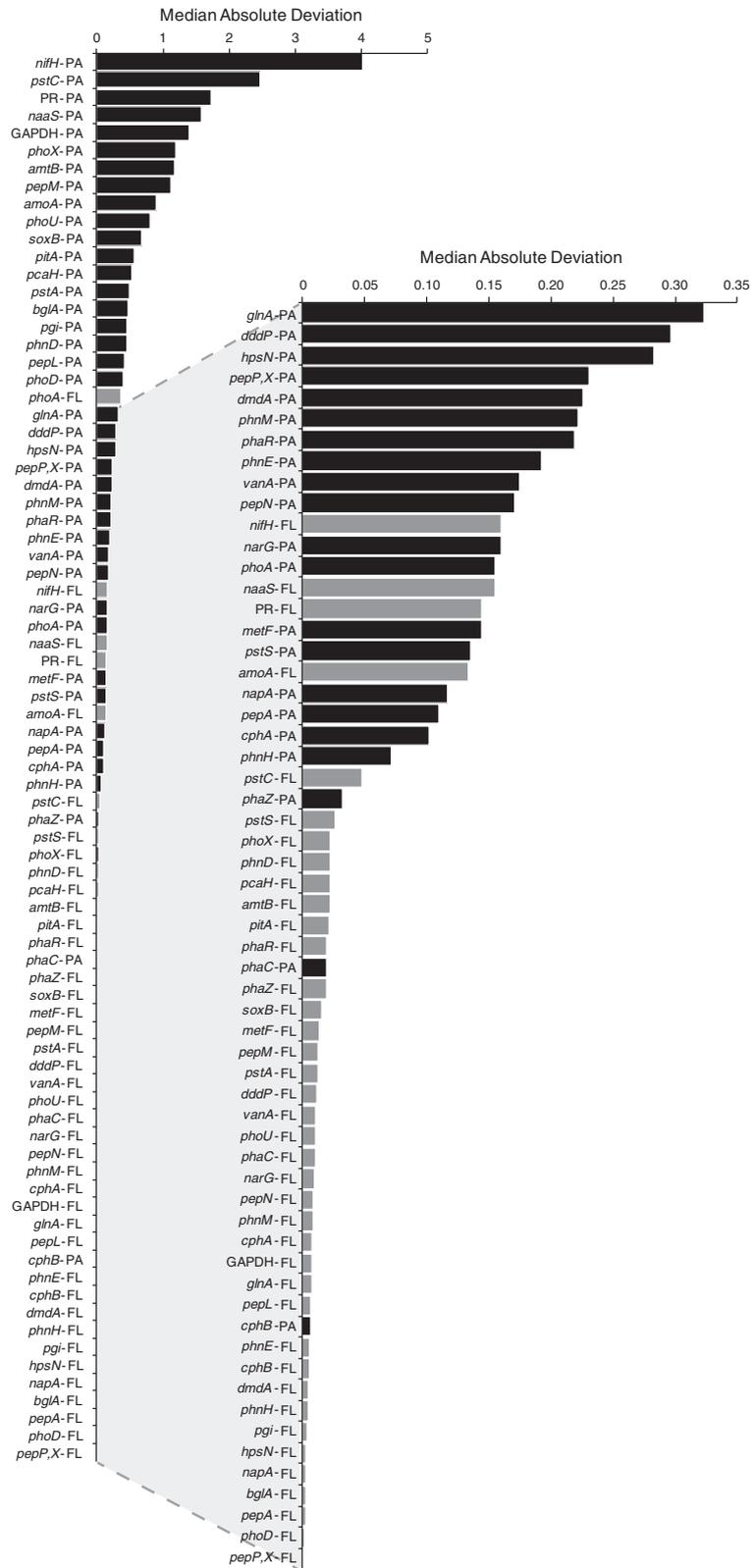
Different levels of baseline gene expression have been predicted for marine bacterial taxa because of variations in genetic capabilities for synthesizing transcription machinery and regulating transcript production (Giovannoni et al., 2005; Polz et al., 2006; Lauro et al., 2009; McCarren et al., 2010; Luo and Moran, 2015). We compared transcript gene copy<sup>-1</sup> ratios for four taxonomic groups in the plume that are hypothesized to have either higher (Alteromonadales and Roseobacter clades) or lower (SAR11 and SAR86 clades) baseline expression based on genome characteristics (Giovannoni et al., 2005; Polz et al., 2006; Lauro et al., 2009; McCarren et al., 2010; Luo and Moran, 2015). The Alteromonadales and Roseobacter groups averaged up to 10-fold higher expression ratios (0.07 and 0.03 transcripts gene copy<sup>-1</sup>, respectively) than did the SAR11 and SAR86 clades (0.002 and 0.006 transcripts gene copy<sup>-1</sup>) (Table 2). A recent study similarly found that Roseobacter and Flavobacteria strains, characterized by large genomes and abundant transcriptional regulators, produced more mRNA per cell than SAR11 and SAR92 strains, typified by small and streamlined genomes (Lankiewicz et al., 2016). We also compared transcript gene copy<sup>-1</sup> ratios for cells in the particle-associated and free-living environments and found that the former had higher transcript gene copy<sup>-1</sup> ratios by a factor of 10 when averaged across all elemental cycling genes (0.88 versus 0.09 transcripts gene copy<sup>-1</sup>) (Table 2).

#### Patterns in element cycle gene expression

Expression of 40 genes involved in C, N, P and S cycling was compared by station and size fraction. The highest median expression ratios for C cycle genes were for *phaP* and *phaR* (0.4 and 0.10 transcripts gene copy<sup>-1</sup>), which mediate C storage as polyhydroxyalkanoate, and *pcaH* (0.10 transcripts gene copy<sup>-1</sup>), which encodes aromatic ring cleavage (Figure 2 and Supplementary Table S3). Among N cycle genes, ammonia oxidation gene *amoA* and nitrogen fixation gene *nifH* had the highest median expression (0.7 and 0.5 transcripts gene copy<sup>-1</sup>). Among P cycle genes, *pstC* and *pstS* of the high-affinity phosphate transporter had the highest median expression (0.7 and 0.5 transcripts gene copy<sup>-1</sup>). Finally, for S cycle genes, *N*-acetyltaurine amidohydrolase (*naaS*) and sulfur-oxidizing protein *soxB* had the highest median expression (0.7 and 0.5 transcripts gene copy<sup>-1</sup>). The most highly expressed genes were not necessarily common in plume microbial communities. For example, *amoA*, *nifH* and *phaP* were present in <2% of cells (based on genome equivalent calculations; Supplementary Table S2). These cells are contributing disproportionately to discrete steps in N and P cycles.

Two major trends were evident from the expression ratio data. The first was that the majority of genes exhibited peak expression ratios in the mesohaline region and lowest expression ratios in oceanic water (Figure 2 and Supplementary Table S3). For example, the average gene expression ratio across all 40 genes was 1.9 transcripts gene copy<sup>-1</sup> for mesohaline Stations 3, 2 and 25 compared with 0.13 transcripts gene copy<sup>-1</sup> for oceanic Station 27.

The second trend was that genes in the particle-associated community showed a larger dynamic range of expression compared with free-living. An analysis of the deviation of expression ratios from the median showed that the most extreme variation was found for *nifH* in particle-associated cells (Figure 3), with expression ratios at Stations 2 and 25 among the highest of any gene surveyed (9 and 32 transcripts gene copy<sup>-1</sup>) while no expression was found at Station 10 even though genes were present. Ammonium transporter gene *amtB* in particle-associated cells also had high deviation in expression (Figure 3) despite the fact that all stations had ammonium concentrations at or below the level of detection (Goes et al., 2014) (Figure 1b and Supplementary Table S4). Spatially variable expression ratios offer insights into the biogeochemical processes with the strongest changes along the axis of the plume. Among the 10 genes with the highest deviation around the median value, three carry out transformations in the N cycle (*nifH*, *amtB* and *amoA*), three in the P cycle (*pstC* and alkaline phosphatase genes *phoX* and *phoU*), two in the C cycle (glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in the glycolysis pathway and peptidase *pepM*) and one in the sulfur cycle (*N*-acetyltaurine



**Figure 3** Variability in the expression ratios of 40 biogeochemically relevant genes, shown as rank order of the median absolute difference by gene and size fraction. PA size fraction, gray bars; FL size fraction, black bars.

amidohydrolase gene *naaS*). A gene allowing energy acquisition from sunlight (proteorhodopsin) was also in the top 10 list. Many of these same genes also had high deviation in free-living cells (*nifH*, *naaS*, proteorhodopsin, *amoA*, *pstC*, *phoX* and *amtB*) but the range of variation was considerably less (Figure 3).

#### *Do expression changes reflect shifts in taxonomy or gene regulation?*

To investigate the relative importance of taxonomic composition versus gene regulation in explaining differences in the expression ratios in the plume, we first identified reference genome bins that were genetically coherent across multiple stations (Supplementary Figure S1). Analyzing taxonomically homogeneous bins eliminated complications from strain or population differences and allowed analysis of gene regulation in a single taxon at different plume locations. This screen for homogeneous genome bins revealed three reference bins with similar population structure across five or more samples: *C. akajimensis* DSM45221, marine Gammaproteobacteria HIMB30, and SAR324 cluster bacterium JCVI-SC AAA005 (Supplementary Figure S1).

Within each homogeneous genome bin, the large variations in the expression ratios across stations suggested substantial regulation of gene expression in response to local conditions (Figure 4). For example, GAPDH gene expression by the genetically homogeneous *C. akajimensis* population varied several orders of magnitude, from 0.004 transcripts gene<sup>-1</sup> for free-living cells at low-salinity Station 23 to 129 transcripts gene<sup>-1</sup> for particle-associated cells at mesohaline Station 3. For genes in the HIMB30 and SAR324 bins, expression levels varied up to 130- and 93-fold among genetically homogeneous populations (Figure 4). The transcription patterns in these reference bins echo the two major patterns observed for the full community: higher expression ratios at mesohaline stations and significantly more dynamic range in the expression for particle-associated cells (Figure 4). SAR11 member *Pelagibacter ubique* HTCC1062 has been shown to change gene expression levels by only threefold between nitrogen-replete and -limited conditions (Smith *et al.*, 2013), yet marine betaproteobacterium NB0046 varied the expression ratios over several orders of magnitude depending on growth stage and nutrient regime (Gifford *et al.*, 2016). Here the expression difference due to transcriptional regulation was >100-fold for some genes within a genetically homogeneous reference genome, considerably larger than the differences in baseline gene expression rates observed among bacterial taxa (3-fold; Table 2). These results invoke a major role for gene regulation in the shifting inventories of elemental cycling transcripts across the plume salinity and nutrient regimes.

#### *How do the stations differ?*

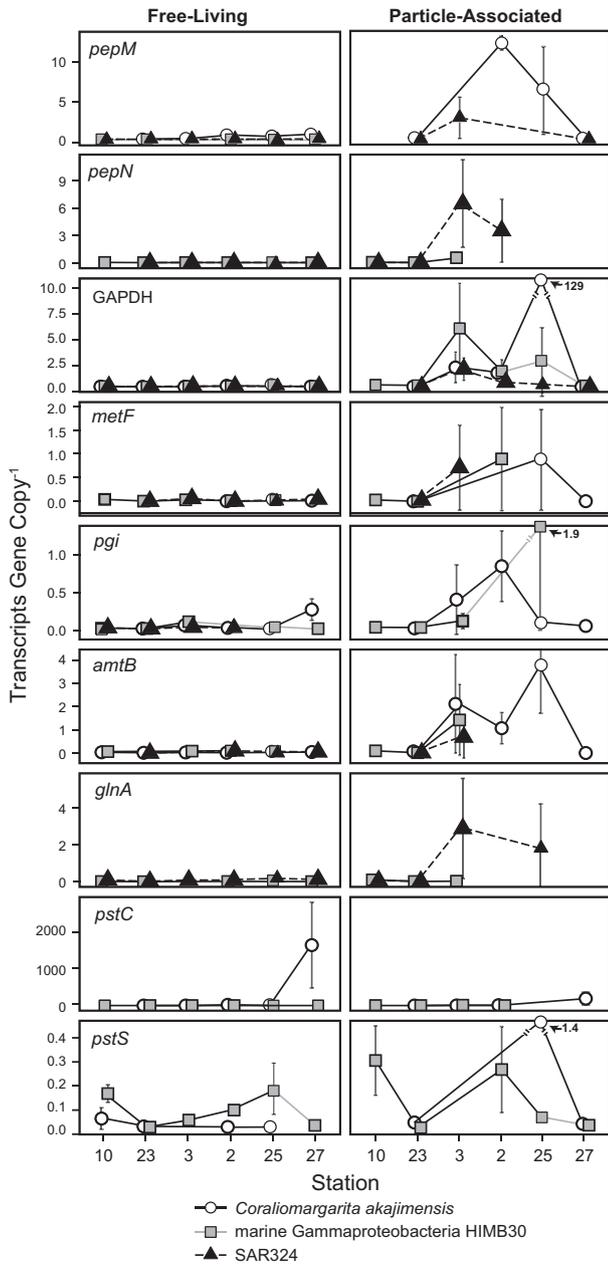
A previous study of Station 10 showed that genomes from particle-associated cells contributed more transcripts for sulfur cycling and aromatic compound degradation than free-living cells, with an overall twofold higher expression ratio, but fewer transcripts for phosphorus acquisition, carbon metabolism and ammonia oxidation (Satinsky *et al.*, 2014a). The five additional stations from more ocean-influenced regions of the Amazon Plume were examined to determine whether transcript inventories (transcripts l<sup>-1</sup>) and gene expression ratios (transcripts gene copy<sup>-1</sup>) (Supplementary Table S3) had consistent gene-specific patterns.

Similar to Station 10, the expression ratios of sulfur cycle genes were higher in cells associated with particulate material at the other stations (Figure 5). These genes included four mediating the utilization of phytoplankton-derived organic sulfur compounds (*dmdA* and *dddP* for dimethylsulfoniopropionate (DMSP), *hpsN* for dihydroxypropanesulfonate (DHPS) and *naaS* for *N*-acetyltaurine) and one catalyzing the oxidation of reduced inorganic sulfur to sulfate (*soxB*). Station 3 was the only location besides Station 10 to have *amoA* transcripts present and similarly had a higher inventory of transcripts for archaeal ammonia oxidation gene *amoA* in the free-living community. Unlike Station 10, transcript inventories and the expression ratios for aromatic compound degradation (vanillate demethylase *vanA* and protocatechuate 3,4-dioxygenase *pcaA*) were not particularly biased toward the particle-associated community at the more oceanic stations (Figure 5).

Genes plotting in the bottom two quadrants of Figure 5 had higher expression ratios in particle-associated cells. Nearly all genes at Stations 2, 3 and 25 were in these bottom quadrants, and 70–92% of genes at the low salinity and oceanic stations were also. Genes plotting in the two right quadrants in Figure 5 have higher transcript inventories in free-living cells. More than 80% of genes plotted in these right quadrants for all stations except mesohaline Stations 2 and 3 (18% and 68%). Thus, despite the generally higher expression ratios in the particle-associated community, transcript inventories were nonetheless higher for many genes in the free-living community owing to a greater abundance of free-living cells at all stations (>68%; Table 1, Supplementary Table S2), with Stations 2 and 3 as the exceptions to this pattern (Figure 5).

#### *Relationships between biogeochemical properties and gene expression patterns*

MINE (Reshef *et al.*, 2011) was used to detect associations between environmental properties of the plume and gene expression ratios. The analysis measured associations between data pairs (an environmental parameter paired with a gene expression ratio), capturing both linear and non-linear relationships and assigning a normalized maximum



**Figure 4** Differential gene regulation by genetically homogeneous bacterial populations at various locations in the plume. Expression ratios are shown for nine biogeochemically relevant genes in three homogeneous taxonomic bins (*Coraliomargarita akajimensis* DSM 45221, marine Gammaproteobacterium HIMB30 and SAR324 cluster bacterium JCVI-SC AAA005).

information coefficient value between 0 (no relationship) and 1 (strongest relationship). Of the 1040 parameter–gene pairs examined for each microenvironment (26 parameters × 40 genes; Supplementary Table S5), 53 high-scoring pairs emerged for genes in particle-associated cells (defined as maximum information coefficient > 0.90) and 104 emerged for genes in free-living cells (Supplementary Figures S2a–e).

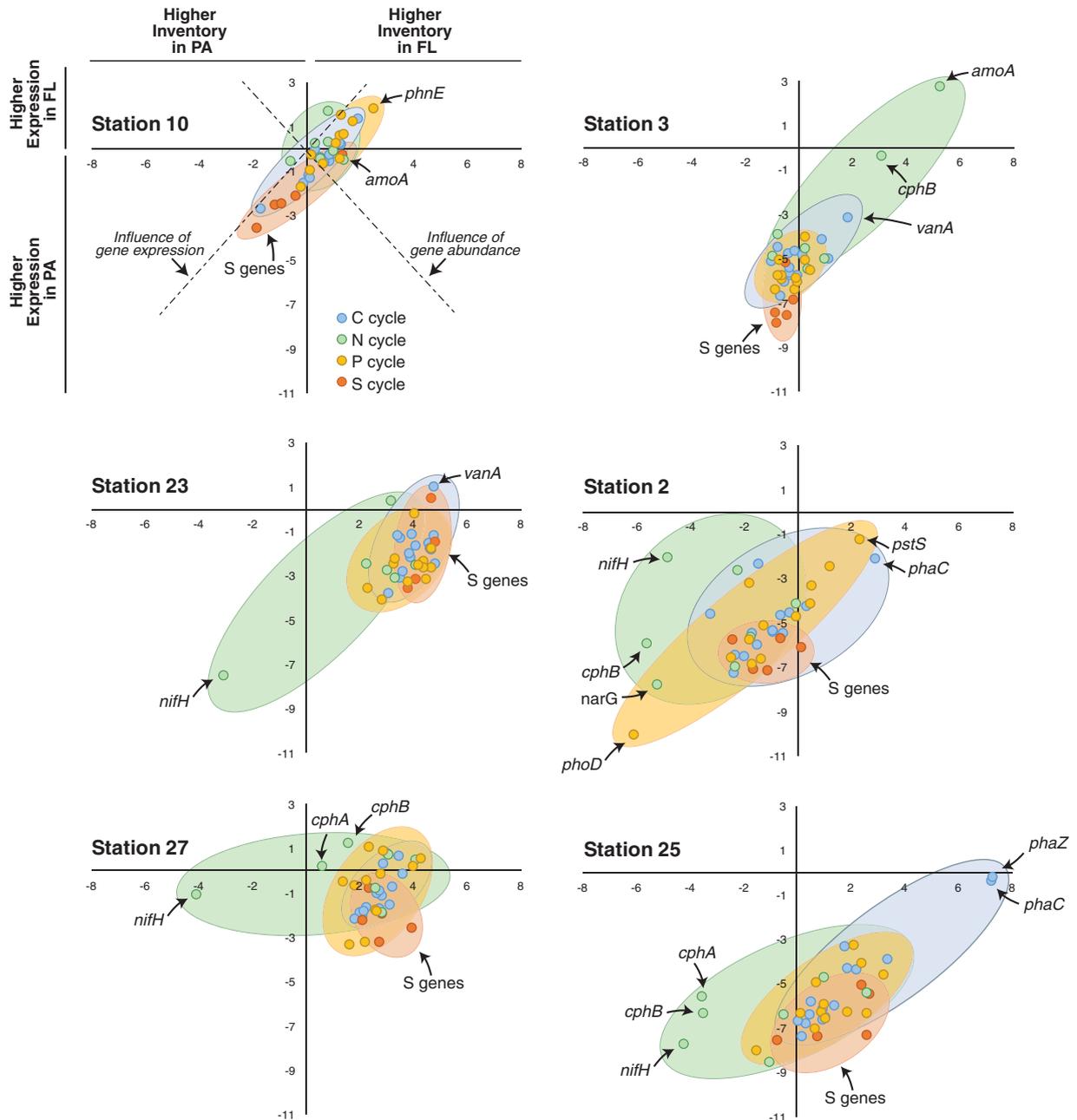
The most prominent pattern was found for the particle-associated community for which associations between phosphate concentration and gene

expression ratios accounted for 34 of the 53 high-scoring pairs. For 19 of these, the association was a negative linear correlation (Pearson’s coefficient,  $P < 0.05$ ) (Supplementary Figure S2a). These associations were driven by high gene expression ratios at mesohaline Stations 3, 2 and 25 where  $[PO_4]$  was  $< 0.2 \mu M$  and low gene expression at low salinity and ocean Stations 10, 23 and 27 where  $[PO_4]$  was  $> 0.3 \mu M$  (Figure 1b and Supplementary Table S5). In contrast, the free-living community had only two high-scoring pairs involving phosphate concentrations (Supplementary Figure S2e).

MINE analysis also identified a high-scoring positive correlation between  $N_2$  fixation rates and the expression of *nifH* and *phnN* (phosphonate utilization) genes ( $P < 0.01$ ) (Supplementary Figure S2b).  $N_2$  fixation typically peaks in mid-salinity regions of the Amazon Plume due to depletion of labile nitrogen while P is still available (Subramaniam et al., 2008; Loick-Wilde et al., 2015), and rate measurements at the time of this study also reflected this pattern;  $N_2$  fixation rates varied 50-fold from  $0.01 \text{ nmol l}^{-1} \text{ h}^{-1}$  at low-salinity Station 10 to  $0.5 \text{ nmol l}^{-1} \text{ h}^{-1}$  at mesohaline Station 25 (Loick-Wilde et al., 2015).

Unexpectedly, none of the particle-associated genes had high-scoring associations with particle composition parameters, including particulate organic carbon and particulate organic nitrogen concentrations. However, three markers of terrestrial influence in the dissolved phase had associations with both *nifH* and *cphB* particle-associated expression ratios: terrestrial contribution to the dissolved organic carbon pool (based on  $\delta^{13}C$  content), river-derived silicate (Si) concentration, and salinity. In all cases, gene expression ratios were highest at intermediate terrestrial influence (Stations 2 and 25) and low elsewhere (Supplementary Figure S2b). Both *nifH* and *cphB* have links to  $N_2$  fixation (*cphB* mediates intracellular nitrogen storage), and based on this pattern and the numerous relationships between  $[PO_4]$  and gene expression ratios described above, we surmise that stimulation of  $N_2$  fixation and an accompanying drawdown of phosphate broadly affected the expression of many biogeochemically relevant genes in regions of intermediate terrestrial influence.

Sulfur cycle gene *dddP* is responsible for the bacterial production of dimethylsulfide from DMSP and had high-scoring associations with turbidity and chlorophyll fluorescence in particle-associated cells (Supplementary Figure S2b), driven by a 60-fold higher expression ratio for *dddP* at Station 3 compared with the average of the other five stations. Highest expression at Station 3 was also observed for the four other sulfur cycle genes, although to a less dramatic extent (ranging from 3- to 25-fold higher for DMSP demethylation gene *dmdA*, sulfide oxidation gene *soxB* and sulfonate catabolism genes *hpsN* and *naaS*). Neither environmental parameters nor phytoplankton community composition (Goes et al., 2014; Carpenter, unpublished data) provided any insights into why flux of sulfur compounds was particularly



**Figure 5** Ratios of expression ( $\log_2(\text{free-living expression ratio}/\text{particle-associated expression ratio})$ ) and transcript inventory ( $\log_2(\text{free-living transcripts } l^{-1}/\text{particle-associated transcripts } l^{-1})$ ) for genes in the free-living versus particle-associated microenvironments. Expression ratios higher in free-living communities are plotted in the two upper quadrants while those higher in particle-associated communities are plotted in the lower quadrants; transcript inventories higher in free-living communities are plotted in the two right quadrants while those higher in particle-associated communities are plotted in the left quadrants. Genes are color-coded by element cycle; blue = C, green = N, light orange = P, dark orange = S. Guidelines for interpreting plots are given in the Station 10 panel. Data are provided in Supplementary Table S3.

high at Station 3. Sulfur cycle gene expression was also high in free-living cells at Station 3 (Figure 2 and Supplementary Figure S2d).

## Conclusion

In the Amazon Plume region of the western tropical North Atlantic, two hypotheses emerged from the

expression ratios of 40 bacterial and archaeal genes representing four major element cycles. First, many biogeochemical processes exhibit peak expression in the mesohaline regions of the plume, including carbon metabolism, nitrogen acquisition and phosphorus acquisition.  $N_2$  fixation by cyanobacterial symbionts of diatoms in regions of intermediate terrestrial influence appears to be a driver of the enhanced biogeochemical activity, leaving a

transcriptional imprint for increased nitrogen storage, utilization of more diverse sources of phosphorus and a peak in carbon catabolism. Indeed *Richelia* symbionts have been estimated to fix up to seven times more nitrogen than needed for their own growth (Foster *et al.*, 2011). Second, down-plume shifts in biogeochemical activities appear to be driven largely by particle-associated cells. We found a larger and more compositionally stable reservoir of transcripts in free-living cells across the six plume stations, overlain by a smaller yet highly dynamic transcript pool in particle-associated cells. Bacteria associated with the particulate matrix may experience more environmental variability along the axis of the plume compared with those living singly in bulk seawater. Although we did not find statistical associations between bulk particulate organic carbon or particulate organic nitrogen parameters and bacterial gene expression ratios on particulate material, prominent shifts in the phytoplankton community composition were occurring through the plume at the time of sampling, varying from coastal diatoms to mesohaline DDAs to oceanic *Trichodesmium* and *Prochlorococcus* (Goes *et al.*, 2014; Zielinski *et al.*, 2016). Moreover, particle-associated gene expression had strong associations with markers of terrestrial influence (Supplementary Figure S2b).

The export of terrestrially derived material from the Amazon River leads to enhanced rates of autotrophic and heterotrophic production over a large region of the tropical Atlantic (Yeung *et al.*, 2012; Medeiros *et al.*, 2015) with effects on both atmospheric CO<sub>2</sub> uptake and sequestration (Cooley *et al.*, 2007). Measures of the number of microbial transcripts produced per gene copy indicate an important role for gene regulation by heterotrophic bacteria, concurring with evidence from other estuarine and plume systems of higher cell-specific activity of particle-associated cells (Crump and Baross, 2000; Kellogg *et al.*, 2011). Gene expression patterns in the Amazon Plume also reiterate a critical influence of N<sub>2</sub> fixation over diverse microbial activities.

## Conflict of Interest

The authors declare no conflict of interest.

## Acknowledgements

We appreciate advice and assistance from Jan Mrazek, Adrian Burd, Roger Nilsen and Karie Sines and the captain and crew of the RV *Knorr* KN-197-08. Resources were provided by the University of Georgia's Georgia Advanced Computing Resource Center. This work was funded by the Gordon and Betty Moore Foundation through grants GBMF2293, GBMF2928 and GBMF538.01.

## References

Amon R, Benner R. (1998). Seasonal patterns of bacterial abundance and production in the Mississippi River

- plume and their importance for the fate of enhanced primary production. *Microb Ecol* **35**: 289–300.
- Beier S, Rivers AR, Moran MA, Obernosterer I. (2015). The transcriptional response of prokaryotes to phytoplankton-derived dissolved organic matter in seawater. *Environ Microbiol* **17**: 3466–3480.
- Church MJ, Wai B, Karl DM, DeLong EF. (2010). Abundances of crenarchaeal *amoA* genes and transcripts in the Pacific Ocean. *Environ Microbiol* **12**: 679–688.
- Coles VJ, Brooks MT, Hopkins J, Stukel MR, Yager PL, Hood RR. (2013). The pathways and properties of the Amazon River Plume in the tropical North Atlantic Ocean. *J Geophys Res-Oceans* **118**: 6894–6913.
- Cooley SR, Coles VJ, Subramaniam A, Yager PL. (2007). Seasonal variations in the Amazon Plume-related atmospheric carbon sink. *Glob Biogeochem Cycles* **21**: GB3014.
- Cottrell MT, Kirchman DL. (2016). Transcriptional control in marine copiotrophic and oligotrophic bacteria with streamlined genomes. *Appl Environ Microb* **82**: 6010–6018.
- Crump BC, Baross JA. (2000). Characterization of the bacterially-active particle fraction in the Columbia River estuary. *Mar Ecol Prog Ser* **206**: 13–22.
- Dagg M, Benner R, Lohrenz S, Lawrence D. (2004). Transformation of dissolved and particulate materials on continental shelves influenced by large rivers: plume processes. *Continental Shelf Res* **24**: 833–858.
- Dagg MJ, Breed GA. (2003). Biological effects of Mississippi River nitrogen on the northern Gulf of Mexico—a review and synthesis. *J Mar Syst* **43**: 133–152.
- Foster RA, Kuypers MM, Vagner T, Paerl RW, Musat N, Zehr JP. (2011). Nitrogen fixation and transfer in open ocean diatom-cyanobacterial symbioses. *ISME J* **5**: 1484–1493.
- Gardner WS, Cotner JB, Eadie BJ, Cavaletto JF, Benner R, Chin-Leo G. (1994). Mineralization of organic material and bacterial dynamics in Mississippi River plume water. *Estuaries* **17**: 816–828.
- Gifford SM, Sharma S, Rinta-Kanto JM, Moran MA. (2011). Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *ISME J* **5**: 461–472.
- Gifford SM, Becker JW, Sosa OA, Repeta DJ, DeLong EF. (2016). Quantitative transcriptomics reveals the growth- and nutrient-dependent response of a streamlined marine methylotroph to methanol and naturally occurring dissolved organic matter. *mBio* **7**: e01279–e01216.
- Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D *et al.* (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242–1245.
- Goebel NL, Turk KA, Achilles KM, Paerl R, Hewson I, Morrison AE *et al.* (2010). Abundance and distribution of major groups of diazotrophic cyanobacteria and their potential contribution to N<sub>2</sub> fixation in the tropical Atlantic Ocean. *Environ Microbiol* **12**: 3272–3289.
- Goes JI, Gomes HdR, Chekalyuk AM, Carpenter EJ, Montoya JP, Coles VJ *et al.* (2014). Influence of the Amazon River discharge on the biogeography of phytoplankton communities in the western tropical north Atlantic. *Prog Oceanogr* **120**: 29–40.
- Hilton JA, Satinsky BM, Doherty M, Zielinski BL, Zehr JP. (2014). Metatranscriptomics of N<sub>2</sub>-fixing cyanobacteria in the Amazon River plume. *ISME J* **9**: 1557–1569.
- Howard EC, Henriksen JR, Buchan A, Reisch CR, Bürgmann H, Welsh R *et al.* (2006). Bacterial taxa that limit sulfur flux from the ocean. *Science* **314**: 649–652.

- 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.
- Kellogg CT, Carpenter SD, Renfro AA, Sallon A, Michel C, Cochran JK *et al.* (2011). Evidence for microbial attenuation of particle flux in the Amundsen Gulf and Beaufort Sea: elevated hydrolytic enzyme activity on sinking aggregates. *Polar Biol* **34**: 2007–2023.
- Kramer JG, Singleton FL. (1992). Variations in rRNA content of marine *Vibrio* spp during starvation-survival and recovery. *Appl Environ Microb* **58**: 201–207.
- Lankiewicz TS, Cottrell MT, Kirchman DL. (2016). Growth rates and rRNA content of four marine bacteria in pure cultures and in the Delaware estuary. *ISME J* **10**: 823–832.
- Lauro FM, McDougald D, Thomas T, Williams TJ, Egan S, Rice S *et al.* (2009). The genomic basis of trophic strategy in marine bacteria. *P Natl Acad Sci USA* **106**: 15527–15533.
- Loick-Wilde N, Weber SC, Conroy BJ, Capone DG, Coles VJ, Medeiros PM *et al.* (2015). Nitrogen sources and net growth efficiency of zooplankton in three Amazon River plume food webs. *Limnol Oceanogr* **61**: 460–481.
- Luo H, Moran MA. (2015). How do divergent ecological strategies emerge among marine bacterioplankton lineages? *Trends Microbiol* **23**: 577–584.
- Maaloe O, Ingraham G, Maaloe O, Neidhardt F. (1983). *Growth of the Bacterial Cell*. Sinauer Association, Inc.: Sunderland, MA, USA.
- Maier T, Schmidt A, Güell M, Kühner S, Gavin AC, Aebersold R *et al.* (2011). Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol Sys Biol* **7**: 511.
- McCarren J, Becker JW, Repeta DJ, Shi Y, Young CR, Malmstrom RR *et al.* (2010). Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proc Natl Acad Sci* **107**: 16420–16427.
- Medeiros PM, Seidel M, Ward ND, Carpenter EJ, Gomes HR, Niggemann J *et al.* (2015). Fate of the Amazon River dissolved organic matter in the tropical Atlantic Ocean. *Glob Biogeochem Cycles* **29**: 677–690.
- Moran MA, Satinsky B, Gifford SM, Luo HW, Rivers A, Chan LK *et al.* (2013). Sizing up metatranscriptomics. *ISME J* **7**: 237–243.
- Moran MA, Kujawinski EB, Stubbins A, Fatland R, Aluwihare LI, Buchan A *et al.* (2016). Deciphering ocean carbon in a changing world. *Proc Natl Acad Sci USA* **113**: 3143–3151.
- Neidhardt F, Umbarger H. (1996). Chemical composition of *Escherichia coli*. In: Neidhardt FC (ed). *Escherichia coli and Salmonella: Cellular and Molecular Biology*. American Society for Microbiology: Washington, DC, pp 13–16.
- Pedrés-Alió C. (2006). Marine microbial diversity: can it be determined? *Trends Microbiol* **14**: 257–263.
- Polz MF, Hunt DE, Preheim SP, Weinreich DM. (2006). Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. *Philos Trans R Soc B* **361**: 2009–2021.
- Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ *et al.* (2011). Detecting novel associations in large data sets. *Science* **334**: 1518–1524.
- Richey JE, Nobre C, Deser C. (1989). Amazon River discharge and climate variability: 1903 to 1985. *Science* **246**: 101–103.
- Satinsky BM, Gifford SM, Crump BC, Moran MA. (2013). Use of internal standards for quantitative metatranscriptome and metagenome analysis. *Methods Enzymol* **531**: 237–250.
- Satinsky BM, Crump BC, Smith CB, Sharma S, Zielinski BL, Doherty M *et al.* (2014a). Microspatial gene expression patterns in the Amazon River Plume. *Proc Natl Acad Sci* **111**: 11085–11090.
- Satinsky BM, Zielinski BL, Doherty M, Smith CB, Sharma S, Paul JH *et al.* (2014b). The Amazon continuum dataset: quantitative metagenomic and metatranscriptomic inventories of the Amazon River plume, June 2010. *Microbiome* **2**: 17.
- Simon N, Cras AL, Foulon E, Lemée R. (2009). Diversity and evolution of marine phytoplankton. *Comptes Rendus Biol* **332**: 159–170.
- Smets W, Leff JW, Bradford MA, McCulley RL, Lebeer S, Fierer N. (2016). A method for simultaneous measurement of soil bacterial abundances and community composition via 16S rRNA gene sequencing. *Soil Biol Biochem* **96**: 145–151.
- Smith DP, Thrash JC, Nicora CD, Lipton MS, Burnum-Johnson KE, Carini P *et al.* (2013). Proteomic and transcriptomic analyses of *Candidatus Pelagibacter ubique* describe the first PII-independent response to nitrogen limitation in a free-living Alphaproteobacterium. *mBio* **4**: e00133–e00112.
- Smith WO, Demaster DJ. (1996). Phytoplankton biomass and productivity in the Amazon River plume: correlation with seasonal river discharge. *Continental Shelf Res* **16**: 291–319.
- Steglich C, Lindell D, Futschik M, Rector T, Steen R, Chisholm SW. (2010). Short RNA half-lives in the slow-growing marine cyanobacterium *Prochlorococcus*. *Genome Biol* **11**: 1.
- Subramaniam A, Yager PL, Carpenter EJ, Mahaffey C, Björkman K, Cooley S *et al.* (2008). Amazon River enhances diazotrophy and carbon sequestration in the tropical North Atlantic Ocean. *Proc Natl Acad Sci* **105**: 10460–10465.
- Taniguchi Y, Choi PJ, Li GW, Chen HY, Babu M, Hearn J *et al.* (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**: 533–538.
- Varaljay VA, Robidart J, Preston CM, Gifford SM, Durham BP, Burns AS *et al.* (2015). Single-taxon field measurements of bacterial gene regulation controlling DMSP fate. *ISME J* **9**: 1677–1686.
- Yeung LY, Berelson WM, Young ED, Prokopenko MG, Rollins N, Coles VJ *et al.* (2012). Impact of diatom-diazotroph associations on carbon export in the Amazon River plume. *Geophys Res Lett* **39**: L18609.
- Zhao Y, Tang H, Ye Y. (2012). RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics* **28**: 125–126.
- Zielinski BL, Allen AE, Carpenter EJ, Coles VJ, Crump BC, Doherty M *et al.* (2016). Patterns of transcript abundance of eukaryotic biogeochemically-relevant genes in the Amazon River Plume. *PLoS One* **11**: e0160929.

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