Title
Linking regional shifts in microbial genome adaptation with surface ocean biogeochemistry

Journal
Philosophical Transactions of the Royal Society B Biological Sciences, 375(1798)

ISSN
0962-8436

Authors
Garcia, Catherine A
Hagstrom, George I
Larkin, Alyse A
et al.

Publication Date
2020-05-11

DOI
10.1098/rstb.2019.0254

Peer reviewed

PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Garcia CA, Hagstrom GI, Larkin AA, Ustick LJ, Levin SA, Lomas MW, Martiny AC. 2020 Linking regional shifts in microbial genome adaptation with surface ocean biogeochemistry. *Phil. Trans. R. Soc. B* **375**: 20190254.
http://dx.doi.org/10.1098/rstb.2019.0254

Accepted: 14 February 2020

One contribution of 19 to a theme issue 'Conceptual challenges in microbial community ecology'.

Subject Areas:
environmental science, ecology, genomics, microbiology

Keywords:
metagenomics, elemental stoichiometry

Author for correspondence:
Adam C. Martiny
e-mail: amartiny@uci.edu

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.4873539.

THE ROYAL SOCIETY PUBLISHING

Linking regional shifts in microbial genome adaptation with surface ocean biogeochemistry

Catherine A. Garcia[1], George I. Hagstrom[3], Alyse A. Larkin[1], Lucas J. Ustick[2], Simon A. Levin[3], Michael W. Lomas[4] and Adam C. Martiny[1,2]

[1]Department of Earth System Science and [2]Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA
[3]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA
[4]Bigelow Laboratory for Ocean Sciences, East Boothbay, ME 04544, USA

CAG, 0000-0001-6192-4539; GIH, 0000-0002-2635-3723; AAL, 0000-0003-4466-0791; LJU, 0000-0002-6261-5241; SAL, 0000-0002-8216-5639; MWL, 0000-0003-1209-3753; ACM, 0000-0003-2829-4314

Linking 'omics measurements with biogeochemical cycles is a widespread challenge in microbial community ecology. Here, we propose applying genomic adaptation as 'biosensors' for microbial investments to overcome nutrient stress. We then integrate this genomic information with a trait-based model to predict regional shifts in the elemental composition of marine plankton communities. We evaluated this approach using metagenomic and particulate organic matter samples from the Atlantic, Indian and Pacific Oceans. We find that our genome-based trait model significantly improves our prediction of particulate C : P (carbon : phosphorus) across ocean regions. Furthermore, we detect previously unrecognized ocean areas of iron, nitrogen and phosphorus stress. In many ecosystems, it can be very challenging to quantify microbial stress. Thus, a carefully calibrated genomic approach could become a widespread tool for understanding microbial responses to environmental changes and the biogeochemical outcomes.

This article is part of the theme issue 'Conceptual challenges in microbial community ecology'.

1. Introduction

Linking genomics and other 'omics measurements with biogeochemical cycles is a widespread challenge in microbial community ecology. Currently, most 'omics observations are used to quantify shifts in diversity and functional potential. By contrast, we rarely use microbial 'omics data to understand and constrain large-scale energy or nutrient fluxes. This lack of convergence between microbial 'omics information and ecosystem or global models may limit our ability to predict future changes to global biogeochemical cycles (table 1).

It is well-established that the cellular and community regulation of elemental requirements and composition (i.e. carbon : nitrogen : phosphorus, C : N : P) are important for linking the global carbon and nutrient cycles [1]. There is an intense debate about the interaction between microbial diversity and environmental changes in regulating C : N : P for both terrestrial and aquatic environments [1,2]. The chemical composition of a cell is affected by many environmental factors, but nutrient availability is emerging as central [3]. Nutrient availability impacts the elemental composition of a community in multiple ways. Physiologically, the overall nutrient level impacts the growth rate [4]. In addition, cells are sensitive to the supply ratio of N versus P (and other nutrients) relative to the biomass ratio [5]. Microbial lineages can also have unique resource requirements and thus experience a shared environment differently at a physiological level. For example, the marine cyanobacterium *Prochlorococcus* appears to have a lower P requirement compared with larger phytoplankton [6] and co-existing diatoms can have unique N : P [7].

Table 1. Mean environmental characteristics for each ocean cruise transect. POC = particulate organic carbon, PON = particulate organic nitrogen, Pro. = *Prochlorococcus*, Syn. = *Synechococcus*, $V_{max[P_i]}$ = maximum PO_4 uptake rate, K_s = half saturation PO_4 concentration. BD = below detection and n.a. = not measured. $V_{max[P_i]}$ and K_s are calculated according to Michaelis–Menton functional kinetics for the whole community [25]; $V_{[P_i]} = (V_{max[P_i]}\ [P_i])/([P_i] + K_s)$.

cruise	POC (μM)	POP (nM)	C : P	PO_4 (nM)	NO_3 (μM)	Pro. abundance (cells ml^{-1})	Syn. abundance (cells l^{-1})	Pro. $V_{max[P_i]}$ (amol $cell^{-1}$ h^{-1})	Pro. K_s (nmol l^{-1})	Syn. $V_{max[P_i]}$ (amol $cell^{-1}$ h^{-1})	Syn. K_s (nM)	NH_4 uptake (nM N h^{-1})	urea uptake (nM N h^{-1})	NO_3 uptake (nM N h^{-1})
AE1319	3	23.2	167.5	36.5	6.1	34 690	12 164	8.9	10.6	39.5	29.9	n.a.	n.a.	n.a.
B46	n.a.	7.1	n.a.	13.6	n.a.	54 894	4307	10.4	5.9	46.8	11.4	n.a.	n.a.	n.a.
NH1418	1.6	10.8	153.1	74.1	2.3	82 423	2858	0.6	54.5	0.6	56.4	n.a.	n.a.	n.a.
I09	2	14.7	135.3	40.5	BD	145 089	3236	n.a.	n.a.	n.a.	n.a.	4.9	4.2	1.3

Thus, the interaction between microbial diversity and nutrient stress plays a complex role in regulating ecosystem C : N : P.

It is a challenge to define and quantify the nutritional environment experienced by microorganisms. First, the concentrations of inorganic phosphorus and nitrogen are commonly below detection limits in many marine environments [8]. Second, most microorganisms can use multiple alternative forms of nutrients [9–12]. Ammonium is energetically the most favoured form of nitrogen. When ammonium is in low supply, microorganisms can shift in some order to urea, nitrate, or organically bound nitrogen [13]. There are several unknowns associated with the use of alternative resources. We rarely quantify the concentration and chemical form of alternative nutrients or the chemical nature of organically bound N or P. Either assumptions are made about what substrate microorganisms are using, or there are difficulties measuring the uptake rate of diverse and poorly defined substrates. Furthermore, the resource costs associated with the use of many alternative nutrients are broadly unknown, leading to ill-defined trade-offs for nutrient assimilation. For example, cells need to invest N when upregulating acquisition proteins, leading to trade-offs between nutrient investments and uptake [14]. Finally, there is variation among individual lineages in the extent they can rely on alternative nutrient forms [15]. Thus, it is currently impossible to predict microbial nutrient use and associated biogeochemical roles even with a perfect chemical characterization of an environment.

Marine microorganisms show clear genomic evidence for adaptation to specific nutritional environments through gene gain and loss [16–18]. Such genomic changes reflect a shift from simple to more complex nutrient forms under limiting conditions. This pattern has been detected in many microorganisms but is clearly illustrated in marine cyanobacteria. In regions with a replete inorganic phosphate supply, streamlined *Prochlorococcus* genomes mainly contain transporters directly associated with inorganic phosphate uptake [19]. However, *Prochlorococcus* adapts to low phosphate supply via the gain of genes associated with regulation and the use of alternative forms. In regions with severe P stress, *Prochlorococcus* genomes contain genes for alkaline phosphatase to cleave off phosphate from organic molecules [20,21]. Here, alkaline phosphatase and a few other proteins are highly induced to use organic P as an alternative P source [19,22]. *Prochlorococcus* adapts to N stress in a parallel fashion, whereby cells from high N areas only contain genes for ammonium uptake [23]. In regions with stronger N stress, *Prochlorococcus* genomes sequentially include genes for urea, nitrite and ultimately nitrate assimilation. Thus, the genome content of *Prochlorococcus* (and other marine microorganisms) closely corresponds to the underlying environmental conditions and thereby describes the cellular strategies for nutrient acquisition [24].

We propose using genomic shifts among microbial communities as a 'biosensor' for *in situ* nutritional environments in order to improve predictions of resource use and C : N : P variability across ocean regions. Specifically, we combine the distribution of genes with a trait model to simulate cellular investment strategies and predict C : N : P. We assume that genome streamlining in cyanobacteria will lead to clear nutrient investment trends. However, increasing cell genomes sizes in the larger cyanobacteria *Synechococcus* reveals a more generalist lifestyle. We show that in comparison with both traditional abiotic and common trait models, the incorporation of nutrient trait variation quantified using metagenomics greatly improves our ability to predict shifts in

C : N : P. This work illustrates how we can use microbial community 'omics observations to improve our understanding of global biogeochemical cycles in ways that would be challenging to achieve with abiotic characterizations alone.

2. Material and methods

(a) Sample collection

Seawater samples were collected from the western Atlantic Ocean (AE1319, Aug/Sep 2013; BV46, Oct 2011), central Pacific Ocean (NH1418, Sep 2014) and the eastern Indian Ocean (IO9N, Mar/Apr 2016) (electronic supplementary material, figure S1 and table S1). On each cruise, samples for DNA, flow cytometry, particulate organic matter, uptake rate kinetics and nutrient measurements were collected as described previously [3,25–28] (table 1). Fifty-four stations were selected for metagenomics analysis where these corresponding measurements were taken. Selected data are already available on BCO-DMO (uptake rate kinetics, nutrient concentrations, cell abundances and particulate elemental concentrations) for the Atlantic AE1319 and BV46 (https://www.bco-dmo.org/project/2178) and Indian Ocean I09N cruises (https://www.bco-dmo.org/project/628972). Results have previously been reported describing the cyanobacterial diversity [28,29], cell quotas and abundances [26,27], uptake rate kinetics [25,26] and particulate organic matter ratios [3] along several transects.

(b) Particulate organic matter

All particulate organic matter samples for carbon (POC), nitrogen (PON) and phosphorus (POP) were collected on pre-combusted (4 h at 500°C) GF/F filters with a nominal pore size of 0.7 µm (electronic supplementary material, table S1). A nylon mesh prefilter with a pore size of 30 µm was used to remove rarer biomass such as larger plankton and particles. POP filters were rinsed with 0.17 M Na_2SO_4 at the time of collection to remove residual dissolved organic phosphorus. All filters were stored frozen until analysis in the laboratory. POC/PON samples were measured using a Flash 1112 EA elemental analyser (Thermo Scientific, Waltham, MA, USA) for the I09 transect against an atropine ($C_{17}H_{23}NO_3$) standard curve (range 0.2–1.5 mg). For the NH1418, AE1319 and BV46 transects, POC/PON samples were measured on either a control equipment 240-XA or 440-XA elemental analyser using acetanilide as a standard [30]. POP samples were analysed using an ash/hydrolysis colorimetric method described previously [31]. Briefly, 2 ml of 0.017 M $MgSO_4$ was added to the filter and KH_2PO_4 standards in acid-washed scintillation vials and dried overnight at 90°C. The filters were exposed to high temperature (500°C) for 2 h and acidified in 0.2 M HCl at 90°C. After a mixed reagent was added, the samples were analysed on a spectrophotometer at 885 nm.

(c) Uptake rate kinetics

On the Atlantic (AE1319, BV46) and Pacific (NH1418) Ocean transects, phosphate uptake rate kinetics were measured for whole community and taxon-specific groups (e.g. *Synechococcus* and *Prochlorococcus*) using methods previously described [25]. Incubations were performed using 10 ml seawater aliquots within 3°C of ambient temperature during the time of collection (approximately 23°C). Kinetics experiments for phosphate were performed with increasing dissolved inorganic phosphorus (DIP) additions up to 100 nM and ended at a final concentration of 100 µM.

On the Indian Ocean GO-SHIP transect (I09N), whole community bottle incubations were performed for uptake of ^{15}N-labelled ammonia, urea and nitrate [26]. The incubations were performed in 2 l polycarbonate bottles over a 6 h period at ambient seawater temperature. N incubations were mixed to a final concentration of 0.03 µM, which is below the detection limit and reflective of the N-limiting conditions throughout the I09N transect.

(d) Cell abundances using flow cytometry

Samples for flow cytometry and cell sorting were collected previously and are presented elsewhere [26–28]. Briefly, the samples were sorted using a FACSJazz or Influx flow cytometer (BD, Franklin Lakes, NJ, USA). Samples were preserved using a 0.5% paraformaldehyde solution (final concentration), kept in the dark for 1 h to fix at 5°C and then stored frozen at −80°C until analysis. Populations of *Synechococcus* were determined with a gate in orange (585 nm), *Prochlorococcus* based on forward scatter and red fluorescence.

(e) Nutrients

For the NH1418, AE1319 and BV46 cruises, phosphate was measured using the MAGIC-SRP high-sensitivity method [32]. Nitrate was measured by using a cadmium reduction assay as previously described [28].

Nutrient data for the I09N cruise were provided by Jim Swift (Scripps Institution of Oceanography, SIO) and Susan Becker (SIO) and are available at https://cchdo.ucsd.edu.

(f) Metagenomics: library and sequencing

For DNA, 4–10 l seawater samples were collected with a 0.22 µm Sterivex filter and preserved with lysis buffer (50 mM Tris-HCl pH 7.6, 20 mM EDTA pH 8.0, 400 mM NaCl, 0.75 M sucrose) and frozen at −80°C until further processing. Whereas a GF/D (2.7 µm nominal pore size) glass fibre prefilter was used for all Pacific and Atlantic sites [28], no prefilter was used for DNA collections for Indian Ocean sites. As a minor percentage of the total community is composed of eukaryotes [26], we assumed this was an acceptable comparison. However, it is possible that we are missing particle associations greater than 2.7 µm in the Atlantic and Pacific Oceans. DNA was extracted as described previously [28,33,34] and diluted (Atlantic and Pacific: 0.5 ng µl^{-1}, Indian: 1 ng µl^{-1}) for sequencing. Metagenomic libraries were prepared using the Nextera Library Prep Kit (Illumina, San Diego, CA) with a modified PCR mixture. DNA with a concentration of between 0.5 and 1 ng µl^{-1} was tagmented using the Nextera DNA Prep Kit tagmentation enzyme and incubated for 10 min at 55°C. The Nextera XT barcodes were annealed to metagenome fragments using the following PCR protocol. For PCR, we used 20 µl of a master mix containing 0.5 µl Phusion High Fidelity buffer (New England Biolabs, Ipswich, MA), 0.5 µl dNTPs (New England Biolabs, Ipswich, MA), 0.25 µl Phusion High Fidelity Polymerase (New England Biolabs, Ipswich, MA) and 14.25 µl of PCR-grade water. Equimolar samples were pooled and the quality was checked and quantified using a Bioanalyzer (Agilent, Santa Clara, CA). The pooled library was sequenced on an HiSeq 4000 (Illumina, San Diego, CA), producing paired end reads (2 × 150 bp). Low-quality reads and adapters were removed using Trimmomatic 0.35 [35] with a sliding window of 4:15 and minimum length set to 36. PhiX was filtered out using BBduk2 tool BBMap (https://sourceforge.net/projects/bbmap/, $k = 31$, hdist = 1). Sequences were aligned and mapped to a curated reference database (electronic supplementary material, tables S2 and S3) using Bowtie 2 [36] with the following settings: -local -D 15 -R 2 -L 15 -N 1 -gbar 1 -mp 3. High-quality contigs were assembled and processed with Anvi'o [37]. Pangenome gene clusters were identified using the DIAMOND algorithm [38] and summarized in Anvi'o. Metagenomes are available through BioProject (SRA PRJNA598881) at the following link: https://www.ncbi.nlm.nih.gov/sra/PRJNA598881.

(g) Nutrient assimilation gene frequencies

Prochlorococcus and *Synechococcus* genes associated with assimilation for iron, nitrogen and phosphorus were identified based on prior studies (electronic supplementary material, S1) [17,21,23,24,39,40]. Several genes of unknown function are listed as *uknX*, but are included because of their association with low P availability in *Prochlorococcus* [41], and close proximity to known

regulatory P assimilation genes in the MED4 genome. Based on these past studies, we filtered out genes if present in all *Synechococcus* and *Prochlorococcus* to detect variation in lineage coverage. We found the relative gene frequency by scaling to the median coverage of single copy core genes (SCCGs) [41] across 54 stations. We identified the relative gene frequency for each nutrient listed in electronic supplementary material, S1, per station, and per taxon (*Synechococcus* and *Prochlorococcus*) as follows:

$$\text{relative gene frequency}_{\text{gene in taxon}} = \sum_{\substack{\text{genomes}\\ \text{in taxa}}} \left[\left(\frac{\text{gene coverage}_{\text{gene}}}{\text{median coverage of SCCG}_{\text{taxa}}}\right)\left(\frac{\text{total reads}_{\text{genome}}}{\text{total reads}_{\text{taxa}}}\right)\right].$$

Next, we conducted three separate principal component analyses (PCAs) for N, P and Fe assimilation genes, respectively (electronic supplementary material, figure S4). Each relative gene frequency was scaled between 0 and 1 across the 54 stations as inputs to each PCA (n x m matrix of n stations and m normalized gene frequencies). A total of four gene indices were produced for each station, where N_{gene} (or P_{gene}) = first component of PCA:

$N_{\text{gene } \textit{Prochlorococcus}}$,
$P_{\text{gene } \textit{Synechococcus}}$,
$N_{\text{gene } \textit{Prochlorococcus}}$
and $P_{\text{gene } \textit{Synechococcus}}$.

These N and P gene indices for *Prochlorococcus* and *Synechococcus* were subsequently incorporated into a trait model to predict C : P.

(h) ATOM-gene model

We developed the ATOM-gene model to predict phytoplankton C : P ratios from temperature, irradiance and metagenomic data on phosphorus and nitrogen nutrient uptake gene abundance. The ATOM-gene model shares its basic structure with the trait-based phytoplankton model developed by Moreno *et al.* [42]. It predicts the C : P of particulate organic matter in the surface ocean using a multi-step process. ATOM-gene first characterizes phytoplankton according to several key functional traits, namely their radius (r) and their allocation of biomass to biosynthetic proteins and ribosomes (E), to photosynthetic proteins (L), to structural components (S), and to nutrient uptake proteins (A). ATOM-gene also represents a luxury nutrient storage pool. Each trait combination corresponds both to a functional response to environmental conditions, and to cell quotas of C, N and P, which we derived from biophysics, physiology and statistical modelling. The functional response determines the growth rate of cells with each trait combination (r, E, L, A) in each possible environment, which consists of temperature (T), irradiance (L) and metagenomic uptake gene abundance indices P_{gene} and N_{gene}. Traditionally, in trait-based phytoplankton models, the functional response to environmental conditions requires nutrient concentrations to calculate growth rates. However, nitrate + nitrite and phosphate nutrient concentrations are frequently below standard assay detection limits. Furthermore, nutrient concentrations were not great predictors across regions. Therefore, we needed genes to detect unseen nutrient stress variability. Here, we treat nutrient concentrations as latent variables, which are not directly observed, and model their concentration using the metagenomic data.

Given the irradiance, temperature and nutrient uptake gene abundances in a given sampling location, ATOM-gene uses the functional responses to determine the trait combination with the fastest growth rate and predicts that these traits and the resulting C : P characterize the plankton community and particulate organic matter at that sampling site.

ATOM-gene is part of a family of trait-based models that we have developed to predict C : P ratios in phytoplankton, and which extend the model in Moreno *et al.* [42] in important ways. First, ATOM-gene does not just model phosphorus availability like [42], but also models nitrogen availability. ATOM-gene includes an additional resource investment pool, representing variable allocations of biomass to surface membrane and periplasmic proteins for nutrient uptake of phosphorus. Secondly, we parameterized the trait-based model in [42] using the point estimates of physiological parameters taken from the literature, only using statistical methods to predict luxury P storage. Here we integrated the entire ATOM-gene model into a Bayesian statistical framework, allowing us to incorporate uncertainty in our understanding of key physiological processes (such as the temperature dependence or different biochemical processes).

Below we describe the model and its parameters. Summaries of the model parameters, and the prior distributions for statistical parameters, can be found in electronic supplementary material, tables S5, 6–S7. Phytoplankton traits determine P : C according to:

$$(\text{P}:\text{C}) = \frac{EP_{\text{E}} + \gamma P_{\gamma} + P_{\text{stor}}}{EC_{\text{prot}} + LC_{\text{prot}} + \gamma C_{\gamma} + \alpha(C_{\text{M}} + AC_{\text{prot}})/2r} (\text{mol}_{\text{C}}/\text{mol}_{\text{P}}).$$

Here P : C is the phosphorus to carbon ratio. P_{E} and P_{γ} are the specific fraction of phosphorus in the biosynthetic protein and structure pool, respectively, with units of gP g^{-1}. Their phosphorus content arises from ribosomes in the case of the biosynthetic apparatus, which we model as having a ribosome fraction of α_{E} and from DNA/RNA in the case of the structural pool, which we model as occupying a total fraction γ_{DNA} of cellular biomass. P_{stor} is the level of luxury P storage, in units of gP g^{-1}. The symbol C_{prot} is the specific fraction of carbon in proteins, with units of gC g^{-1}, C_{DNA} is the specific fraction of carbon in DNA, C_{γ} is composed of C_{lip} and C_{DNA}, C_{lip} is the specific fraction of carbon in lipids and γ_{lip} is the fraction of cellular biomass in lipids. The fraction of cellular biomass in the inner and outer membranes and periplasmic space is α/r, which we assume is half membrane and half periplasmic space. A is the fraction of the periplasmic space occupied by proteins. C_{M} is the carbon fraction of the inner and outer membranes, which we assume are composed partially of proteins and partially of phospholipids. mol_{P} and mol_{C} are the molar masses of phosphorus and carbon.

The traits must satisfy several constraints. The sum of allocations to cytoplasmic components should equal the cytoplasmic fraction of the cell:

$$E + L + \gamma_{\text{DNA}} + \gamma_{\text{lip}} = 1 - \frac{\alpha}{r}.$$

Furthermore, the fraction of the periplasmic volume allocated to proteins satisfies $2rA_{\min} < A < 1$.

To predict the stoichiometry in a given environment, ATOM-gene selects the trait combination with the fastest growth rates. Environmental conditions and traits translate into rates of biosynthesis μ_{E}, photosynthesis μ_{L}, nitrogen uptake μ_{N} and phosphorus uptake μ_{P}, with overall growth rate determined by the slowest of these processes:

$$\mu = \min(\mu_{\text{E}}, \mu_{\text{L}}, \mu_{\text{N}}, \mu_{\text{P}}).$$

The biosynthesis rate depends linearly on the investment E:

$$\mu_E = k_{\text{S}}(T)E,$$

where k_{S} is the specific synthesis rate of the synthetic apparatus at 25°C, and the biosynthetic efficiency decreases with temperature with a $Q_{10\text{k}} = 2$ (where Q_{10} is the 10°C temperature dependence of a process, $Q_{10\text{k}}$ being the Q_{10} for biosynthesis). The photosynthesis functional response comes from Geider *et al.* [43] (see the formulation in Moreno *et al.* [42]):

$$\mu_{\text{L}} = \frac{f(I,T)L}{1 + \phi_{\text{S}}},$$

where we allow the photosynthesis rate to have a non-trivial temperature dependence. Here T is the temperature in degrees centigrade, I is the irradiance measured in µmolphotons m^{-2} s^{-1} and ϕ_S is the carbon cost of synthesis in gC gC^{-1}. The functional response $f(I,T)$ to light is described in Moreno *et al.* [42] and depends on temperature according to a $Q_{10,photo}$ (where $Q_{10,photo}$ is the Q_{10} for light). We assume diffusion-limited growth to derive the nitrogen- and phosphorus-dependent growth rates:

$$\mu_N = \frac{4\pi D_N [N_{model}] r}{Q_N}, \quad \mu_P = \frac{4\pi D_P [P_{model}] r A}{Q_P}.$$

$$Q_N = \frac{4\pi r^3 mol_N}{3\rho p_{dry}((E + L + \alpha A/(2r))N_{prot} + \gamma_{DNA} N_{DNA} + \alpha/(2r)N_M)}$$

and $$Q_P = \frac{4\pi r^3 mol_P}{3\rho p_{dry}(EP_E + \gamma_{DNA} P_{DNA})}.$$

We treat the concentrations of bioavailable nitrogen and phosphate as latent variables, modelled using the gene frequencies for nitrogen and phosphate uptake genes in *Prochlorococcus* and *Synechococcus*, respectively.

$$\log [N_{model}] = \log [N_0] - c_N N_{gene},$$
$$\log [P_{model}] = \log [P_0] - c_P P_{gene}.$$

The terms N_0, P_0, c_N and c_P are model parameters, and N_{gene} and P_{gene} are the gene indices introduced earlier. The diffusion coefficients (D_N,D_P) decrease with temperature using $Q_{10D} = 1.5$ (where Q_{10D} is the Q_{10} for diffusion). ATOM-gene then finds the trait combination with the largest μ. At the optimal solution either

$$\mu_E = \mu_L = \mu_N < \mu_P \quad \text{(N limitation)},$$
$$\mu_E = \mu_L = \mu_P < \mu_N \quad \text{(P limitation)},$$
$$\text{or } \mu_E = \mu_L = \mu_P = \mu_N \quad \text{(co-limitation)}.$$

ATOM-gene subsequently determines C : P from this optimal strategy. If the strategy is N-limited, then we assume that the cell does luxury P storage proportional to the modelled P concentration:

$$P_{stor} = C_{stor} [P_{model}] \max(0, \mu_c - \mu),$$

where μ_c is a growth rate cutoff above which luxury storage stops.

We selected a prior probability distribution over model parameters (electronic supplementary material, table S3) and implemented ATOM-gene within the STAN probabilistic programming language [44]. We integrated C : P, N and P gene indices, temperature and irradiance (averaged over the top 50 m) and calculated the posterior probability distribution over model parameters assuming a log-normal probability distribution for C : P:

$$(C:P)_{obs} \sim \text{lognormal}((C:P)_{ATOM\text{-}gene}(I, T, N_{gene}, P_{gene}, \sigma)).$$

We performed this Bayesian optimization for the gene indices computed from both *Prochlorococcus* and *Synechococcus*, leading to a statistical model of C : P.

(i) Galbraith–Martiny and P-regression model

The Galbraith–Martiny model [45] calculates P : C as a linear function of phosphate concentration:

$$(P:C)_{GM} = 6.9 \times 10^3 [P_{obs}] + 6.0 \times 10^{-3}.$$

We also created a P-regression based model (Preg) by refitting the Galbraith–Martiny (G-M) model just to the dataset gathered here, assuming a log-normal error model:

$$(P:C)_{Preg} \sim \text{lognormal}(\kappa [P_{obs}] + [P_0], \sigma).$$

(j) Yvon-Durocher model and *T*-regression model

The Yvon-Durocher model [46] expresses phytoplankton C : P as an exponential function of temperature:

$$\log (C:P)_{YD} = \Pi(T - 15) + b,$$

where $\Pi = 0.037°C^{-1}$ and $b = 5.010$. We also created a *T*-regression based model by refitting the Yvon-Durocher model to the dataset gathered here, assuming log-normal errors:

$$(C:P)_{Treg} \sim \text{lognormal}(\Pi(T - 15) + b, \sigma).$$

(k) Moreno–Hagstrom model

The Moreno–Hagstrom model [42] uses the radius (r) and allocation of biomass to biosynthesis (E) and photosynthesis (L) to model C : P, by calculating the trait combination that leads to maximal growth for each combination of irradiance (I), temperature (T) and phosphorus (P). The Moreno–Hagstrom model models luxury P storage as a linear function of P, so that:

$$(C:P)_{MH} = \frac{1}{(C:P)_{structure} + f_{stor}[P_{obs}]}.$$

It should be noted the relationship between polyphosphate storage and ambient P concentrations has been demonstrated to have an inverse correlation in subtropical North Atlantic *Synechococcus* [47], but the direction appears to be regionally dependent [48].

3. Results

We quantified the variation in the carbon to phosphorus (C : P) elemental stoichiometry across ocean environmental gradients in the Atlantic, Indian and Pacific Oceans (figure 1). Generally, C : P ratios decreased towards colder water and higher nutrient concentrations. This pattern was present in the temperate region in the North Atlantic (figure 1*a*) and equatorial upwelling region in the Pacific Ocean (figure 1*b*). In the Indian Ocean C : P decreased toward lower phosphate concentrations and warmer water (figure 1*c*) and thus showed the opposite relationship to temperature [3]. Statistical models based solely on phosphate (G-M) or temperature (Y-D) were unable to capture the C : P trends in the Indian Ocean and showed significant biases (figure 2). All models overestimated C : P in large parts of the Indian Ocean and either over- or underestimated C : P in the equatorial Pacific Ocean. This bias remained even if we refitted the G-M and Y-D models' observations from this study suggesting a structural bias. We next tested a more complex previously published trait-based model [42], but this model had strong bias, too. Thus, existing models driven by common abiotic factors were unable to predict shifts in the elemental stoichiometry of marine communities.

The incorporation of genomically derived resource acquisition traits into our model greatly improved the prediction of regional shifts in elemental stoichiometry (figure 2, $R^2 = 0.51$ for ATOM-Syn. gene, $R^2 = 0.26$ for ATOM-Pro. gene). The models incorporating genomically derived traits remained superior in a comparison based on information criteria computed using cross-validation [49] (electronic supplementary material, table S7). We derived resource acquisition traits in *Prochlorococcus* and *Synechococcus* (the two most abundant phytoplankton in these samples) [26–28] from metagenomes. We then used relative gene frequency of nitrogen and phosphorus acquisition genes to develop an index for the

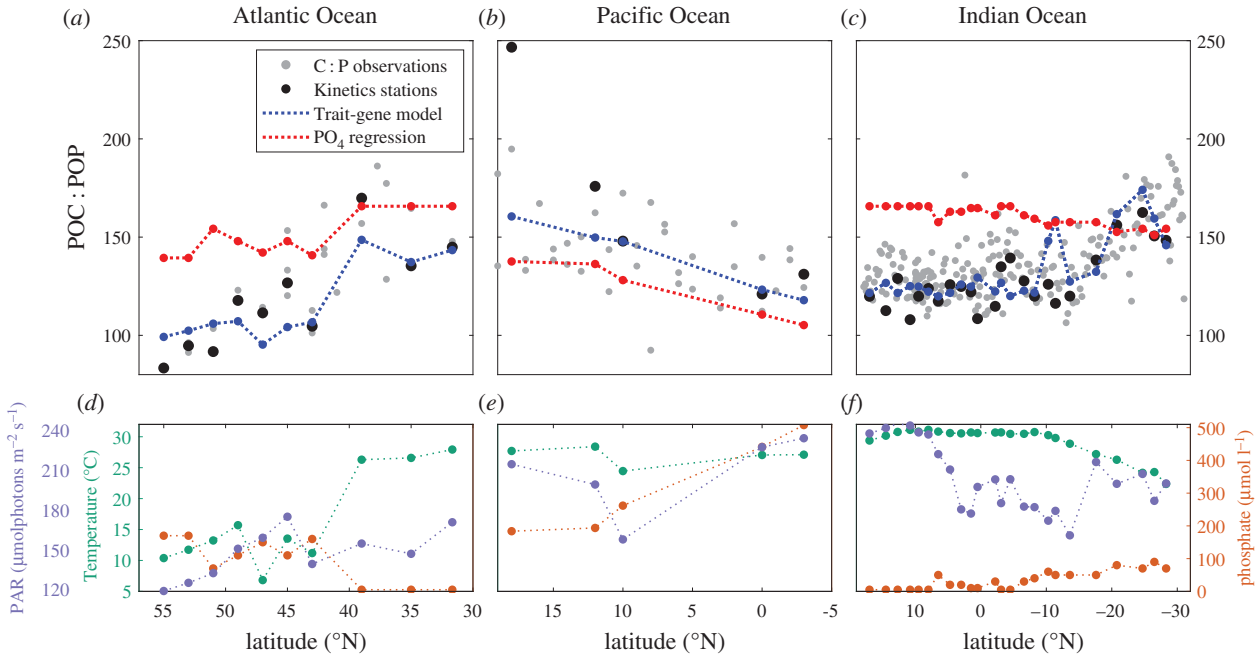

Figure 1. Observations and predictions of seston elemental stoichiometry. *In situ* measurements of particulate organic matter C : P are shown in grey, with selected stations in black where nutrient uptake incubations were performed for the (*a*) Atlantic, (*b*) Pacific and (*c*) Indian Oceans [3]. Predicted C : P is shown by the ATOM-Syn. trait-gene model (blue) and Galbraith–Martiny [43] phosphate regression model (red). Additional environmental variables of temperature (green), photosynthetically active radiation (purple) and phosphate (orange) are shown below for the (*d*) Atlantic, (*e*) Pacific and (*f*) Indian Oceans from electronic supplementary material, table S3. PAR = photosynthetically active radiation. (Online version in colour.)

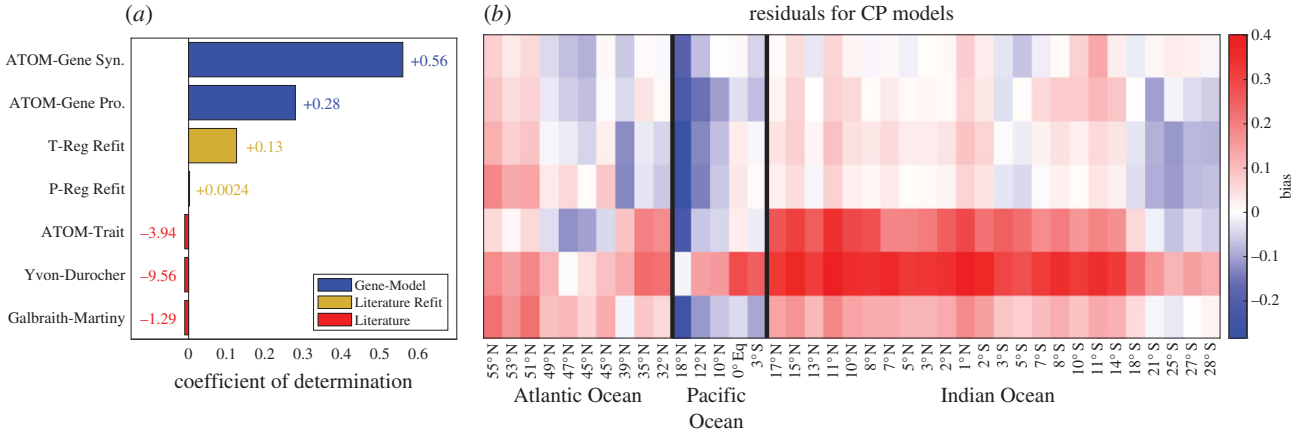

Figure 2. Trait model C : P bias. Statistical results for the predicted C : P models showing (*a*) coefficient of determination, (*b*) residuals ($\log_{10}$(predictions) – $\log_{10}$(observations)) across stations where surface C : P measurements were taken. Red indicates a positive bias, and blue negative bias. Since the distribution of C : P data looks much more log-normal we plotted the bias of the log-transformed data and models, and computed the percentage of variance of the log-transformed data that the models explained. The coefficient of determination was calculated as: $R^2 = 1 - (\text{mean}((\log_{10}(\text{observations}) - \log_{10}(\text{predictions}))^2)/\text{mean}((\log_{10}(\text{observations}) - \text{mean}(\log_{10}(\text{predictions})))^2))$. Extreme negative R^2 are plotted as a set number for legibility, with the correct value provided for the literature models (ATOM-Trait, Yvon-Durocher and Galbraith–Martiny). (Online version in colour.)

induction of nutrient acquisition machinery for each nutrient and lineage (electronic supplementary material, figure S4). This index assumes cyanobacterial lineages adapt to their environment through genome streamlining and the presence or absence of nutrient acquisition genes is directly related to nutrient stress. We found that shifts in adaptation and investment strategies for nutrient uptake led to lower bias in all the regions (figures 1 and 2). For example, this was the only model that captured the latitudinal gradient in C : P in the Indian Ocean (figure 1). ATOM-gene is a nonlinear model and predicts elevated C : P when either the N or P gene indices are close to the maximum. The difference between the North Atlantic subtropical gyre and the North Indian is that the gene indices diverge more in the subtropical North Atlantic. The P gene index is notably higher in the subtropical North Atlantic than the north Indian Ocean. Thus, the nutrient limitation is more extreme in the subtropical North Atlantic, compared with the north Indian Ocean. Similarly, the south Indian Ocean has higher C : P because the N gene index peaks there (and the same is true in a few North Pacific data points). Thus, the ATOM-gene model was able to incorporate a previously unknown pattern of nutrient gene frequencies to predict the regional shifts in C : P.

The frequency of nutrient acquisition genes helped resolve variation in nutrient stress at very low nutrient concentrations. We observed a significant correlation between

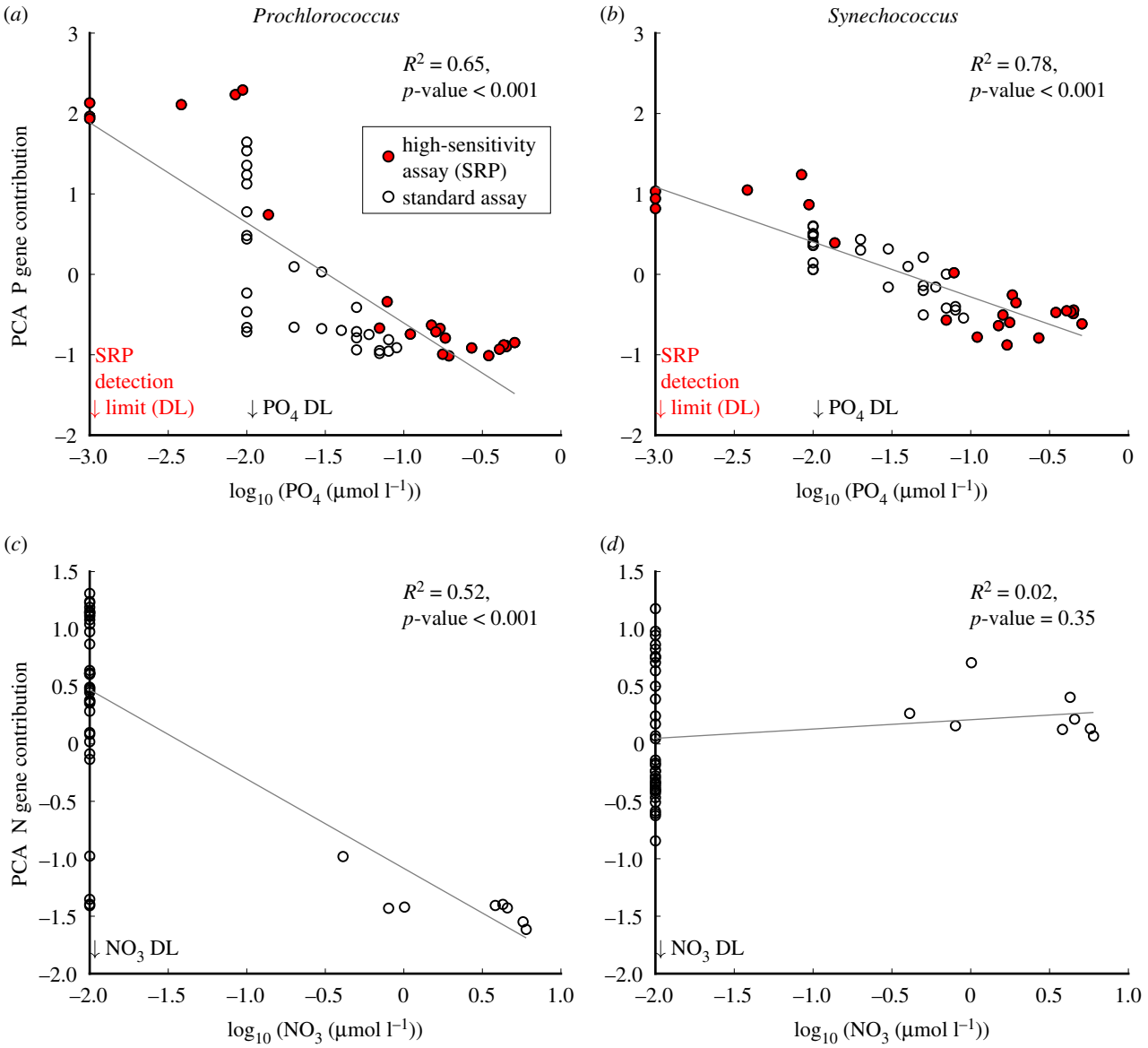

Figure 3. PCA component 1 versus nutrient concentrations. *In situ* nutrient concentrations for phosphate and nitrate are plotted against the first principal component calculated from relative gene frequencies for (*a*) *Prochlorococcus* phosphorus assimilation genes ($R^2 = 0.65$, *p*-value < 0.001), (*b*) *Synechococcus* phosphorus assimilation genes ($R^2 = 0.52$, *p*-value < 0.001), (*c*) *Prochlorococcus* nitrogen assimilation genes ($R^2 = 0.78$, *p*-value < 0.001) and (*d*) *Synechococcus* nitrogen assimilation genes ($R^2 = 0.02$, *p*-value = 0.35). High-sensitivity phosphate measurements (filled red circles) were made using a MAGIC-SRP assay [32]. Otherwise nitrate and phosphate observations were taken using standard methods (open circles) [50]. DL = detection limit. (Online version in colour.)

shifts in nutrient acquisition gene frequencies and the ambient nutrient concentration (figure 3). This was seen for both phosphorus and nitrogen acquisition genes and their respective inorganic nutrient concentrations. However, the ambient nutrient concentration of phosphorus and especially nitrogen was below detection limit in many samples. Additionally, we observed higher relative gene frequencies for iron in the subtropical Indian Ocean, equatorial Pacific and the North Atlantic in *Prochlorococcus* metagenomes (figure 4*a*). Whereas higher iron stress in the Indian Ocean overlaps with low macronutrient availability, high macronutrient availability is typical of the equatorial Pacific and temperate North Atlantic, as shown by N and P relative gene frequencies (figure 4). Here we detected large variations in gene frequencies, suggesting corresponding shifts in nutrient stress. Thus, metagenomic analyses across diverse ocean regions provided a high-sensitivity quantification of nutrient stress.

The frequency of *Prochlorococcus* acquisition genes suggested regional shifts in nutrient stress by both a single and multiple nutrients. As seen in earlier studies, we detected a high frequency of P acquisition genes for *Prochlorococcus* in the subtropical North Atlantic Ocean below 39° N, where phosphate concentrations were low (figure 4*a*) [41]. This included genes responsible for the regulation and uptake of dissolved organic P, and for arsenate detoxification, and several of unknown function. We also saw elevated P acquisition genes for *Prochlorococcus* in the north Indian Ocean and Bay of Bengal (between 1° and 17° N). By contrast, P acquisition genes were low in all samples from the Pacific Ocean and south Indian Ocean. *Prochlorococcus* N acquisition genes showed a different biogeographical pattern. Urea acquisition genes were frequent in all samples with the exception of the high nitrate areas in the equatorial Pacific Ocean and temperate waters in the North Atlantic Ocean. Nitrite and nitrate acquisition genes were frequent throughout the Indian Ocean

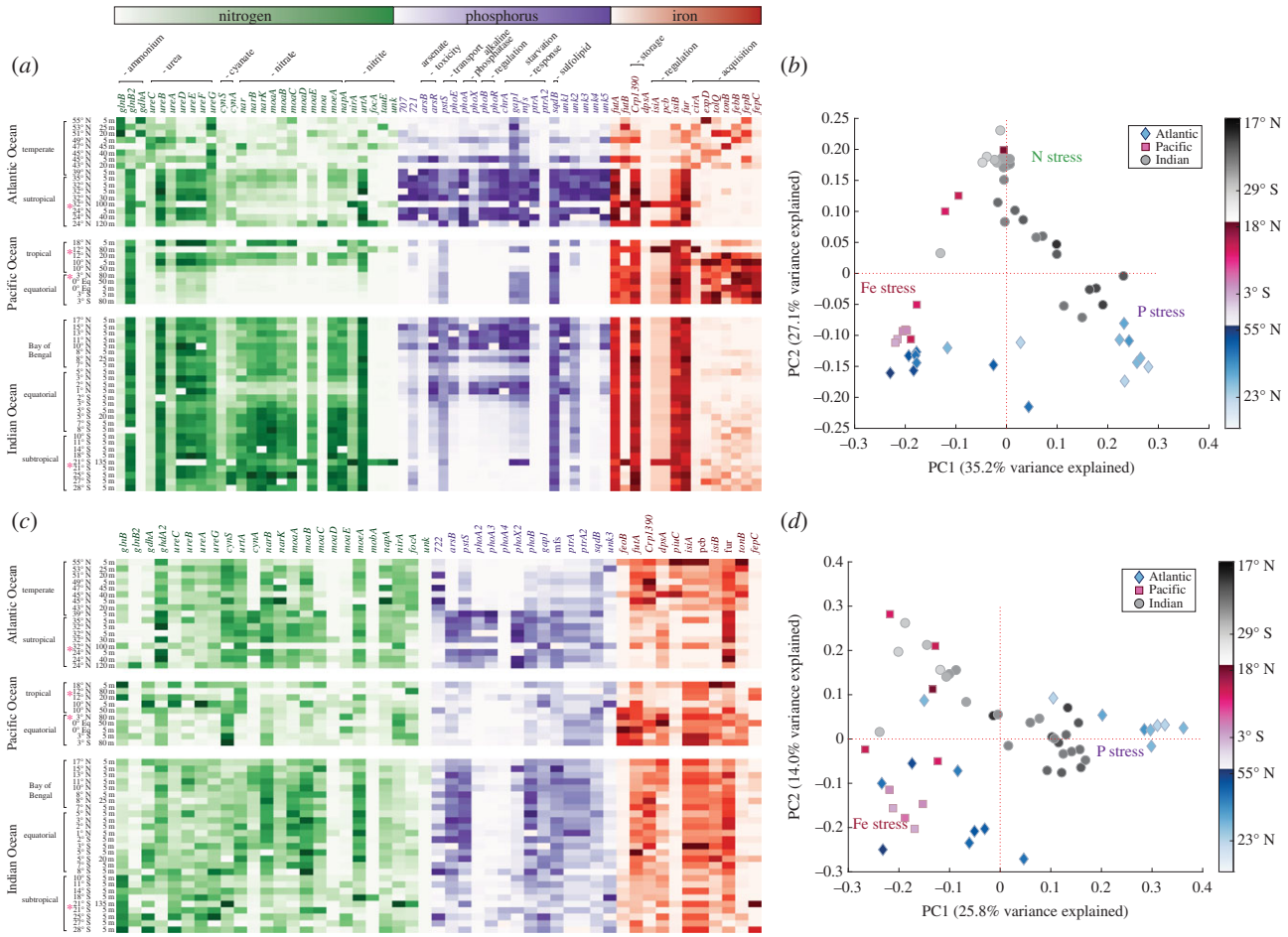

Figure 4. Variation among relative gene frequencies between stations. Green = nitrogen, purple = phosphorus, red = iron. Relative gene frequencies are normalized between 0 and 1 for each gene across stations, with white representing values near 0. Matrices based on normalized gene frequency are significantly correlated (Mantel test $R = 0.65$, p-value < 0.001). The heatmaps for (*a*) *Prochlorococcus* and (*c*) *Synechococcus* cluster the relative gene frequencies along the top according to functional role for each station row. The PCA plots for (*b*) *Prochlorococcus* and (*d*) *Synechococcus* show the variation among stations that is solely attributed to differences in relative gene frequencies. If the overall contribution of N, P or Fe genes clusters along one direction, we have added a label 'N stress', 'P stress' or 'Fe stress' in panels (*b*) and (*d*). *Prochlorococcus* relative gene frequencies cluster the stations according to these three nutrient stressors. *Synechococcus* relative gene frequencies cluster the stations mainly along two (Fe and P stress), with weak contributions from N relative gene frequencies. A red asterisk (*) indicates samples deeper than 50 m. (Online version in colour.)

(with the exception of samples on the Equator) and in the northern part of the Pacific Ocean transect. However, nitrite and nitrate genes were less common in the North Atlantic subtropical waters. Iron acquisition genes were common in the equatorial Pacific Ocean. Thus, we detected a clear biogeography of genes involved in N, P and Fe in *Prochlorococcus*.

We observed a partial correspondence between the frequency of nutrient acquisition genes in *Prochlorococcus* and *Synechococcus*, suggesting some lineage-specific adaptations to specific ocean environmental conditions (figure 4*a*). Overall, the regional shifts in *Prochlorococcus* and *Synechococcus* genome content were significantly correlated (Mantel test $R = 0.65$, p-value < 0.001). In *Synechococcus*, there was also a high frequency of P acquisition genes in the subtropical North Atlantic Ocean and north Indian Ocean (figure 4*c*). However, it appeared that the Indian Ocean area with high P acquisition genes spread further south in *Synechococcus* compared with *Prochlorococcus*. N acquisition genes were also frequent in nearly all samples for *Synechococcus*, whereas the genes were more geographically restricted in *Prochlorococcus*. There was some evidence of increase in *Synechococcus* iron acquisition genes in the equatorial Pacific Ocean, but the pattern was not strong. This method is favourable within the relatively stable environments inhabited by *Synechococcus* and *Prochlorococcus*, leading to the selection for specialized genotypes. The gene index results are more distinct for *Prochlorococcus* (figure 4), likely owing to their higher degree of genomic streamlining. Thus, the biogeographical shifts in nutrient acquisition genes were more pronounced for *Prochlorococcus* compared with *Synechococcus*.

The variation in nutrient acquisition genes may be linked to shifts in stress by one or more nutrients (figure 4*b*,*d*; electronic supplementary material, figure S4). The frequency of nutrient acquisition genes suggested P stress but also some N co-stress in the western North Atlantic Ocean and north Indian Ocean. The North Pacific Ocean and south Indian Ocean appeared to be N stressed. The equatorial Pacific Ocean was iron stressed. However, the gene frequencies suggested that a brief transition region around 10° N in the North Pacific Ocean experienced co-stress by N and Fe. *Synechococcus* appeared to be stressed by N in temperate North Atlantic Ocean waters whereas *Prochlorococcus* appeared more stressed by iron. Similarly, *Synechococcus* showed evidence of P stress in parts of the south Indian Ocean but this was not seen in *Prochlorococcus*. Shifts in the relative gene frequency corresponded to shifts in clade ecotypes (electronic supplementary material, figure S2).

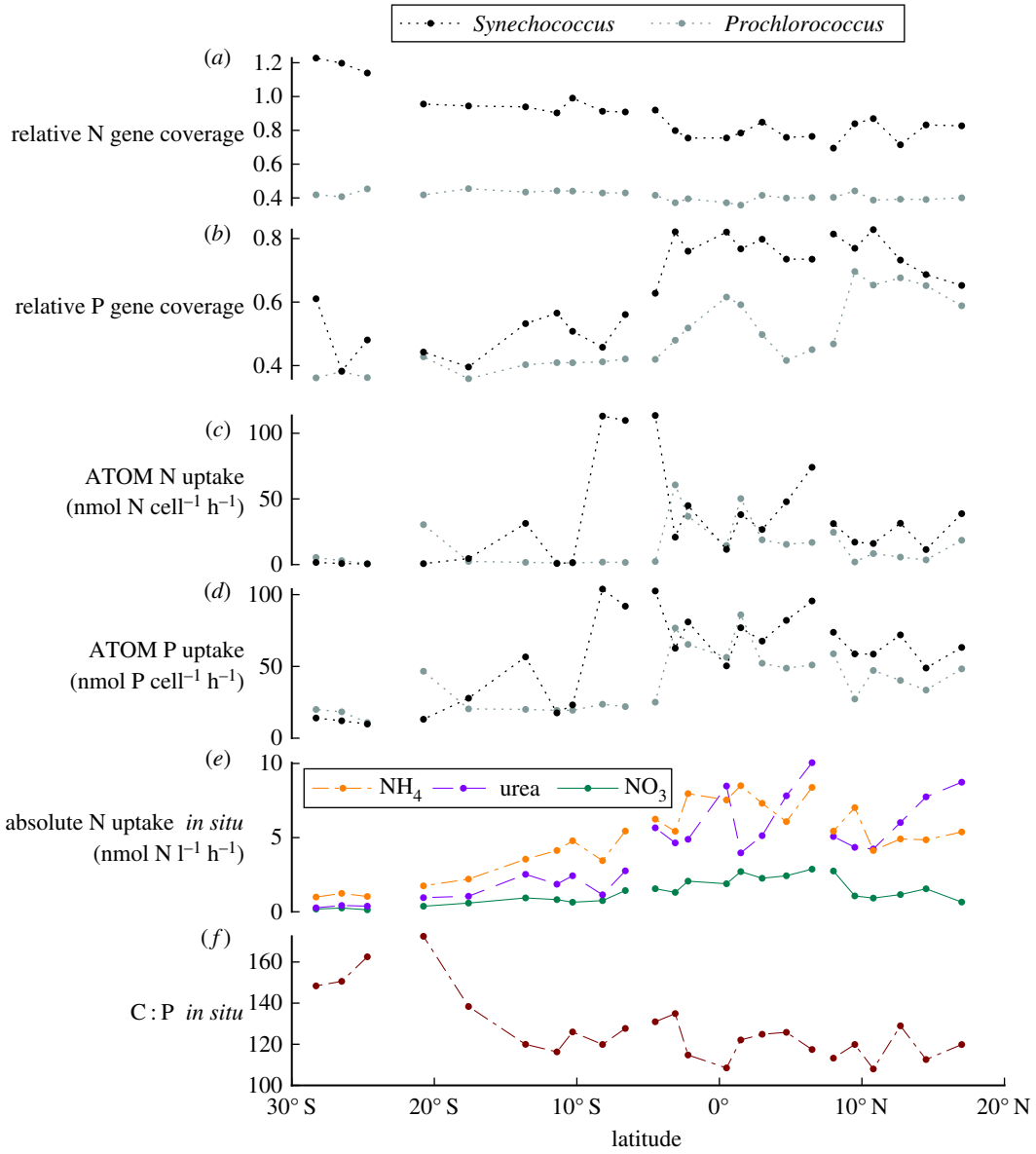

Figure 5. Evaluation of nutrient stress indices against ATOM-gene and *in situ* uptake parameters in the Indian Ocean. Relative gene frequencies of (*a*) nitrogen and (*b*) phosphorus genes are shown for *Prochlorococcus* (blue) and *Synechococcus* (black). ATOM-gene estimates for (*c*) N uptake and (*d*) P uptake normalized to cell volume are compared with the *in situ* parameters of (*e*) absolute uptake of N species (nitrate+green, urea+purple, ammonium+gold) and (*f*) the ratio of particulate organic carbon to phosphorus. *In situ* uptake rates and C : P are presented in [3,26]. Absolute uptake rates measure the accumulation of a substrate within particles. (Online version in colour.)

Thus, metagenomic analyses of phytoplankton populations suggested regional shifts in stress by one or multiple nutrients.

We used additional ecosystem measurements to verify the predictions from ATOM-gene and the overall resource investment strategies. In the Indian Ocean, uptake kinetics for the ATOM-gene model were positively correlated with observed uptake rates for nitrate, ammonium and urea (figure 5; electronic supplementary material, table S4). The implied nutrient distributions matched our observations of increasing N northwards and vice versa for P into the subtropical Indian Ocean gyre. Increases in N and P uptake rates, cellular investment in photosynthesis and biosynthesis, and cell volume corresponded to reduced nitrogen stress (electronic supplementary material, table S3). The aforementioned parameters were significantly correlated to higher *in situ* N uptake rates and lower relative N gene frequency for *Prochlorococcus* and *Synechococcus*. Phosphorus stress appeared to have little impact on C : P and cellular uptake traits in the Indian Ocean, unlike the other two ocean basins (electronic supplementary material, figure S5). Draws from the posterior-predictive distribution are shown for model parameters (electronic supplementary material, figures S6 and S7) and C : P (electronic supplementary material, figures S8 and S9). We give summaries of the posterior distribution over model parameters in electronic supplementary material, tables S8 and S9, where $\hat{R} = 1$ suggests the convergence of the Markov chain Monte Carlo integrator. Although P investment increased into the subtropical Indian Ocean gyre, there was little influence on P luxury uptake and storage (electronic supplementary material, figure S10). Only larger cells in the temperate North Atlantic exhibited P storage in the ATOM-gene model. The small number of data points with metagenome information prevented tight inference of parameter values, but the posterior distribution favours the hypothesis that the effect of nutrient stress on cell size and ribosomal content is the

strongest driver of C : P in the regions sampled, with smaller than expected roles for temperature and luxury storage. This is reflected by the posterior favouring small values of the luxury storage parameter and higher values of the Q_{10} for photosynthetic processes. Consequently, the interaction between N and P stress as seen in the genomic observations could be the underlying mechanism leading to latitudinal shifts in C : P.

4. Discussion

Linking 'omics with global biogeochemistry is a major research challenge and opportunity [51–54]. A great deal of molecular data are being generated [55,56], but there is limited current application of this new knowledge towards understanding large-scale changes in the Earth system [57]. Trait-based approaches are attractive for scaling from individual organisms to key ecosystem functions using a model intermediate [58,59]. We here use this approach as an intermediate for linking genomic information with ocean biogeochemical processes. By quantifying the spatial variation due to differences in nutrient assimilation genes, we improved our predictions of C : P across three major ocean basins (figures 1 and 2). The ATOM-gene model allowed for multiple nutrient indexes (N and P), where *in situ* nutrient concentrations were undetectable, resulting in significant improvements to the existing trait model [42]. Importantly, the gene index quantifies cyanobacterial adaptation to nutrient stressors in regions for which we have limited knowledge (e.g. the central Indian Ocean). Nutrient stress may occur through diffusive limitation at low ambient concentrations, the magnitude of nutrient fluxes, the ratio of nutrient supply or nutrient co-limitation. Additionally, both *Synechococcus* and *Prochlorococcus* can use different P and N sources [60]. Thus, genome shifts integrate these unknowns through the selective pressure to retain particular genes in nutrient-poor biomes.

The frequency of nutrient assimilation genes greatly improved our spatial understanding of nutrient stress and elemental stoichiometry of marine communities. In particular, the results showed surprising patterns of P and N stress in the less studied Indian Ocean. Our results support a recent analysis of *Synechococcus* and *Prochlorococcus* elemental quotas, suggesting a gradient of N, P and Fe stress in the Indian Ocean [61]. The Bay of Bengal showed evidence of P stress but lower N : P and C : P ratios. We attribute this contradictory observation to an interaction between N and P stress as the upregulation of P uptake proteins is restricted by N stress [62]. Culture studies have shown that N and P stress interact in controlling the overall cellular physiology and C : N : P [5]. However, it has been a challenge to translate these findings to field communities. Some of this confusion originates from difficulties in constraining external N and possibly P sources from atmospheric deposition and N-fixation. This leads to a poorly constrained *in situ* N : P supply ratio. It is unclear why we see the evidence of increased P stress near the Bay of Bengal, but it is tempting to attribute it to elevated N-fixation [8,63]. Similar to recent observations of dissolved and particulate Fe, we saw indications of Fe stress via *Prochlorococcus* Fe assimilation genes in the subtropical Indian Ocean gyre [61,64]. We also saw a high presence of Fe assimilation genes in regions with low C : P, where *Synechococcus* and *Prochlorococcus* cell abundances remained elevated [28]. As expected, this was seen for the equatorial Pacific high nutrient–low chlorophyll (HNLC) region [65]. Our data also support past studies indicating that the temperate western North Atlantic Ocean [66] and the southern Indian Ocean [61] could experience some iron stress. Thus, our genomic techniques are unveiling regions where we have a limited understanding of trace-metal stress.

Our approach is based on an assumption of rapid adaptation leading to direct association between genome content and environmental conditions [67–70]. Tropical and subtropical ocean regions have fast bacterial turnover, leading to rapid selection and genome streamlining [71]. However, environments with slow bacterial turnover may include ecotypes or genes that reflect past environmental conditions. Different lineages may also experience unique stress [72], whereas we here only analysed the abundant marine cyanobacteria. Our dataset includes few representative stations from high latitudes, where light or temperature may be the dominant selective factors [73,74]. In such conditions, transcriptomics or proteomics may be more applicable. However, these techniques suffer from their own caveats like strong diel cycles [75,76] or low correlation between RNA and protein expression [77,78]. Thus, the exact link between 'omics measurements and biogeochemical processes needs to be tailored to the system of interest.

'Omics techniques can be powerful for understanding the environmental conditions experienced by microorganisms. This principle is also applied in other ecosystem settings. A high presence of proteobacteria in the human gut may be an indicator of an imbalance in the redox potential and 'ecosystem' dysbiosis [79]. Similarly, the presence of ammonia monooxygenase may be indicative of nitrification [80]. In many ecosystems, it can be very challenging to quantify microbial physiology and stress. Thus, a carefully calibrated genomic approach could become a widespread tool to understand microbial responses to environmental changes and the biogeochemical outcomes.

Data accessibility. Uptake rate kinetics, nutrient concentrations, cell abundances and particulate elemental concentrations are available on BCO-DMO for the Atlantic AE1319 and BV46 cruises (https://www.bco-dmo.org/project/2178) and for the Indian I09 cruise (https://www.bco-dmo.org/project/628972). Nutrient data for the I09N cruise were provided by Jim Swift (Scripps Institution of Oceanography, SIO) and Susan Becker (SIO) and are available at https://cchdo.ucsd.edu. Metagenomes are available through BioProject (SRA PRJNA598881) at the following link: https://www.ncbi.nlm.nih.gov/sra/PRJNA598881. Additional environmental data and particulate elemental concentrations are available in the electronic supplementary material. A table of gene information and genomes is provided as well.

Authors' contributions. C.A.G., A.C.M. and G.I.H. wrote the manuscript. C.A.G. and A.A.L. prepared libraries and analysed sequencing data. C.A.G., L.J.U. and A.C.M. developed the nutrient gene index. G.I.H. developed the ATOM-gene model. C.A.G., A.A.L., M.W.L. and A.C.M. assisted in sample collection. M.W.L. provided uptake kinetics. A.C.M., M.W.L. and S.A.L. designed the project. Everyone edited the manuscript.

Competing interest. We have no conflict of interest.

Funding. This study was funded by the National Aeronautics and Space Administration (grant no. NESSF16R), NIH (grant no. T32AI141346), Simons Foundation (grant no. 395890) and National Science Foundation – Division of Ocean Sciences (grant nos 1559002, 1756054 and 1848576).

Acknowledgements. We would like to thank Jim Prosser and Jennifer Martiny for organizing this special issue. We would like to thank Claudia Weihe for guidance in preparing metagenome libraries, and the scientists and crews aboard the NH1418, AE1319, BV46 and I09 cruises for their effort.

References

1. Sterner RW, Elser JJ. 2002 *Ecological stoichiometry: The biology of elements from molecules to the biosphere*. Princeton, NJ: Princeton University Press. See http://books.google.com/books?hl=en&lr=&id=53NTDvppdYUC&pgis=1.
2. Moreno AR, Martiny AC. 2018 Ecological stoichiometry of ocean plankton. *Ann. Rev. Mar. Sci.* **10**, 43–69. (doi:10.1146/annurev-marine-121916-063126)
3. Garcia CA, Baer SE, Garcia NS, Rauschenberg S, Twining BS, Lomas MW, Martiny AC. 2018 Nutrient supply controls particulate elemental concentrations and ratios in the low latitude eastern Indian Ocean. *Nat. Commun.* **9**, 4868. (doi:10.1038/s41467-018-06892-w)
4. Monod J. 1950 Technique, theory and applications of continuous culture. *Ann. Inst. Pasteur* **79**, 390–410.
5. Klausmeier CA, Litchman E, Levin SA. 2004 Phytoplankton growth and stoichiometry under multiple nutrient limitation. *Limnol. Oceanogr.* **49**, 1463–1470. (doi:10.4319/lo.2004.49.4_part_2.1463)
6. Martiny AC, Pham CTA, Primeau FW, Vrugt JA, Moore JK, Levin SA, Lomas MW. 2013 Strong latitudinal patterns in the elemental ratios of marine plankton and organic matter. *Nat. Geosci.* **6**, 279–283. (doi:10.1038/ngeo1757)
7. Alexander H, Jenkins BD, Rynearson TA, Dyhrman ST. 2015 Metatranscriptome analyses indicate resource partitioning between diatoms in the field. *Proc. Natl Acad. Sci. USA* **112**, E2182–E2190. (doi:10.1073/pnas.1421993112)
8. Martiny AC *et al.* 2019 Biogeochemical controls of surface ocean phosphate. *Sci. Adv.* **5**, eaax0341. (doi:10.1126/sciadv.aax0341)
9. Shilova IN *et al*. 2017 Differential effects of nitrate, ammonium, and urea as N sources for microbial communities in the North Pacific Ocean. *Limnol. Oceanogr.* **62**, 2550–2574. (doi:10.1002/lno.10590)
10. Guidot A, Verner MC, Debaud JC, Marmeisse R. 2005 Intraspecific variation in use of different organic nitrogen sources by the ectomycorrhizal fungus *Hebeloma cylindrosporum*. *Mycorrhiza* **15**, 167–177. (doi:10.1007/s00572-004-0318-1)
11. Tapia-Torres Y, Rodríguez-Torres MD, Elser JJ, Islas A, Souza V, García-Oliva F, Olmedo-Álvarez G. 2016 How to live with phosphorus scarcity in soil and sediment: lessons from bacteria. *Appl. Environ. Microbiol.* **82**, 4652–4662. (doi:10.1128/AEM.00160-16)
12. Sosa OA, Casey JR, Karl DM. 2019 Methylphosphonate oxidation in *Prochlorococcus* strain MIT9301 supports phosphate acquisition, formate excretion, and carbon assimilation into purines. *Appl. Environ. Microbiol.* **85**, e00289-19. (doi:10.1128/AEM.00289-19)
13. Herrero A, Muro-Pastor AM, Flores E. 2001 Nitrogen control in cyanobacteria. *J. Bacteriol.* **183**, 411–425. (doi:10.1128/JB.183.2.411-425.2001)
14. Bonachela JA, Allison SD, Martiny AC, Levin SA. 2013 A model for variable phytoplankton stoichiometry based on cell protein regulation. *Biogeosciences* **10**, 4341–4356. (doi:10.5194/bg-10-4341-2013)
15. Zimmerman AE, Allison SD, Martiny AC. 2014 Phylogenetic constraints on elemental stoichiometry and resource allocation in heterotrophic marine bacteria. *Environ. Microbiol.* **16**, 1398–1410. (doi:10.1111/1462-2920.12329)
16. Martiny JBH, Jones SE, Lennon JT, Martiny AC. 2015 Microbiomes in light of traits: a phylogenetic perspective. *Science* **350**, aac9323. (doi:10.1126/science.aac9323)
17. Scanlan DJ *et al*. 2009 Ecological genomics of marine picocyanobacteria. *Microbiol. Mol. Biol. Rev.* **73**, 249–299. (doi:10.1128/mmbr.00035-08)
18. Morris JJ, Lenski RE, Zinser ER. 2012 The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *mBio* **3**, e00036-12. (doi:10.1128/mBio.00036-12)
19. Martiny AC, Coleman ML, Chisholm SW. 2006 Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. *Proc. Natl Acad. Sci.* **103**, 12 552–12 557. (doi:10.1073/pnas.0601301103)
20. Coleman ML, Chisholm SW. 2010 Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc. Natl Acad. Sci. USA* **107**, 18 634–18 639. (doi:10.1073/pnas.1009480107)
21. Martiny AC, Huang Y, Li W. 2009 Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environ. Microbiol.* **11**, 1340–1347. (doi:10.1111/j.1462-2920.2009.01860.x)
22. Antelmann H, Scharf C, Hecker M. 2000 Phosphate starvation-inducible proteins of *Bacillus subtilis*: proteomics and transcriptional analysis. *J. Bacteriol.* **182**, 4478–4490. (doi:10.1128/JB.182.16.4478-4490.2000)
23. Martiny AC, Kathuria S, Berube PM. 2009 Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proc. Natl Acad. Sci. USA* **106**, 10 787–10 792. (doi:10.1073/pnas.0902532106)
24. Berube PM *et al*. 2015 Physiology and evolution of nitrate acquisition in *Prochlorococcus*. *ISME J.* **9**, 1195–1207. (doi:10.1038/ismej.2014.211)
25. Lomas MW, Bonachela JA, Levin SA, Martiny AC. 2014 Impact of ocean phytoplankton diversity on phosphate uptake. *Proc. Natl Acad. Sci. USA* **111**, 17 540–17 545. (doi:10.1073/pnas.1420760111)
26. Baer SE, Rauschenberg S, Garcia CA, Garcia NS, Martiny AC, Twining BS, Lomas MW. 2018 Carbon and nitrogen productivity during spring in the oligotrophic Indian Ocean along the GO-SHIP IO9 N transect. *Deep Sea Res. II Top. Stud. Oceanogr.* **161**, 81–91. (doi:10.1016/J.DSR2.2018.11.008)
27. Baer SE, Lomas MW, Terpis KX, Mouginot C, Martiny AC. 2017 Stoichiometry of *Prochlorococcus*, *Synechococcus*, and small eukaryotic populations in the western North Atlantic Ocean. *Environ. Microbiol.* **19**, 1568–1583. (doi:10.1111/1462-2920.13672)
28. Kent AG, Baer SE, Mouginot C, Huang JS, Larkin AA, Lomas MW, Martiny AC. 2019 Parallel phylogeography of *Prochlorococcus* and *Synechococcus*. *ISME J.* **13**, 430–441. (doi:10.1038/s41396-018-0287-6)
29. Larkin AA, Garcia CA, Ingoglia KA, Garcia NS, Baer SE, Twining BS, Lomas MW, Martiny AC. 2019 Subtle biogeochemical regimes in the Indian Ocean revealed by spatial and diel frequency of *Prochlorococcus* haplotypes. *Limnol. Oceanogr.* **65**, S220–S232. (doi:10.1002/lno.11251)
30. Steinberg DK, Carlson CA, Bates NR, Johnson RJ, Michaels AF, Knap AH. 2001 Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-scale look at ocean biology and biogeochemistry. *Deep Sea Res. II Top. Stud. Oceanogr.* **48**, 1405–1447. (doi:10.1016/S0967-0645(00)00148-X)
31. Lomas MW, Burke AL, Lomas DA, Bell DW, Shen C, Dyhrman ST, Ammerman JW. 2010 Sargasso Sea phosphorus biogeochemistry: an important role for dissolved organic phosphorus (DOP). *Biogeosciences* **7**, 695–710. (doi:10.5194/bg-7-695-2010)
32. Karl DM, Tien G. 1992 MAGIC: a sensitive and precise method for measuring dissolved phosphorus in aquatic environments. *Limnol. Oceanogr.* **37**, 105–116. (doi:10.4319/lo.1992.37.1.0105)
33. Larkin AA, Martiny AC. 2017 Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. *Environ. Microbiol. Rep.* **9**, 55–70. (doi:10.1111/1758-2229.12523)
34. Boström KH, Simu K, Hagström Å, Riemann L. 2004 Optimization of DNA extraction for quantitative marine bacterioplankton community analysis. *Limnol. Oceanogr. Methods* **2**, 365–373. (doi:10.4319/lom.2004.2.365)
35. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. (doi:10.1093/bioinformatics/btu170)
36. Langmead B, Salzberg SL. 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. (doi:10.1038/nmeth.1923)
37. Eren AM, Esen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015 Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319. (doi:10.7717/peerj.1319)
38. Buchfink B, Xie C, Huson DH. 2014 Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60. (doi:10.1038/nmeth.3176)
39. Malmstrom RR *et al*. 2013 Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *ISME J.* **7**, 184–198. (doi:10.1038/ismej.2012.89)
40. Robidart JC, Magasin JD, Shilova IN, Turk-Kubo KA, Wilson ST, Karl DM, Scholin CA, Zehr JP. 2019 Effects of nutrient enrichment on surface microbial

community gene expression in the oligotrophic North Pacific Subtropical Gyre. *ISME J.* **13**, 374–387. (doi:10.1038/s41396-018-0280-0)

41. Martiny AC, Ustick L, Garcia CA, Lomas MW. 2019 Genomic adaptation of marine phytoplankton populations regulates phosphate uptake. *Limnol. Oceanogr.* **65**, S340–S350. (doi:10.1002/lno.11252)
42. Moreno AR, Hagstrom GI, Primeau FW, Levin SA, Martiny AC. 2018 Marine phytoplankton stoichiometry mediates nonlinear interactions between nutrient supply, temperature, and atmospheric CO_2. *Biogeosciences* **15**, 2761–2779. (doi:10.5194/bg-15-2761-2018)
43. Geider RJ, Macintyre HL, Kana TM. 1996 A dynamic model of photoadaptation in phytoplankton. *Limnol. Oceanogr.* **41**, 1–15. (doi:10.4319/lo.1996.41.1.0001)
44. Carpenter B *et al*. 2017 Stan: a probabilistic programming language. *J. Stat. Softw.* **76**, 1–32. (doi:10.18637/jss.v076.i01)
45. Galbraith ED, Martiny AC. 2015 A simple nutrient-dependence mechanism for predicting the stoichiometry of marine ecosystems. *Proc. Natl Acad. Sci. USA* **112**, 8199–8204. (doi:10.1073/pnas.1423917112)
46. Yvon-Durocher G, Dossena M, Trimmer M, Woodward G, Allen AP. 2015 Temperature and the biogeography of algal stoichiometry. *Glob. Ecol. Biogeogr.* **24**, 562–570. (doi:10.1111/geb.12280)
47. Martin P, Dyhrman ST, Lomas MW, Poulton NJ, Van Mooy BAS. 2014 Accumulation and enhanced cycling of polyphosphate by Sargasso Sea plankton in response to low phosphorus. *Proc. Natl Acad. Sci. USA* **111**, 8089–8094. (doi:10.1073/pnas.1321719111)
48. Li J, Dittrich M. 2019 Dynamic polyphosphate metabolism in cyanobacteria responding to phosphorus availability. *Environ. Microbiol.* **21**, 572–583. (doi:10.1111/1462-2920.14488)
49. Vehtari A, Gelman A, Gabry J. 2017 Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432. (doi:10.1007/s11222-016-9696-4)
50. Hood E, Sabine C, Sloyan B (eds). 2010 The GO-SHIP *Repeat hydrography manual: a collection of expert reports and guidelines*. Report no. 14, ICPO Publication Series no. 134. IOCCP. See http://www.go-ship.org/HydroMan.html.
51. Mock T, Daines SJ, Geider R, Collins S, Metodiev M, Millar AJ, Moulton V, Lenton TM. 2016 Bridging the gap between omics and Earth system science to better understand how environmental change impacts marine microbes. *Glob. Chang. Biol.* **22**, 61–75. (doi:10.1111/gcb.12983)
52. Coles VJ *et al*. 2017 Ocean biogeochemistry modeled with emergent trait-based genomics. *Science* **358**, 1149–1154. (doi:10.1126/science.aan5712)
53. Hennon GMM, Dyhrman ST. 2019 Progress and promise of omics for predicting the impacts of climate change on harmful algal blooms. *Harmful Algae.* **91**, 101587. (doi:10.1016/j.hal.2019.03.005)
54. Caputi L *et al*. 2019 Community-level responses to iron availability in open ocean plankton ecosystems. *Global Biogeochem. Cycles* **33**, 391–419. (doi:10.1029/2018GB006022)
55. Sunagawa S *et al*. 2015 Structure and function of the global ocean microbiome. *Science* **348**, 1261359. (doi:10.1126/science.1261359)
56. Venter JC *et al*. 2004 Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74. (doi:10.1126/science.1093857)
57. Moran MA. 2015 The global ocean microbiome. *Science* **350**, aac8455. (doi:10.1126/science.aac8455)
58. Kiørboe T, Visser A, Andersen KH. 2018 A trait-based approach to ocean ecology. *ICES J. Mar. Sci.* **75**, 1849–1863. (doi:10.1093/icesjms/fsy090)
59. Talmy D, Blackford J, Hardman-Mountford NJ, Dumbrell AJ, Geider RJ. 2013 An optimality model of photoadaptation in contrasting aquatic light regimes. *Limnol. Oceanogr.* **58**, 1802–1818. (doi:10.4319/lo.2013.58.5.1802)
60. Moore LR, Post AF, Rocap G, Chisholm SW. 2002 Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol. Oceanogr.* **47**, 989–996. (doi:10.4319/lo.2002.47.4.0989)
61. Twining BS, Rauschenberg S, Baer SE, Lomas MW, Martiny AC, Antipova O. 2019 A nutrient limitation mosaic in the eastern tropical Indian Ocean. *Deep Sea Res. II Top. Stud. Oceanogr.* **166**, 125–140. (doi:10.1016/j.dsr2.2019.05.001)
62. Bonachela JA, Raghib M, Levin SA. 2011 Dynamic model of flexible phytoplankton nutrient uptake. *Proc. Natl Acad. Sci. USA* **108**, 20 633–20 638. (doi:10.1073/pnas.1118012108)
63. Wang WL, Moore JK, Martiny AC, Primeau FW. 2019 Convergent estimates of marine nitrogen fixation. *Nature* **566**, 205–211. (doi:10.1038/s41586-019-0911-2)
64. Grand MM, Measures CI, Hatta M, Hiscock WT, Landing WM, Morton PL, Buck CS, Barrett PM, Resing JA. 2015 Dissolved Fe and Al in the upper 1000 m of the eastern Indian Ocean: a high-resolution transect along 95°E from the Antarctic margin to the Bay of Bengal. *Global Biogeochem. Cycles* **29**, 375–396. (doi:10.1002/2014GB004920)
65. Coale KH *et al*. 1996 A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific Ocean. *Nature* **383**, 495–501. (doi:10.1038/383495a0)
66. Rijkenberg MJA, Middag R, Laan P, Gerringa LJA, Van Aken HM, Schoemann V, De Jong JTM, De Baar HJW. 2014 The distribution of dissolved iron in the West Atlantic Ocean. *PLoS ONE* **9**, e101323. (doi:10.1371/journal.pone.0101323)
67. Partensky F, Garczarek L. 2010 *Prochlorococcus*: advantages and limits of minimalism. *Ann. Rev. Mar. Sci.* **2**, 305–331. (doi:10.1146/annurev-marine-120308-081034)
68. Tripp HJ, Bench SR, Turk KA, Foster RA, Desany BA, Niazi F, Affourtit JP, Zehr JP. 2010 Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. *Nature* **464**, 90–94. (doi:10.1038/nature08786)
69. Giovannoni SJ *et al*. 2005 Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**, 1242–1245. (doi:10.1126/science.1114057)
70. Swan BK *et al*. 2013 Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc. Natl Acad. Sci. USA* **110**, 11 463–11 468. (doi:10.1073/pnas.1304246110)
71. Giovannoni SJ, Thrash JC, Temperton B. 2014 Implications of streamlining theory for microbial ecology. *ISME J.* **8**, 1553–1565. (doi:10.1038/ismej.2014.60)
72. Alexander H, Rouco M, Haley ST, Wilson ST, Karl DM, Dyhrman ST. 2015 Functional group-specific traits drive phytoplankton dynamics in the oligotrophic ocean. *Proc. Natl Acad. Sci. USA* **112**, E5972–E5979. (doi:10.1073/pnas.1518165112)
73. Dickman EM, Vanni MJ, Horgan MJ. 2006 Interactive effects of light and nutrients on phytoplankton stoichiometry. *Oecologia* **149**, 676–689. (doi:10.1007/s00442-006-0473-5)
74. Thomas MK, Kremer CT, Litchman E. 2016 Environment and evolutionary history determine the global biogeography of phytoplankton temperature traits. *Glob. Ecol. Biogeogr.* **25**, 75–86. (doi:10.1111/geb.12387)
75. Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP, Moran MA. 2009 Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environ. Microbiol.* **11**, 1358–1375. (doi:10.1111/j.1462-2920.2008.01863.x)
76. Ottesen EA, Young CR, Gifford SM, Eppley JM, Marin R, Schuster SC, Scholin CA, DeLong EF. 2014 Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science* **345**, 207–212. (doi:10.1126/science.1252476)
77. Jayapal KP, Philp RJ, Kok Y-J, Yap MGS, Sherman DH, Griffin TJ, Hu W-S. 2008 Uncovering genes with divergent mRNA-protein dynamics in *Streptomyces coelicolor*. *PLoS ONE* **3**, e2097. (doi:10.1371/journal.pone.0002097)
78. Maier T, Schmidt A, Güell M, Kühner S, Gavin AC, Aebersold R, Serrano L. 2011 Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol. Syst. Biol.* **7**, 511. (doi:10.1038/msb.2011.38)
79. Shin NR, Whon TW, Bae JW. 2015 Proteobacteria: microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol.* **33**, 496–503. (doi:10.1016/j.tibtech.2015.06.011)
80. Francis CA, Roberts KJ, Beman JM, Santoro AE, Oakley BB. 2005 Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc. Natl Acad. Sci. USA* **102**, 14 683–14 688. (doi:10.1073/pnas.0506625102)