nature microbiology

LETTERS

PUBLISHED: 29 FEBRUARY 2016 | ARTICLE NUMBER: 16005 | DOI: 10.1038/NMICROBIOL.2016.5

Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom

David M. Needham and Jed A. Fuhrman*

Marine phytoplankton perform approximately half of global carbon fixation, with their blooms contributing disproportionately to carbon sequestration[1], and most phytoplankton production is ultimately consumed by heterotrophic prokaryotes[2]. Therefore, phytoplankton and heterotrophic community dynamics are important in modelling carbon cycling and the impacts of global change[3]. In a typical bloom, diatoms dominate initially, transitioning over several weeks to smaller and motile phytoplankton[4]. Here, we show unexpected, rapid community variation from daily rRNA analysis of phytoplankton and prokaryotic community members following a bloom off southern California. Analysis of phytoplankton chloroplast 16S rRNA demonstrated ten different dominant phytoplankton over 18 days alone, including four taxa with animal toxin-producing strains. The dominant diatoms, flagellates and picophytoplankton varied dramatically in carbon export potential. Dominant prokaryotes also varied rapidly. Euryarchaea briefly became the most abundant organism, peaking over a few days to account for about 40% of prokaryotes. Phytoplankton and prokaryotic communities correlated better with each other than with environmental parameters. Extending beyond the traditional view of blooms being controlled primarily by physics and inorganic nutrients, these dynamics imply highly heterogeneous, continually changing conditions over time and/or space and suggest that interactions among microorganisms are critical in controlling plankton diversity, dynamics and fates.

Phytoplankton blooms are typically triggered by increases in nutrients and irradiance, in concert with partial stratification, leading to algal growth outpacing grazing, infection, sinking and other losses and resulting in biomass accumulation in the upper photic zone[4]. Most studies rely on chlorophyll or coarsely resolving phytoplankton identification methods to observe bloom progression and demise, with few following the day-to-day changes in species composition. However, different phytoplankton, even close relatives, can have very different properties, including interactions with each other or with various heterotrophs[5,6]. Bloom-associated prokaryotic taxa are capable of rapidly using various organic products of healthy, senescent or decaying phytoplankton[7]. These utilizations may be classified as positive/mutualistic, parasitic or pathogenic[5,6]. Heterotrophic bacteria are reported to consume ~15–85% of fixed carbon during a bloom, clearly influencing its fate[8]. Roseobacteria, Flavobacteria and Gammaproteobacteria are among the relatively few bacterial taxa reported to respond to blooms quickly, with successional patterns over several days to weeks[7,9–11]. Notably, archaea have largely been ignored as possible agents in phytoplankton dynamics, although one recent study indicated a euryarchaeal association with eukaryotic phytoplankton[12] and another reported high euryarchaeal abundance on a spring sample date[13]. The extent and speed of variation in the composition of phytoplankton, bacteria and archaea on a day-to-day scale remains relatively unknown, but is important for understanding the mechanisms controlling the base of marine food webs in a changing ocean[3].

We sampled almost daily during and shortly after a region-wide spring bloom (in March 2011, Supplementary Fig. 1) at the site of the San Pedro Ocean Time Series (SPOT; latitude 33°33′ N, longitude 118°24′ W; 900 m total water depth; 20 km offshore of southern California), followed by daily-to-weekly sampling through August. Strong winds in the region (>25 km h^{-1} on average) a few days before the start of the study probably stimulated phytoplankton growth by mixing nutrients to the surface (although the nutrients were almost depleted at our initial sampling), with a chlorophyll concentration of 10.3 μg l^{-1} on the first day of sampling, 12 March (Supplementary Fig. 2). By 17 March, the concentration of chlorophyll dropped to ~2 μg l^{-1} for several days while nutrients continued to decrease (Supplementary Fig. 2). On 20 March, a storm brought 2.5 cm of rain and a 1 °C drop in the surface water temperature, coinciding with a large increase in inorganic nutrients (most probably from wind-based deepening of the mixed layer), producing a secondary increase in chlorophyll (up to 3.1 μg l^{-1}) that was observed region-wide (Supplementary Fig. 1) and lasted for about a week. Over the remainder of the time series (through August), temperatures generally increased and chlorophyll concentrations remained low (<2.5 μg l^{-1}; average of 0.67 μg l^{-1}) (Supplementary Figs 1–3).

The microbial communities constituting the bloom were measured by high-throughput sequencing of rRNA genes, thus bypassing some of the shortcomings of microscopy where close relatives are not easily distinguished and small organisms have few identifying morphological characteristics. We separately analysed the <1 μm and >1 μm size fractions to distinguish free-living from larger and attached microorganisms, and used a three-domain approach shown by mock community analysis to measure 16S rRNA gene abundances with minimal bias (observed versus expected $r^2 = 0.95$; ref. 14). We analysed 16S rRNA genes of chloroplasts, not only protistan nuclear 18S rRNA, to allow assessment of the dynamics of photosynthetic protists without the uncertainties of distinguishing heterotrophic and autotrophic protists, considering ambiguous identifications due to incomplete databases and numerous protists with unknown physiology[15]. The number of chloroplast 16S rRNA gene copy numbers per cell varies between taxa, with most small taxa having 1–2, but very large taxa can have dozens to hundreds (Supplementary Table 1). Chloroplast numbers, like other cell components (including all nuclear genes), vary by a factor of two over a phytoplankton life cycle (as the cell prepares to divide), and environmental factors may also influence the number of chloroplasts per cell, particularly in taxa with many chloroplasts (for example, very large diatoms[16]). For the taxa at our location, the chloroplast 16S rRNA gene copy numbers would be expected to vary between taxa less than for the 18S genes (which range from <10 to ~50,000 copies per cell[15,17]) and are highly variable even

Department of Biological Sciences, University of Southern California, Los Angeles, California 90089-0371, USA. *e-mail: fuhrman@usc.edu

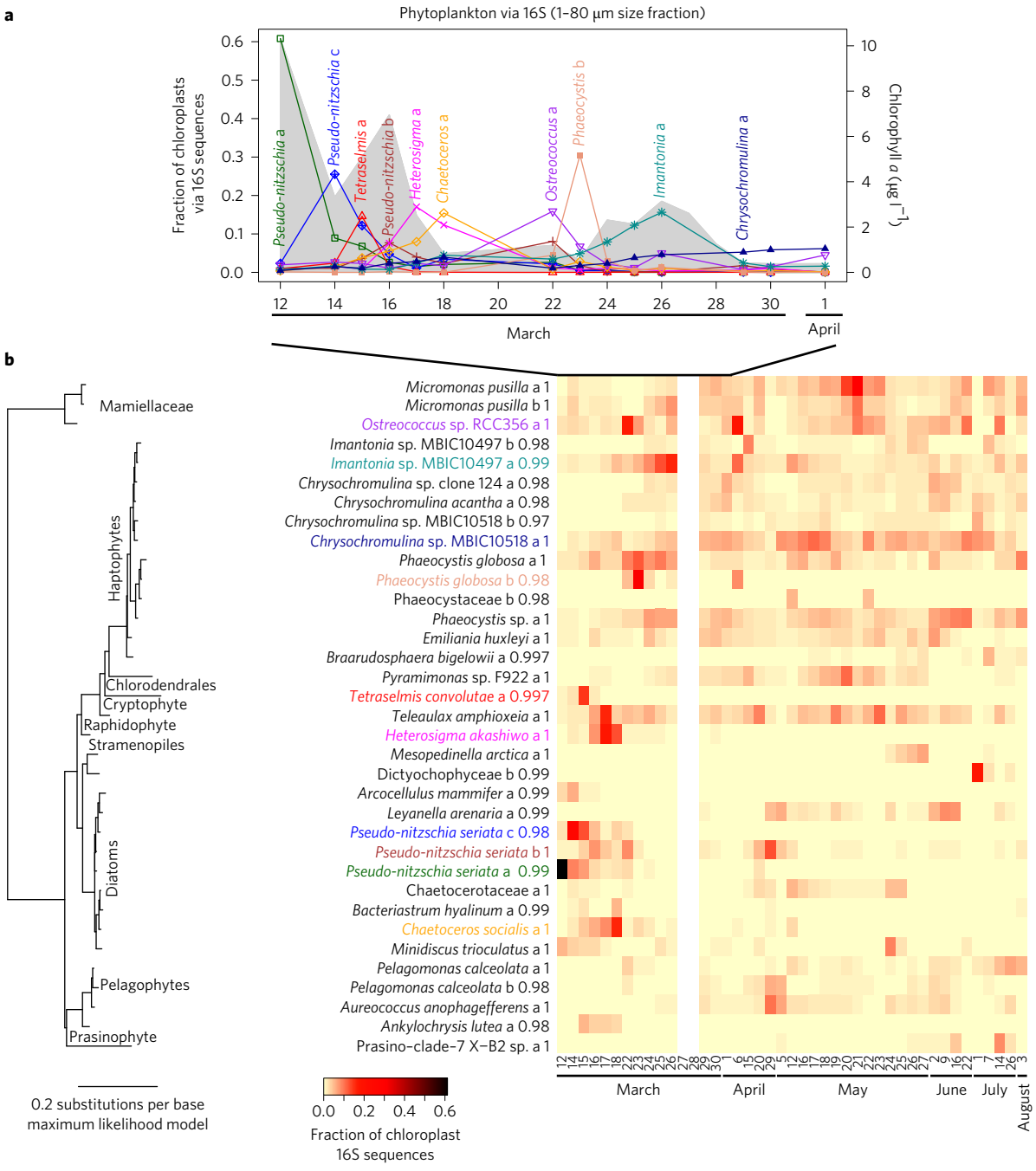

Figure 1 | Dynamics of eukaryotic phytoplankton taxa (99% operational taxonomic units (OTUs)) as measured by 16S rRNA sequences of chloroplasts. **a**,**b**, The results show near-daily change of the dominant taxa during March (**a**) and continued dynamics of the diverse community but relative stability after March (**b**). In **a**, only the taxa that became the most represented sequence on at least a single day are shown. A grey background indicates chlorophyll concentration. In **b**, all phytoplankton taxa that were >5% on any day, or >1% on average are shown. The taxonomy information for each OTU is the name associated with their best sequence match within the NCBI or PhytoRef database (see Methods). The letters following the classification indicate the rank (a, most abundant; b, less; and so on) of the OTU within equivalent classifications. The value following the letter indicates the fraction of sequence similarity to its database match. Note the *Micromonas pusilla* OTUs were 100% matches to different strains (Supplementary Fig. 10). Full taxonomic classifications are provided in Supplementary Data Set 1.

among similarly sized organisms (Supplementary Table 1). Thus, the chloroplast 16S rRNA gene abundance, tracking the organelles responsible for photosynthesis, may provide a view of phytoplankton dominance that is less distorted by genomic variation.

As expected, diatoms initially dominated the phytoplankton bloom, and smaller flagellates or non-motile picophytoplankton dominated subsequently (Fig. 1 and Supplementary Figs 4–6). The temporal and taxonomic resolution of our methods allowed us to obtain greatly enhanced details of the classical view of phytoplankton dynamics. The most abundant taxon, characterized via sequence abundance, with classification via best match of the 16S rRNA genes of chloroplasts to NCBI and PhytoRef databases, changed on each of the six sampling days during the primary bloom (12 to 18 March), and included three *Pseudo-nitzschia* types (diatoms; one was a 100% match to *Pseudo-nitzschia seriata*): *Chaetoceros socialis* (diatom), *Tetraselmis convolutae*

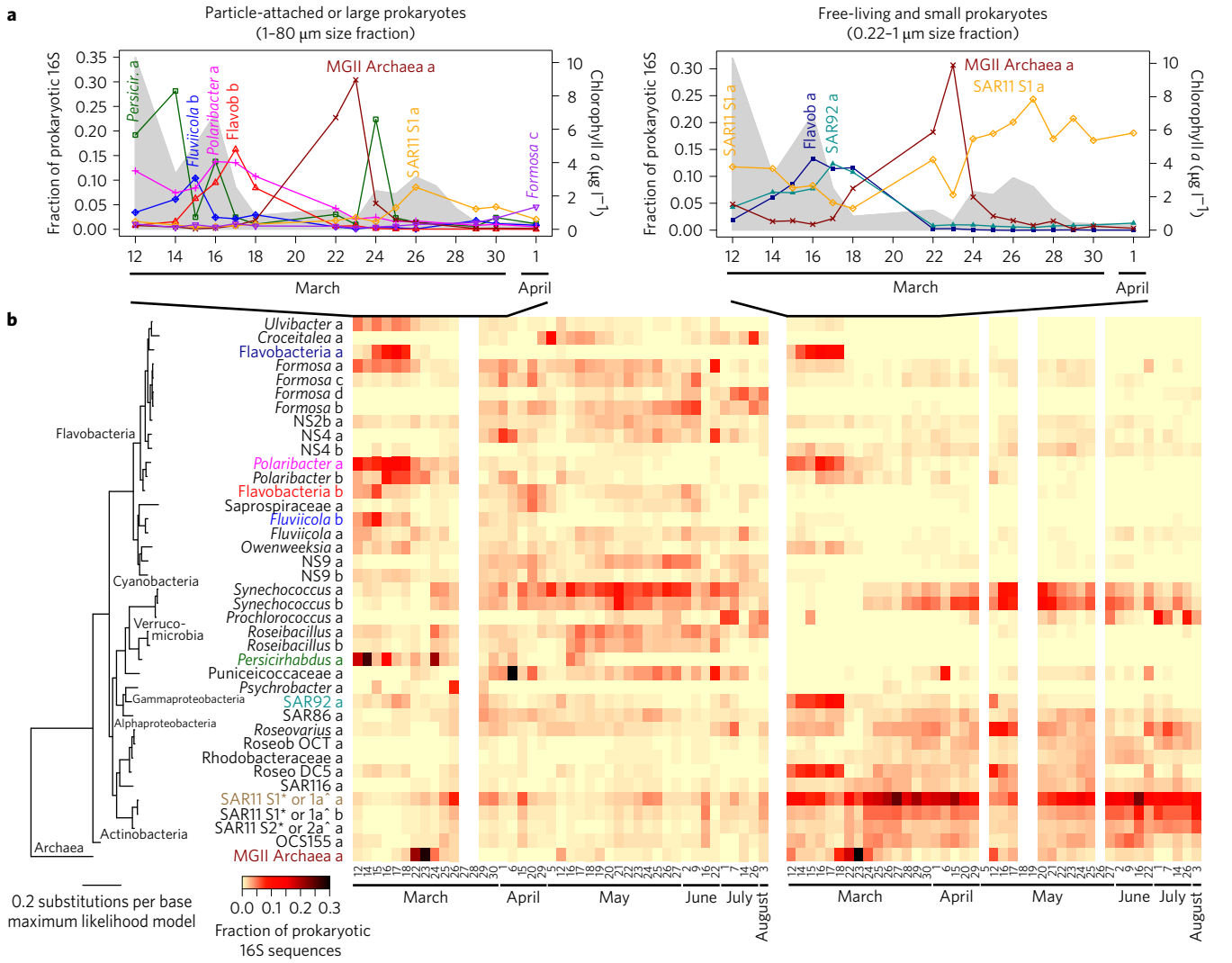

Figure 2 | Dynamics of bacterial and archaeal 99% OTUs. a,**b**, Results are shown for March (**a**) and over the full time series selected as in Fig. 1 (**b**). Left and right panels show results for the 1–80 µm size fraction and the 0.22–1 µm size fraction, respectively. Full classifications are provided in Supplementary Data Set 1. SAR11 taxonomy includes subclade classifications from the SILVA database (*) and from ref. 30 (^).

(prasinophyte, possibly a radiolarian symbiont) and *Heterosigma akashiwo* (raphidophyte). After the initial bloom subsided, *Phaeocystis globosa* (prymnesiophyte) and *Ostreococcus lucimarinus* (chlorophyte) became the most highly represented types. Subsequently, the increase in chlorophyll on 24 to 28 March was dominated by small phytoplankton: prymnesiophytes (*Imantonia* and *Chrysochromulina*) and *Ostreococcus* (Fig. 1). Throughout the next five months, the initial bloom types were rare, and a broad and fluctuating variety of mostly small phytoplankton dominated (Fig. 1b and Supplementary Figs 4–6).

Phytoplankton dynamics via 18S rRNA gene sequences were highly concordant with the 16S rRNA gene results (Mantel rho = 0.90, $P \leq 0.001$) with most taxa strongly correlated between the approaches (Supplementary Fig. 5) and with nine successive dominants over 18 days. Via 18S sequences, the dinoflagellate *Lingulodinium* (= *Gonyaulax*) *polyedrum* was dominant on 15 March (Supplementary Figs 4 and 6); dinoflagellates are the only major group not represented in the 16S rRNA gene plastid data due to aberrant sequences[18]. Additionally, via 18S rRNA gene sequences, the coccolithophore *Emiliania huxleyi* was found to be dominant on 16 March, a second *Phaeocystis* was dominant on 25 March, and the diatom *Guinardia* was dominant from 29 March to 1 April. Although via 16S chloroplast rRNA genes these had smaller proportions, their 16S and 18S rRNA gene temporal variations were extremely similar (Spearman's $r = 0.8–0.9$; Supplementary Fig. 5). When considering all phytoplankton, the temporal patterns and taxon names occasionally differed between 16S and 18S rRNA genes, most probably because of a variation in copy numbers, but also due to differential coverage of the databases, different conservation of these genes between strains, and primer biases. Note that *Lingulodinium* and *Guinardia* (a large diatom) are expected to have particularly high 18S rRNA gene copy numbers[19].

The phytoplankton variation we observed has profound implications for ecosystem function. The particular blooming species observed include well-known 'harmful algal bloom' organisms *Pseudo-nitzschia* (amnesic shellfish poisoning), *Heterosigma* (fish poisoning) and *Lingulodinium* (paralytic shellfish poisoning)[20], while *Phaeocystis* and *E. huxleyi* are major emitters of the dimethyl sulfide precursor DMSP[21]. Export would vary considerably in concert with rapidly changing composition[22]. Diatoms, for example, with their relatively heavy silica frustules, generally sink rapidly upon nutrient depletion, although the rates can depend on cell size and shape[23]. *E. huxleyi*, often calcified, can aggregate and sink rapidly. The mineral-rich exteriors of diatoms and *E. huxleyi* can enhance flocculation of other organic material, forming 'marine snow', further increasing export[24]. The domination of motile and/or

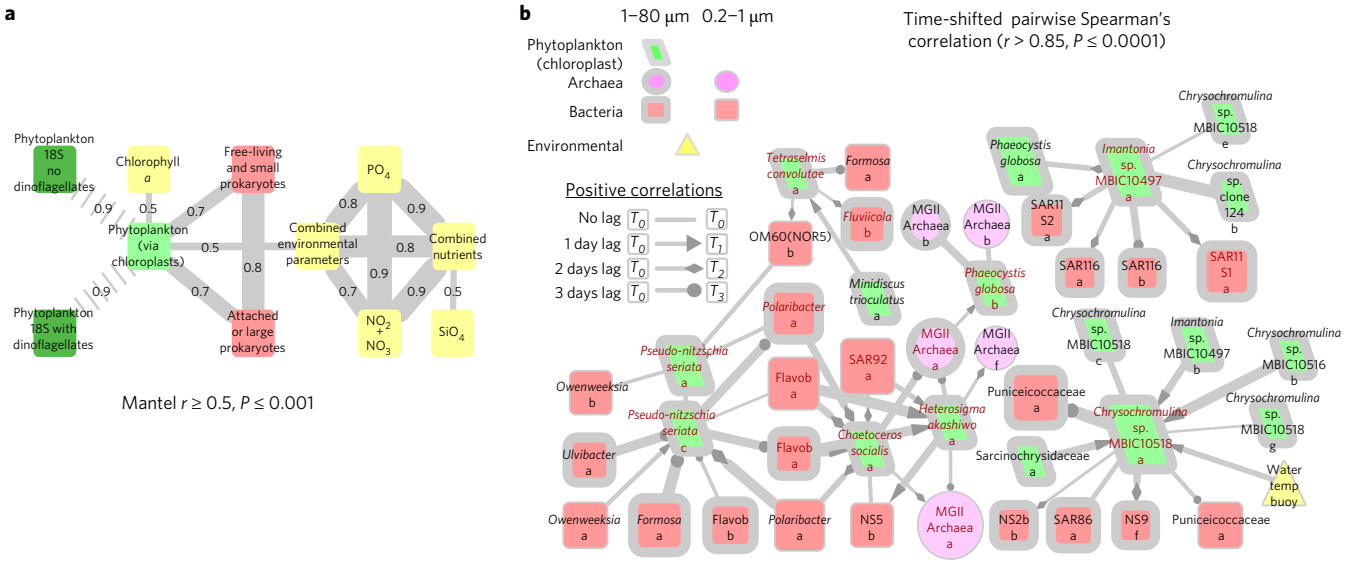

Figure 3 | Correlations among biotic and abiotic parameters indicate stronger relationships between populations of microorganisms than to other bulk environmental parameters. a, Mantel tests show biological communities, consisting of all eukaryotic phytoplankton and small or large prokaryotic size fractions, are more correlated to one another than to environmental parameters. For simplicity, correlations of 18S other than to chloroplasts (dashed edge) are not shown. 'Combined environmental parameters' includes a combination of those shown in yellow + wave height. **b**, Microbial association network shows that pairwise time-shifted Spearman's correlations to the dominant phytoplankton during the bloom (from Fig. 1a) are often to bacteria, with time lags. Taxa top ranked in abundance in either size fraction have red labels. In **a** and **b**, the width of the lines connecting parameters represents correlation strength, and in **b**, the size of the nodes represents the average OTU abundance.

smaller, non-mineralized taxa, for example, *Ostreococcus*, *Phaeocystis* and *Chrysochromulina*, result in reduced export[22].

Like the phytoplankton, the prokaryotic community composition was also highly dynamic during and after the bloom (Fig. 2). During the first six days of sampling, sequences from a Verrucomicrobium taxon, three Flavobacteria taxa and a SAR92 taxon became most abundant for at least one day in either the large and particle-attached or free-living fractions (Fig. 2). This shows higher fine-scale variation of bloom-associated taxa than previously documented[11]. Additionally, although archaea are not traditionally thought to be rapidly growing organisms, we observed a sharp increase in Marine Group II (MGII) euryarchaeal taxa around 23 March, with a single taxon reaching an estimated relative abundance of 30% of the prokaryotic community, and the top four MGII archaeal taxa making up about 40%, cumulatively (Fig. 2 and Supplementary Fig. 7). This suggested the archaea were briefly the numerically dominant organism overall at that time and location. As the level of chlorophyll increased from 24 March, the larger and particle-attached prokaryotes were dominated initially by the same Verrucomicrobium taxon as the primary bloom. Following this, SAR11 members, along with cyanobacteria *Synechococcus* and *Prochlorococcus*, generally dominated the prokaryotic abundance through the summer, and the taxa dominant during the primary bloom were rare (Fig. 2b and Supplementary Fig. 7).

Some of the shifts in community structure could have been due to spatial variation, including currents moving patches with different dominant taxa past the sample location, vertical or lateral mixing[25], and/or microscale patchiness and aggregation. Estimates of apparent net growth rates from our data suggest that most observed increases were within expected potential growth rates of phytoplankton and prokaryotes (up to a few doublings per day). However, some individual taxa on particular days increased more rapidly than would be expected from anticipated growth rates, therefore suggesting spatial variability (Supplementary Fig. 8). This confounds efforts to estimate individual growth rates. Nevertheless, we have observed that, at a single geographic location, six distinct phytoplankton taxa can become the apparent dominant alga over only six days during bloom conditions, along with five different bacteria. A study covering part of our time series at the same location (in May) reported that the protistan community was more variable over short time periods (days) than over short distances (km)[26], suggesting that temporal variation dominated. Phytoplankton in March were particularly dynamic, with the rate of community change being three times as fast as in May (Supplementary Fig. 9).

To determine what factors best explained the observed variation we used partial Mantel tests, factoring out time as a variable, which showed that biological communities were more highly correlated to one another than to environmental parameters (Fig. 3a and Supplementary Table 2). Local similarity analysis (LSA)-based microbial association networks[24] looked for potential microbial interactions (for example, mutualism, cross-feeding, competition, parasitism, predation and allelopathy) among particular taxa and for correlations with environmental parameters. This indicated several time-delayed correlations (time-shifted Spearman's $r > 0.85$) between the dominant phytoplankton taxa in March and specific prokaryotes. There were fewer correlations between phytoplankton and environmental parameters, with none to nutrients (Fig. 3b). The nitrogen-fixing symbiotic cyanobacterium UCYN-A was well correlated to three different protists, including its known host[27] (Supplementary Table 3), suggesting possible additional hosts. There were several correlations between specific diatoms and Flavobacteria (Fig. 3b). Dominant MGII archaea were correlated contemporaneously to *Phaeocystis* and with delay to *Chaetoceros* and *Heterosigma* (Fig. 3b). Although the onset of spring phytoplankton blooms appears to be driven initially by physical conditions and inorganic nutrients, we interpret the stronger relationships seen among communities afterwards (despite new nutrient additions) as indicating that biologically mediated phenomena strongly influence the structure of the entire microbial community, including phytoplankton, during and after bloom initiation.

It is too early to characterize the exact nature of the microbial associations our analysis indicated. Microbial interactions result

from direct physical associations or are mediated by substrates released from phytoplankton and prokaryotes during growth (including vitamins), grazing, parasitism, infection and allelopathy[5–7,23,28,29]. Also, many 'phytoplankton' are mixotrophic, potentially feeding on other microorganisms[28]. Although there is no doubt validity in the traditional view that nutrients and physics are critical in getting blooms started and that nutrient exhaustion, sinking and grazing (also parasites and viruses) are important factors leading to declines of bloom biomass, our results suggest that a multitude of microbial interactions are probably important in determining the exact types of phytoplankton that dominate at a given time and location. The extent that particular interactions contribute to bloom and post-bloom dynamics remains to be explored and will enhance our understanding of how nutrients and energy are transformed during these common, but ephemeral events.

Our results augment the traditional view of phytoplankton blooms by examining the microbial communities with higher phylogenetic and temporal resolution than has usually been the case in the past. We find that species dominance in blooms and their aftermath can be highly fluid and mercurial, helping to explain the high diversity of marine plankton and that high-resolution sampling of community composition is important for understanding the mechanisms of blooms and their overall impacts. The many intercorrelations among organisms, yet relatively few correlations to inorganic nutrient concentrations, are consistent with the hypothesis that positive and negative interactions among microorganisms are major drivers of the changing composition and productivity of communities, exerting influence on coexistence, diversity, biogeochemistry and global biogeographic patterns[30].

Methods

Location and sampling. For each sampling point, a single water sample (10–20 l) was collected from the top 1 m of the water column at the San Pedro Ocean Time-series location (SPOT: latitude 33°33′ N, longitude 118°24′ W) by multiple bucket casts generally around 08:15 local time from the USC commuter boat, *Miss Christie*. Water samples were stored in a cooler until arrival at the Wrigley Institute of Environmental Science, Catalina Island, around 8:45, where processing began immediately. When not logistically ideal to collect on the morning crossing (23 March, 29 March, 5 May), water was collected in the afternoon and filtered at the University of Southern California in Los Angeles within 2–3 h. Samples were occasionally collected via the SPOT monthly sampling cruise (15 March, 29 April, 24 May, 22 June, 20 July, 3 August), sampled around 12:00 and filtered by 18:00. For the collection of cellular material, seawater was sequentially filtered through an 80 µm mesh, 47 mm Type A/E glass fibre filter (~1.0 µm pore) and 47 mm Durapore (polyvinylidene fluoride, PVDF, Millipore). The AE and Durapore filters were stored at −80 °C until extraction.

Environmental measurements. Temperature, salinity and pH were determined immediately upon sampling in an on-board bucket by YSI 63 instrument. Chlorophyll *a* concentrations were determined by triplicate 0.5–1 l filtrations with a GF/F filter, frozen at −80 °C, and extracted with 90% acetone and analysed fluorometrically. Bacterial and viral abundances were determined via SYBR green epifluorescent microscopy[31]. However, many slides were not of high quality due to being made on 13 mm anodiscs, because of a temporary manufacturer shortage of 25 mm filters. Data from low-quality microscopy slides are not reported. Satellite sea-surface chlorophyll *a* imagery for the southern Californian region was downloaded from National Oceanic and Atmospheric Administration Coast Watch using eight-day averages. Daily averages for precipitation, air temperature and wind speed were downloaded from http://www.wunderground.com/weather-forecast/US/CA/Catalina.html. Wave height and water temperature data were collected from the National Buoy Data Center from a buoy located in the San Pedro Channel at 33°37′ N 118° 19′ W (SCRIPPS Station 4622, http://www.ndbc.noaa.gov/station_page.php?station=46222), which is approximately 11 km from our sampling location.

DNA extraction. DNA was extracted from a single Durapore filter with an SDS and phenol:chloroform extraction[32,33] for each sample date. DNA was extracted from the AE filter with a NaCl/cetyl trimethylammonium bromide (CTAB) extraction[34]. Briefly, lysis was performed with 0.7 M NaCl, 1% CTAB and 0.5 mm zirconia/silica beads. The lysis solution (with beads) was vortexed for 1 min and heated to 70 °C for 5 min (repeated three times). DNA was purified from lysates with a 2× volume of phenol, 1:1 volumes of phenol and chloroform:isoamyl alcohol and a 2× volume (24:1) of chloroform isoamyl alcohol. The DNA was precipitated with a 2.2× volume of 95% ethanol and 0.25× of 10.5 M ammonium acetate, and left overnight at −20 °C. Precipitate was collected via centrifugation and resuspended in tris-EDTA (TE) and stored at −80 °C.

PCR, barcoding and sequencing preparation. Salt-free purified primers were ordered purified from Eurofins Operon. The forward primer construct consisted of (5′ to 3′) a 'generic' Illumina flow cell adapter, Illumina sequencing primer, four random bases, a five base barcode and the 16S forward primer 515F (GTGCCAGCMGCCGCGGTAA). The reverse primer construct consisted of (5′ to 3′) a 'generic' Illumina flow cell adapter, a six base index, Illumina sequencing primer and the 16S reverse primer 926R (CCGYCAATTYMTTTRAGTTT). Triplicate 25 µl polymerase chain reactions (PCRs) of single DNA extracts diluted to 1–2 ng μl^{-1} were performed as follows: 1× Invitrogen Platinum Taq HiFi Buffer (catalogue no. 11304–029), 0.2 mM dNTPs (Promega catalogue no. U1515), 0.4 µM of each primer, 2 mM $MgSO_4$ (Invitrogen, 11304–029) and 1U Invitrogen Platinum HiFi Taq (11304–029). Thermocycling conditions consisted of an initial denaturation of 95 °C for 120 s; 25 cycles of 95 °C for 45 s, 50 °C for 45 s, 68 °C for 90 s; and a final elongation step of 68 °C for 300 s. Triplicates were re-combined and were run on agarose gel to confirm expected amplification or non-amplification (for no template controls). PCR reactions were then clean and concentrated with 1× (vol:vol) Ampure XP magnetic beads. Purified products from samples were pooled in equimolar concentrations. All sequences were sent to the University of California Davis Genome Center and sequenced with generic sequencing primers on an Illumina Miseq. The 0.22–1 µm and 1.0–80 µm size fractions were analysed with 2 × 260 and 2 × 300 sequencing, respectively. Technical replicates, that is, replicated triplicate reactions from the same DNA extract, were PCR amplified and barcoded discretely for some sample dates from the 0.22 µm size fraction (17 March, 18 March, 22 March, 23 March, 24 March, 21 May and 1 July). Relative abundance values for these replicates were averaged before reporting in the manuscript.

Sequence analysis. Raw data, processed data, mapping files and all commands and scripts employed in sequence/data analysis and figure generation are available via Public Databases and/or FigShare (Supplementary Table 4).

For 16S sequence analysis, paired-end reads were merged using usearch7 *fastq_mergepairs*[35], where reads were truncated at the first base with a *q*-score below 5 and a 10% mismatch was allowed over the expected region of overlap (about 100–200 bp depending on the data set). Merged reads shorter than 336 or longer than 486 were removed. QIIME script *split_libraries_fastq.py*[36] was then used to de-index the merged sequence files allowing no mismatches to the index. The de-indexed sequences were then fully demultiplexed via the five-base inline barcode and filtered, with no mismatches to the in-line barcode allowed; sequences with an average *q*-score of <25 were removed. Chimeric sequences were identified and removed by both *de novo* and a reference database enabled chimera-checking using *identify_chimeric_seqs.py*. Sequences were *de novo* clustered at 99% ID by UCLUST via *pick_otus.py* and the most abundant sequence of each operational taxonomic unit (OTU) was picked for classification via *pick_rep_set.py*. Taxonomy was assigned to all sequences with UCLUST by searching against both SILVA (release 111)[37] and greengenes (release 13_8)[38] database using *assign_taxnomy.py* default settings. In general, we used the SILVA classifications for bacteria and archaea throughout the manuscript, but both greengenes and SILVA are available in Supplementary Data Set 1. Additionally, the sequences were classified via a BLASTn search of the non-redundant nucleotide NCBI database (downloaded July 2014), excluding sequences from environmental samples or metagenomes; the best BLASTn match was used as primary classification when the SILVA taxonomy was 'Unassigned', which was rare (Supplementary Data Set 1). 16S sequences classified as mitochondria or chloroplast by greengenes were removed from the bacterial/archaeal OTU table.

Due to inconsistencies in the SAR11 taxonomy between the SILVA database and a recent report (Vergin *et al.* 2013)[30], to avoid confusion we included the Vergin *et al.* taxonomy in addition to the SILVA taxonomy in Fig. 2 and Supplementary Fig. 7. We determined the 'Vergin' subclades of the SAR11 representative sequences from Fig. 2 by aligning and phylogenetic reconstruction of the V4/V5 region of 65 SAR11 sequences reported in Supplementary Fig. 1 of the Vergin *et al.* article. We generated the alignment using the MAFFT algorithm[39] G-INS-I defaults within Geneious version 6.1.6. Maximum likelihood phylogenetic trees were estimated using PHYML[40] with the HKY85 Substitution model and 100 bootstraps. We confirmed similar topology between their tree and ours and determined the subclade of the sequences by visual inspection of the phylogenetic tree (see Supplementary Table 4 for links to alignment and the phylogenetic tree).

Classification of chloroplast sequences. A separate OTU table was generated for OTUs classified as chloroplast by the greengenes classification. Because the taxonomic assignments of chloroplasts via SILVA and greengenes were generally restricted to the Phylum or Class level, we used the BLASTn NCBI classification as the primary way to classify the chloroplast sequences. Additionally, we classified the chloroplast sequences by BLASTn search of the PhytoRef database (downloaded July 2015)[15]. Both classifications are reported for each OTU (Supplementary Data Set 1), together with the percent similarity of matches. Because the exact taxonomic identity often varied slightly between the two databases, we established rules for

determining which classification to use in the body of this manuscript and figures. If a chloroplast sequence had a better match to the PhytoRef database and the PhytoRef match included Genus/Species, we used the PhytoRef classification. If the NCBI match was <95% and the PhytoRef was >95%, we used the PhytoRef classification. Because the NCBI taxonomy often classified the host of kleptoplasts as the chloroplast OTU identity, we used PhytoRef; that is, we used Phytoref for those OTUs classified by the NCBI database as *Phalacroma* or *Virgulinella* (known kleptoplastic organisms).

To examine the classification of the dominant chloroplast OTU sequences (Fig. 1), we assessed their phylogenetic accuracy by aligning all of the representative sequence to the database sequence from which the OTU was classified. Then, a maximum likelihood phylogenetic tree was generated using PHYML[40] with the HKY85 Substitution model and 100 bootstraps (Supplementary Fig. 10). Before alignment, one OTU was removed because the NCBI sequence did not cover the full 16S region of interest; the OTU was a 100% match over 335 base pairs to Pyramimonas, gi: 658131508.

For 18S analysis, the raw, demultiplexed sequences (EMBL Study accession PRJEB10834) were quality filtered with Trimmomatic (LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:225). Forward reads were then trimmed to 290 bases and the reverses were trimmed to 225 bases due to variation in the quality of each read. Paired ends were removed if both ends did not exceed these thresholds. Because 18S fragment size for this primer set is typically 575–595 bp, we were able to separate 18S from 16S by collecting the unmerged sequences with USEARCH8, *fastq_mergepairs*. To use both ends of the read, we converted the reverse read to its reverse complement and pasted the sequences together with a degenerate base 'N' separating the two reads (which does not significantly influence k-mer based classification such as RDP[41]). We then established 99% OTUs with QIIME implementation of the UCLUST algorithm, using *pick_otus.py*. The most abundant sequence from each OTU was assigned taxonomy in QIIME with the kmer-based RDP method against the Protistan Ribosomal Reference database (PR2)[42] and by BLASTn search of the same non-redundant nucleotide NCBI database as used for 16S classifications. We report both classifications (Supplementary Data Set 1), but default to the PR2 classification for primary reporting in the manuscript unless (1) there was no PR2 classification beyond Family and the NCBI classification was >95% ID, (2) if the PR2 was classified as 'Radial-centric-basal-Coscinodiscophyceae' or 'Raphid-pennate', or (3) if PR2 classification was <95% ID and NCBI was >95% ID. When the NCBI classification reported the best match to a host taxon of a symbiotic organism, not the original organism, we used the best hit that did not include 'symbiont' in the classification; for example, *Chrysochromulina* sequences were originally found to be classified as Thraustochytrid symbionts. Additionally, we used the second best hit for taxa classified as 'prasinophyte symbionts of a radiolarian' because the second best hit was *Tetraselmis* sp. (prasinophyte) at very high similarity, often >99% similarity.

To enable a direct comparison between the chloroplast 16S and 18S analyses, we first selected only those lineages from the 18S that corresponded to those found in the chloroplast 16S data and generated an OTU table of just these lineages: Chlorophta, Archaeaplastida, Stramenopiles (excluding MAST) and Haptophytes. In an effort to broaden the 18S analysis to other photosynthetic organisms, while excluding heterotrophs, we generated an OTU table of the preceding groups, plus dinoflagellates that were classified to species levels that are autotrophic or mixotrophic, not temporarily kleptoplastic alone (determined via literature search, Supplementary Table 5). The photosynthetic dinoflagellate species that were most common in our data were *Alexandrium fundyense*, *Blastodinium contortum*, *Cochlodinium fulvescens*, *Gonyaulax polygramma*, *Gonyaulax spinifera*, *Heterocapsa pygmea*, *Lingulodinium polyedrum*, *Neoceratium furca*, *Protoperidinium denticulatum* and *Karlodinium micrum*.

Statistical analyses. Heatmaps were generated via the heatmap3 package in R (3.1.1). The OTU representative sequences were aligned in Geneious (v6.1.6)[43] using the MAFFT algorithm[39] G-INS-I defaults. Maximum likelihood phylogenetic trees were estimated using PHYML[40] with the HKY85 Substitution model and 100 bootstraps.

Calculation of apparent net growth rates. We calculated the apparent net growth rates for the dominant OTUs (those present in Figs 1 and 2) for all pairs of adjacent days in the chloroplast 16S and bacteria and archaeal 16S data where an OTU increased in apparent abundance (relative abundance) from one day to the next. We used the equation $\ln(N_{o+1}/N_o)/\ln(2)$ to determine the number of replications necessary to account for increases in abundances between adjacent days, where N_o and N_{o+1} are the estimated abundances on the initial and the following (adjacent) dates. For the phytoplankton analysis we calculated the apparent net growth rates for both the relative abundances as in Fig. 1 as well as the 'estimated abundance' by multiplying the relative abundances by the chlorophyll concentration observed. For the bacteria and archaea, we only calculated apparent net growth rates for the relative abundances.

Community and environmental similarity. For the community decay in similarity and Mantel tests, the Bray-Curtis similarity was derived via the *vegdist* function in Vegan(2.2–1)[44]. For dates and environmental parameters (*z*-score-transformed), Euclidean distance matrices were calculated via the R base *dist* function. For preparation for Mantel tests, communities and environmental data with missing data were removed to generate a fully overlapping data set (36 samples total). Partial Mantel tests were performed for all combinations of the communities and isolated and combined environmental variables in Vegan using *mantel.partial*, factoring out the date's distance matrix.

LSA. Pairwise correlation matrices were generated using eLSA[45,46] for the OTUs and environmental parameters for both the March time series and the full time series (performed separately). The top 60 OTUs from >1 µm prokaryotic, 0.22–1 µm size fraction, >1 µm chloroplasts and 18S photosynthetic organisms were included that were present in over 25% of samples, corresponding to taxa that were no less than 0.1% on average.

For the March daily time series dates for which we were not able to sample, we inserted a time-series step to enable detection of delayed correlations, by linearly interpolating the missing data from adjacent days. For the March time series, time-lagged correlations were examined with a maximal delay of three days. For the full time series, we did not enable detection of time-lagged correlations because the data were not evenly spaced. We used eLSA's 'mixed' option to determine *P* values; this only performs permutations ($n = 1{,}000$) when *P* values are explicitly determined to be <0.05 (refs 45,46). Values of *q* were calculated to control for false positives[47]. Correlation matrices were imported into Cytoscape (v2.8.3)[48] together with an attribute table of taxonomy and abundance information for mapping onto nodes (Supplementary Data Set 1).

Accession numbers. Demultiplexed sequence files are available at EMBL under study accession no. PRJEB10834. Raw sequence data, QIIME mapping files, all commands used in data analysis and figure generation, processed data sequence files and environmental data are available via FigShare (https://figshare.com/s/ba6029898908bf7b80d5; see Supplementary Table 4 for links and descriptions).

Received 18 October 2015; accepted 15 January 2016; published 29 February 2016

References

1. Ducklow, H. W., Steinberg, D. K. & Buesseler, K. O. Upper ocean carbon export and the biological pump. *Oceanography* **14,** 50–58 (2001).
2. Kirchman, D. L. *Processes in Microbial Ecology* (Oxford Univ. Press, 2012).
3. Hutchins, D. A., Mulholland, M. R. & Fu, F. Nutrient cycles and marine microbes in a CO_2-enriched ocean. *Oceanography* **22,** 128–145 (2009).
4. Miller, C. B. *Biological Oceanography* (Blackwell, 2004).
5. Amin, S. A., Parker, M. S. & Armbrust, E. V. Interactions between diatoms and bacteria. *Microbiol. Mol. Biol. Rev.* **76,** 667–684 (2012).
6. Amin, S. A. *et al.* Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522,** 98–101 (2015).
7. Buchan, A., LeCleir, G. R., Gulvik, C. A. & González, J. M. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nature Rev. Microbiol.* **12,** 686–698 (2014).
8. Ducklow, H. W., Kirchman, D. L., Quinby, H. L., Carlson, C. A. & Dam, H. G. Stocks and dynamics of bacterioplankton carbon during the spring bloom in the eastern North Atlantic Ocean. *Deep-Sea. Res. II* **40,** 245–263 (1993).
9. Mayali, X., Palenik, B. & Burton, R. S. Dynamics of marine bacterial and phytoplankton populations using multiplex liquid bead. *Environ. Microbiol.* **12,** 975–989 (2010).
10. Lindh, M. V. *et al.* Disentangling seasonal bacterioplankton population dynamics by high-frequency sampling. *Environ. Microbiol.* **17,** 2459–2476 (2015).
11. Teeling, H. *et al.* Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336,** 608–611 (2012).
12. Orsi, W. D. *et al.* Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. *ISME J.* **9,** 1747–1763 (2015).
13. Mincer, T. J. *et al.* Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. *Environ. Microbiol.* **9,** 1162–1175 (2007).
14. Parada, A., Needham, D. M. & Fuhrman, J. A. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time-series and global field samples. *Environ. Microbiol.* http://dx.doi.org/10.1111/1462-2920.13023 (2015).
15. Decelle, J. *et al.* PhytoREF: a reference database of the plastidial 16S rRNA gene of photosynthetic eukaryotes with curated taxonomy. *Mol. Ecol. Resour.* **15,** 1435–1445 (2015).
16. Mann, D. G. in *Cytology, Genetics and Molecular Biology of Algae* (eds Chaudhary, B. R. & Agrawal, S. B.) 249–274 (SPB Academic, 1996).
17. Zhu, F., Massana, R., Not, F., Marie, D. & Vaulot, D. Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol. Ecol.* **52,** 79–92 (2005).
18. Koumandou, V. L., Nisbet, R. E. R., Barbrook, A. C. & Howe, C. J. Dinoflagellate chloroplasts—where have all the genes gone? *Trends Genet.* **20,** 261–267 (2004).

19. De Vargas, C. *et al.* Eukaryotic plankton diversity in the sunlit ocean. *Science* **348,** 1261605 (2015).
20. Rossini, G. P. *Toxins and Biologically Active Compounds from Microalgae: Biological Effects and Risk Management* Vol. 2 (CRC Press, 2014).
21. Schoemann, V., Becquevort, S., Stefels, J., Rousseau, V. & Lancelot, C. Phaeocystis blooms in the global ocean and their controlling mechanisms: a review. *J. Sea Res.* **53,** 43–66 (2005).
22. Bienfang, P. K. Sinking rates of heterogeneous, temperate phytoplankton populations. *J. Plankton Res.* **3,** 235–253 (1981).
23. Granéli, E., Salomon, P. S. & Fistarol, G. O. in *Algal Toxins: Nature, Occurrence, Effect and Detection* 159–178 (Springer, 2008).
24. Klaas, C. & Archer, D. E. Association of sinking organic matter with various types of mineral ballast in the deep sea: implications for the rain ratio. *Glob. Biogeochem. Cycles* **16,** 1–14 (2002).
25. Fuhrman, J. A., Cram, J. A. & Needham, D. M. Marine microbial community dynamics and their ecological interpretation. *Nature Rev. Microbiol.* **13,** 133–146 (2015).
26. Lie, A. A. Y., Kim, D. Y., Schnetzer, A. & Caron, D. A. Small-scale temporal and spatial variations in protistan community composition at the San Pedro Ocean time-series station off the coast of southern California. *Aquat. Microb. Ecol.* **70,** 93–110 (2013).
27. Thompson, A. W. *et al.* Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science* **337,** 1546–1550 (2012).
28. Worden, A. Z. *et al.* Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* **347,** 1257594 (2015).
29. Lima-Mendez, G. *et al.* Ocean plankton. Determinants of community structure in the global plankton interactome. *Science* **348,** 1262073 (2015).
30. Vergin, K. L. *et al.* High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. *ISME J.* **7,** 1322–1332 (2013).
31. Noble, R. T. & Fuhrman, J. A. Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquat. Microb. Ecol.* **14,** 113–118 (1998).
32. Needham, D. M. *et al.* Short-term observations of marine bacterial and viral communities: patterns, connections and resilience. *ISME J.* **7,** 1274–1285 (2013).
33. Fuhrman, J. A., Comeau, D. E., Hagstrom, A. & Chan, A. M. Extraction from natural planktonic microorganisms of DNA suitable for molecular biological studies. *Appl. Environ. Microbiol.* **54,** 1426–1429 (1988).
34. Countway, P. D., Gast, R. J., Savai, P. & Caron, D. A. Protistan diversity estimates based on 18S rDNA from seawater incubations in the western North Atlantic. *J. Eukaryot. Microbiol.* **52,** 95–106 (2005).
35. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* **10,** 996–998 (2013).
36. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7,** 335–336 (2010).
37. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41,** D590–D596 (2013).
38. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72,** 5069–5072 (2006).
39. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30,** 3059–3066 (2002).
40. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704 (2003).
41. Jeraldo, P. *et al.* IM-TORNADO: a tool for comparison of 16S reads from paired-end libraries. *PLoS ONE* **9,** e114804 (2014).
42. Guillou, L. *et al.* The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41,** D597–D604 (2013).
43. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–1649 (2012).
44. Oksanen, A. J. *et al. vegan: Community Ecology Package* (2015); http://vegan.r-forge.r-project.org/.
45. Xia, L. C. *et al.* Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates. *BMC Syst. Biol.* **5,** S15 (2011).
46. Xia, L. C., Ai, D., Cram, J. A., Fuhrman, J. A. & Sun, F. Efficient statistical significance approximation for local similarity analysis of high-throughput time series data. *Bioinformatics* **29,** 230–237 (2013).
47. Storey, J. D. The positive false discovery rate: a Bayesian interpretation and the *q*-value. *Ann. Stat.* **31,** 2013–2035 (2003).
48. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13,** 2498–2504 (2003).

Acknowledgements

The authors acknowledge the USC Wrigley Institute of Environmental Science and thank R. Marinelli, S. Conner, G. Boivin and the crew of the *Miss Christie* for sampling opportunities and laboratory space, and J. Chang, C. Chow, A. Lie, S. McCallister and E. Teel for sampling assistance. The authors thank D. Caron, F. Corsetti, J. Heidelberg, E. Webb, N. Ahlgren, L. Berdjeb, J. Cram, D. Comeau, L. Gómez Consarnau, E. Fichot, M. Lee, A. Parada, B. Phillips, G. Ramirez and E. Sieradzki for discussion and feedback on the manuscript, along with R. Sachdeva who also provided computer support. This work was supported by NSF grants 1031743 and 1136818, grant GBMF3779 from the Gordon and Betty Moore Foundation Marine Microbiology Initiative and a National Science Foundation Graduate Student Research Fellowship to D.M.N.

Author contributions

D.M.N. and J.A.F. designed the study, performed the research, analysed the data and wrote the paper.

Additional information

Supplementary information is available online. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.A.F.

Competing interests

The authors declare no competing financial interests.