Simultaneous quantification of active carbon and nitrogen fixing communities and estimation of rates using fluorescence in situ hybridization and flow cytometry

Allison S. McInnes1*, Alicia K. Shepard1, Eric J. Raes2, Anya M. Waite2†, and Antonietta Quigg1,3

1Department of Oceanography, Texas A&M University, College Station, Texas, USA
2The Oceans Institute, University of Western Australia, Crawley, WA, Australia
3Department of Marine Biology, Texas A&M University at Galveston, Galveston, Texas, USA

Running Head: Simultaneous quantification of C and N fixation

# Address correspondence to Allison S. McInnes: Email: allison.mcinnes@uwa.edu.au;
*Present address: Allison S. McInnes The Oceans Institute, University of Western Australia, Crawley, WA, Australia
†Present address: Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Am Handelshafen 12, 27570 Bremerhaven, Germany
Abstract

Understanding the inter-connectivity of oceanic carbon and nitrogen cycles, specifically carbon and nitrogen fixation, is essential in elucidating the fate and distribution of carbon in the ocean. Traditional techniques measure either organism abundance or biochemical rates. As such, measurements are performed on separate samples and at different time scales. Here we developed a method to simultaneously quantify organisms while estimating rates of fixation across time and space for both carbon and nitrogen.

Tyramide signal amplification (TSA) fluorescence in situ hybridization (FISH) of mRNA for functionally specific oligonucleotide probes *rbcL* (RuBisCO, carbon fixation) and *nifH* (nitrogenase, nitrogen fixation) was combined with flow cytometry to measure abundance and estimate activity. Cultured samples representing a diversity of phytoplankton (cyanobacteria, coccolithophores, chlorophytes, diatoms and dinoflagellates), as well as environmental samples from the open ocean (Gulf of Mexico, USA and south eastern Indian Ocean, Australia) and an estuary (Galveston Bay, Texas, USA) were successfully hybridized.

Strong correlations between positively tagged community abundance and $^{14}$C/$^{15}$N measurements are presented. We propose that these methods can be used to estimate carbon and nitrogen fixation in environmental communities. Utilization of mRNA TSA-FISH to detect multiple active microbial functions within the same sample will offer increased understanding of important biogeochemical cycles in the ocean.
Introduction

Carbon (C) is the biogeochemical currency of the ocean; its importance exemplified in measuring either carbon dioxide (CO$_2$) sequestration (about half of anthropogenic CO$_2$ is sequestered by the oceans; 1, 53) or potential fisheries yields (average catch of 68 Mt yr$^{-1}$ equating to primary production of 3.5 Gt C yr$^{-1}$; 6). Primary production converts atmospheric CO$_2$ into organic carbon, the critical first step in introduction of C into the food web.

Determination of the fate of fixed C can be estimated by examination of the source of nitrogen (N) (11). In recent decades, it has become increasingly clear that the primary source of new N in the oligotrophic ocean is microbially-mediated biological N$_2$ fixation (14, 18, 19). In theory, at an ecosystem level, if accurate measurements of biological N$_2$ fixation can be made, C sequestration and food web production can also be estimated (see 12).

However, disparate techniques are used for quantification of biomass and rates. Biomass measurements for phototrophs include chlorophyll a, pigment analysis using high performance liquid chromatography (27) and flow cytometry (37, 39). Methods to measure rates of primary production include light/dark bottles (16, 61), $^{14}$C uptake (55), stable isotopes of oxygen ($^{16}$O, $^{18}$O and $^{17}$O – 2, 38) or fluorescence kinetics (32, 33). Similarly, for nitrogen fixation, organisms are enumerated by labeling via 4', 6-Diamidino-2-Phenylindole, Dihydrochloride and other nucleic acid stains, while determination of the rates of targeted processes, are executed via acetylene reduction reactions (22) or $^{15}$N$_2$ uptake (41, 42, 43).

Because quantification of different carbon- or nitrogen-fixing populations and estimates of their respective rates have used fundamentally different techniques, understanding of the connectivity between these two cycles is necessarily replete with assumptions. To improve our understanding, microbial biologists need not only to determine the composition of the microbial community (DNA) but to include process
measurements including activity (rRNA), and recently, specific activity (mRNA) within these populations. Fluorescence in situ hybridization (FISH) has been extensively employed to determine abundance and general activity (using rRNA probes) of prokaryotes in aquatic environments (30). Transcript abundance (qPCR of mRNA) of the carbon-fixing enzyme RuBisCO (9, 28, 47) and the nitrogen-fixing enzyme nitrogenase (40) correlate with their respective rate measurements. mRNA FISH is a tool capable of community quantification while simultaneously quantifying the number of organisms performing a specific function (50). Due to the low abundance of target sequences (mRNA << rRNA; 47) amplification of the signal using tyramide signal amplification (TSA) is necessary. Most studies have been limited to very specific groups (34, 35). This study targeted processes that span taxonomic and genetic variability. By simultaneously incorporating universal oligonucleotide probes complimentary to multiple functions actively expressed in the sample population, we can begin to understand the connectivity of carbon and nitrogen fixation via direct measurement.

The aim of this study was to develop an inclusive method employing TSA-FISH with general oligonucleotide probes complimentary to rbcL (codes for the large sub unit of RuBisCO, carbon fixation) and nifH (catalytic subunit of dinitrogenase reductase, nitrogen fixation) mRNA transcripts allowing quantification of C and N₂ fixing communities, and simultaneous estimation of C and N₂ fixation rates within those communities. Application of this method reduces the number of assumptions and allows a more direct understanding of the connectivity of these two dynamic cycles.

Materials and Methods
Probe design. In this study, our aim was to target C and N\textsubscript{2} fixation across the entire community, rather than tagging a unique species. This necessitated, firstly, finding a conserved target region in the gene sequences across a diversity of organisms, and secondly, allowing enough generality in hybridization optimization to tag all functional representatives. We addressed the first issue during probe design (Tables A1 & A2). We compiled 113/124 (rbcL/nifH) sequences including 7/4 Phyla representing 17/15 orders and 26/37 genera from GenBank (January 2010; http://www.ncbi.nlm.nih.gov/nucleotide/). Multiple sequences were selected for many species to include inter-species variability as well as between groups variability. Sequences were compiled covering all major taxa (cyanobacteria, diatom, dinoflagellates, haptophytes, and green algae among others for \textit{rbcL}; from the four major \textit{nifH} clusters described in 66) so that the probes would be as near to universal as possible at the time of design. These were aligned using MEGA5/ClustalW (gap opening penalty of 15, gap extension penalty of 6.6 [http://www.megasoftware.net/]). The most conserved region was chosen and further analyzed using OligoCalc (v3.26 [http://www.basic.northwestern.edu/biotools/OligoCalc.html]) to ensure no hairpin formation potential, 3’ complementarity, or self-annealing. Comparisons of these probe sequences were performed against the GenBank database to verify specificity to primary producers and nitrogen fixing organisms. NON338, a sequence commonly employed as a negative control, was used in this study (34, 58).

To eliminate background fluorescence from nonspecific binding of the amplification reagent used in visualization, horseradish peroxidase was directly attached (56) to the 5’ end of the oligonucleotide probes (reverse compliment of the target sequence, Tables A1 & A2) (Life Technologies Invitrogen custom oligonucleotide probes). Because we tagged both prokaryotes and eukaryotes, a traditional positive control was not used. Instead, we verified
target presence with SYBR Green I nucleic acid stain (1:10000, Molecular Probes). After
initial determination of accurate rbcL probe hybridization to Synechococcus sp. (data not
shown), this culture and probe was run as a positive control verifying chemical and probe
functionality.

**Maintenance and collection of laboratory cultures.** In order to ensure accuracy
while maintaining generality, rbcL probe specificity was determined using a diversity of
cultured phytoplankton, including: cyanobacteria (Synechococcus sp.), a dinoflagellate
(Thoracosphaera hemii), diatoms (Thalassiosira oceanica and Amphora coffeaeiformis), a
coccolithophore (Emiliana huxleyi), and a green algae (Dunaliella tertiolecta) (Table S3).
Cultures were grown at optimal temperatures according to Provasoli-Guillard National
Center for Culture of Marine Phytoplankton in f/2 media (20) on a 12:12 light:dark cycle at
19°C or 24°C and 130-150 μmol photons m⁻² s⁻¹. Media recipes on the culture collection
website were followed, with Gulf of Mexico seawater used as the base for the f/2 media
(20). The bacterium Escherichia coli, grown in LB broth (LB Broth Base Tablets 50 ml, TRU
MEASURE, Sigma Aldrich L7275-100TAB) at 37°C and was used as a negative control.
Samples were preserved with molecular grade paraformaldehyde (PFA, 4x final
concentration) overnight at 4°C (62). Cells were then collected via centrifugation (17000 × g,
5 min) and PFA was removed; cells were re-suspended in absolute ethanol and stored at -80
°C.

**Collection of environmental samples**

Environmental samples were collected to test hybridization of probes for both rbcL
and nifH. Samples were collected in the Gulf of Mexico (between 26-27°N, 86-92°W, June
2011) and the south eastern Indian Ocean (between 13-32°S, 100-125°W, Aug-Sept 2012)
where biological nitrogen fixation has been documented (Fig. S1; 10, 52). Samples were also
collected from Galveston Bay (Texas, USA; 29.34° N, 94.50° W) representing an environment where biological nitrogen fixation is unlikely (Fig. S1). Water samples were concentrated using gentle vacuum filtration (<2 mmHg) onto a Whatman Nucleopore polycarbonate track-etched membrane 0.2 µm filter (25 mm, noting volume for concentration calculations after analysis). Membrane filters were placed in a microcentrifuge tube with phosphate buffered saline (PBS, autoclaved and 0.2 µm filtered; 750 µl) and cells were vortexed off the filters so they could be analyzed using flow cytometry. Samples were preserved with molecular grade paraformaldehyde (PFA, 4x final concentration) overnight at 4°C. Cells were then collected via centrifugation (17,000 × g, 5 min) and PFA was removed; cells were re-suspended in absolute ethanol and stored at -80 °C.

Prior to collection of environmental samples, various protocols were tested to determine the optimum concentrating procedure. Duplicate water samples were collected from Galveston Bay and an initial chlorophyll a measurement was made using a Turner Instruments 10-AU fluorometer. Samples were then filtered (100 ml) onto Whatman Nucleopore polycarbonate track-etched membrane 0.2 µm filters (25 mm). Reverse filtration and vortexing the filter in PBS were tested for removing the samples from filters. Reverse filtration resulted in a 70-77% recovery while vortexing recovered between 95-99%. Vortexing in PBS was then used as the standard procedure for all environmental samples.

**Method comparison: 14C incorporation rates over diel cycles.** Carbon and nitrogen fixation rates vary on a diel cycle (3, 41). Laboratory cultures of *Synechococcus* sp., *T. hemii*, and *T. oceanica* were grown in triplicate batch cultures using the same media and growth conditions described above. Samples were harvested every 3 hours over a 24 h cycle and processed for 14C uptake using the small bottle method (36); additional aliquots were preserved for mRNA FISH at the same time. Samples were inoculated with 14C sodium
bicarbonate (final conc. 1 μCi ml⁻¹) for 20 min at 16 light intensities (0-1800 μmol photons m⁻² s⁻¹) at 24°C in a photosynthetron. Triplicate blanks and total counts (to determine specific activity) were prepared. Incubations were terminated with buffered formalin (50 μl) and all samples acidified for 24 h. Radioactivity was measured using a Beckman LS8100 scintillation counter.

To elucidate the importance of diel periodicity in C fixation, the median value for replicate samples from the TSA FISH analysis were interpolated (linear) over the 24 h cycle. These data were then integrated over various time periods (12 h light and entire 24 h) to compare with standard protocols (31). The amplitude (minimum subtracted from maximum) was calculated to demonstrate the intensity of the diel variation.

**Method comparison: N¹⁵ incorporation rates.** Environmental samples were solely used for optimization and method comparison for nifH probes because no laboratory samples were available. Samples for analysis of nifH mRNA expression versus¹⁵N₂ uptake were those collected in the eastern Indian Ocean during two cruises in August and September of 2012 (Fig. S1). Aliquots for mRNA FISH were collected every 3 hours, while ¹⁵N₂ uptake measurements were made at every other time point (every 6 hours) over a 24 h cycle. Direct addition of ¹⁵N₂ tracer-enriched seawater was used to estimate N₂ fixation rates (17, 42) The tracer was prepared by degassing and filtering (0.2 μm Sterivex filter) YBC-II media (7, 52) and stored in 3 L gas tight Tedlar bags which were then spiked with 0.8-1 mL ¹⁵N₂ (98 atom%; Aldrich) gas per 100 mL YBC-II media. Incubations were initiated by introducing ¹⁵N₂ tracer-enriched seawater aliquots of 2.6 % of the total incubation volume. Bottles were incubated for 6 hours. Experiments were stopped by collecting the suspended particles from each bottle by gentle vacuum filtration (pressure drop < 2 mmHg) through a 25 mm precombusted GF/F filter and snap frozen in liquid nitrogen and stored at -80°C.
Filters for total N and δ\(^{15}\)N isotope analysis were dried, acidified, and dried again overnight at 60°C. Determination of total N and δ\(^{15}\)N was carried out using a continuous flow system consisting of a SERCON 20-22 mass spectrometer connected with an Automated Nitrogen Carbon Analyzer (Sercon, UK). Samples for natural particulate organic nitrogen (PON) were obtained by gentle vacuum filtering of 4 L water samples onto precombusted GF/F filters. N\(_2\) fixation rates (in nmol N L\(^{-1}\) h\(^{-1}\)) were calculated following Dugdale and Goering (1967). Detailed explanation on fixation rate calculations can be found in the Protocols for the Joint Global Ocean Flux Study (31).

**Fluorescence in situ hybridization for flow cytometry.** The hybridization and detection protocol used is based on Pernthaler and Amann (2004), with some modifications (Table 1). Preserved laboratory and environmental samples were collected via centrifugation (17000 × g, 5 min) and ethanol was removed. Cells were re-suspended in 1x phosphate buffered saline (PBS; 0.14 M NaCl, 2.7 mM KCl, 10.1 mM Na\(_2\)HPO\(_4\), 1.8 mM KH\(_2\)PO\(_4\)), and centrifuged again to remove PBS (this process will hereafter be referred to as a wash/washed step). Endogenous peroxidase activity and RNAses were quenched/removed by incubation of samples in 0.1% diethylpyrocarbonate (DEPC) for 12 min at room temperature (RT, 50) followed by a wash to remove DEPC. Cells were permeabilized by incubation in 5 mg ml\(^{-1}\) lysozyme for 1 h at 35°C, followed by a wash in 1x PBS, and re-suspended in 1% sodium dodecyl sulfate (SDS) for 1 h at RT (49). Post permeabilization, the cells were incubated in 0.01% fresh H\(_2\)O\(_2\) for 10 min at RT to quench any newly exposed peroxidases.

Specificity of the oligonucleotide probes was optimized using various formamide concentrations (25) instead of various incubation temperatures because the horseradish peroxidase (HRP) attached to the 5’ end of the probes is unstable above 35°C (48). It should
be noted that the equation used to determine specificity of oligonucleotide probes, at its most stringent, allows for 20% mismatch (25). Cell suspensions were incubated at 35°C for 1 h in hybridization buffer, containing 1-150 µL formamide, 30 µL 20x sodium citrate buffer (SSC, Amresco), 60 µL 10% w/v dextran sulfate, 30 µL 10% w/v blocking solution (component D, Invitrogen TSA Kit #6, T-20916), 15 µL 4 mg mL⁻¹ yeast RNA, 6 µL 10 mg mL⁻¹ sheared salmon sperm DNA, and 9-159 µL autoclaved Milli Q water. Oligonucleotide probes were diluted in hybridization buffer and incubated at 35°C for 5 min, then added to cell suspensions (final concentration between 100-500 ng µL⁻¹) and incubated for 24 h at 35°C. Cell suspensions were washed with Milli Q, then with wash buffer (0.2x SSC, 0.01% w/v SDS) at 35°C for 30 min. Probes were detected using 1:100 Alexa 647 labeled tyramide (Invitrogen TSA Kit #6) in amplification buffer (0.1% blocking solution, 1% dextran sulfate 2M NaCl, 0.0015% fresh H₂O₂) for 30 min at RT. Cells were washed and re-suspended in PBS and counter stained with 30 mM tripotassium citrate and SYBR Green I nucleic acid stain (Invitrogen, Catalogue #S7563) 1:10000 then incubated for 15 min at 35°C.

**Flow cytometry.** Samples were enumerated using a Beckman Coulter Gallios flow cytometer equipped with 488 nm and 638 nm lasers. Fluorescence was measured using bandpass filters corresponding to emissions for SYBR (filter: 525/30) and Alexa 647 (filter: 695/30). Measurements for forward scatter (FS, roughly equivalent to size), and side scatter (SS, indicator of granularity) were also collected (26). The concentration of 15 µm beads (Coulter CC Size standard L15; 6602797) was determined using a haemocytometer. These were added to samples prior to flow cytometric analysis so that accurate sample volumes could be determined. Raw data was processed using Kaluza (V6). Cells were distinguished from other particulates using FS versus SYBR fluorescence plots. These events were then plotted using FS versus red fluorescence (Alexa 647). Gates were drawn on control plots to...
ensure that unhybridized cells were not counted. These gates were transferred to sample plots (with corresponding formamide concentrations). Positive results, if present, in the negative controls were subtracted from the corresponding samples.

Results

**Method Optimization.** Successful *in situ* hybridization of mRNA depends on optimization of a number of steps including permeabilization of cells and probe specificity, as well as minimization of background fluorescence and non-specific binding (Table 1). The first obstacle in probing a great diversity of organisms is the accompanying variety of cell walls in the target population, making multispecies permeabilization optimization difficult. Previous studies have used minimal treatments to avoid cell loss (Table 1). These treatments were attempted initially, but binding did not occur in our samples (data not shown). When permeabilization was increased (incubation in 5 mg mL⁻¹ lysozyme for 1 h at 35°C, washed, and incubated in 1% SDS for 1 h at RT) probes bound to targets. This permeabilization treatment was successful for all cultures and environmental samples, including phytoplankton with coccoliths (*E. huxleyi*), theca (*T. heimii*) or silicified frustules (*A. coffeaeformis, T. oceanica*) (Fig. 1). *T. hemii* showed the greatest visual loss of the pellet; nonetheless cells were successfully hybridized, indicating that the target cells themselves were not lost.

Minimizing background fluorescence is difficult when targeting photosynthetic organisms given the intrinsic fluorescence of the cells. Storage in absolute ethanol and subsequent permeabilization steps reduced autofluorescence. In addition, the sensitivity of the photomultiplier tubes in the flow cytometer were adjusted so that intrinsic fluorescence was accounted for with unhybridized controls. Quenching of endogenous peroxidase activity...
is also essential in reducing background fluorescence by limiting binding of tyramide-
fluorophore to non-target areas. Optimization of this step was essential as the chemicals
used can reduce target abundance, as well as impacting target integrity, and insufficient
quenching of peroxidase activity can lead to difficulties differentiating between positive
target and background fluorescence (Fig. S2). We therefore tested nine concentrations of
\( \text{H}_2\text{O}_2 \) (0-1% for 10 min at RT) on *Synechococcus* sp. and found maximum differentiation of
positive target to background fluorescence was achieved at 0.01% \( \text{H}_2\text{O}_2 \) (Fig. S2). This
concentration was then used for all subsequent measurements.

Separation of the microbial community from background fluorescence was
accomplished using SYBR Green I as a counter-stain. SYBR has a strong affinity to double
stranded DNA and binds to single stranded DNA and RNA at lower affinities (39) allowing
differentiation of cells to background particulate. SYBR-positive cells were selected and
subsequently analyzed for red fluorescence (Alexa 647). Red positive events are cells that
have been successfully hybridized with mRNA tags. Thus, the total microbial community
(SYBR positive cells) and the proportion showing positive target fluorescence (red positive
cells) could be quantified.

Because incubation temperature and probe concentration both affect hybridization,
we also tested a matrix of formamide concentrations and probe concentrations on five
different samples for *rbcl*: *Synechococcus* sp., *T. hemii*, *T. oceanica*, Gulf of Mexico (open
ocean), and Galveston Bay (estuary) (Fig. S3). Formamide concentrations of 25% yielded
maximum hybridization for all of the samples: *Synechococcus* sp. (75% of total community),
*T. hemii* (64% of total community), *T. oceanica* (66% of total community), Gulf of Mexico
(19% of total community) and Galveston Bay (31% of total community) (Fig. S3). The effect
of probe concentration varied between samples and appeared to be primarily a function of
cell concentration (data not shown). It is important to create a “probe driver situation”, which occurs when all target sequences are hybridized and excess probe is removed in subsequent washes (47). Hybridization specificity was purposefully less stringent here than previous studies (4, 34, 47) in order to be general enough to capture processes across taxonomically distinct groups. Therefore, the need for negative controls was paramount. *E. coli* (negative control) did not hybridize using the methods described in this study (<5%). Further, the cultures used for verification of the positive detection using the *rbcL* probe were monospecific, but not axenic; and the bacterial population present was not tagged (Fig. S5). Additionally, negative control probes were simultaneously applied to corresponding samples to ensure there was no non-specific binding of the oligonucleotide probes (<3%, dependent upon probe concentration) (Fig. S5).

Similar matrix combinations were run for *nifH* oligonucleotide probe on samples where biological nitrogen fixation should be occurring (Gulf of Mexico) and on samples where there should be no biological nitrogen fixation (Galveston Bay; Fig. S4). Successful detection of active biological nitrogen fixation was achieved using a general *nifH* oligonucleotide probe (Fig. S4) in samples from the Gulf of Mexico where nitrogen fixation has been documented (10) but not from Galveston Bay. Galveston Bay samples showed minimal hybridization (<4%) by comparison to the 10-22% hybridization in the open ocean Gulf of Mexico samples, indicating specificity of the *nifH* probe to *nifH* mRNA transcripts (Fig. S4). Hybridization was attempted using a negative control oligonucleotide probe on all of the same treatments (variation in formamide and probe concentration) with less than 10% positive results (data not shown). Optimal results (highest positive hybridization using the *nifH* probe with minimum hybridization using the negative probe) were achieved at a 35% formamide concentration.
To account for intrinsic fluorescence and non-specific binding of the fluorophore, control samples (with no probe added) were run for all formamide concentrations (first column in Fig. S5). Control samples showed very limited hybridization, indicating that removal of intrinsic fluorescence and quenching of endogenous peroxidase activity were effective (Fig. S5, column 1). Hybridization with the NON388 was limited (<5%, lower 12 panels in Fig. S5). Hybridization increased with increasing probe concentration, and decreased with increasing formamide concentration (increasing specificity, upper 12 panels in Fig. S5). The increase in fluorescence of positively hybridized targets is sufficient to easily differentiate these from background fluorescence.

**Detection of active primary producers using TSA FISH.** Cultured phytoplankton, representing a diversity of photosynthetic organisms were successfully tagged using a general rbcL oligonucleotide probe (Fig. 1); this technique was also shown to be successful in samples from both coastal and open ocean environments (Fig. S2). The cultures were chosen to represent the diverse evolutionary history of phytoplankton (Table S3), and were all small enough to be analyzed using flow cytometry (< 50 μm). The variation in the total number of target cells detected (35-70%, Fig. 1) may be due to differential peaks in transcription of target mRNA due to the time of day the cells were harvested (3-4 hours after the start of the light cycle). It is known that environmental diel expression of rbcL peaks between 2-6 hours after the start of the light cycle (51, 59, 60). All differences in positive versus negative probe samples were significant (t-test, p < 0.05; Fig. 1) except for *D. tertiolecta* (p=0.20, independent samples t-test) and *T. oceanica* (p=0.06, independent samples t-test). One of the triplicate samples of *D. tertiolecta* did not hybridize, yet the other two were significantly higher than the negative probe. *T. oceanica* showed a limited difference in positive versus negative hybridization. This may be due to
limited signal as a result of low number of target cells present. *E. coli* (negative control) showed no hybridization using the *rbcL* probe (Fig. 1), supporting accuracy in the specificity of the *rbcL* probe to transcripts of *rbcL* mRNA. The high degree of hybridization across a diversity of cultured organisms reveals that this technique may be successfully employed for a wide variety of samples.

Samples from estuarine (Galveston Bay) and open ocean waters (Gulf of Mexico and the south eastern Indian Ocean) were also used to test the effectiveness of the general *rbcL* probe to quantify active primary producers in environmental samples – all environmental samples were successfully tagged (Fig. S3). Of the total microbial plankton (0.2-20 µm), picoautotrophs represent ~24-40% in Galveston Bay, ~15-40% in the Gulf of Mexico (A. K. Shepard unpublished data), and 33% in the south eastern Indian Ocean (52). Maximum percentages of the microbial population measured expressing *rbcL* were similar to the average percentages of the community: ~38% of the microbial population collected in Galveston Bay, ~10% of the microbial population from the Gulf of Mexico, 10% (33% maximum) in the south eastern Indian Ocean (Fig. S3). This suggests conservation of the relative community composition in environmental samples.

**Detection of active nitrogen fixers using TSA FISH.** 10% of the microbial population sampled in the Gulf of Mexico hybridized using the *nifH* probe. Enumeration of *nifH* expression in the Indian Ocean, was 10,500 cells mL⁻¹ (19% of the total community) with an average of 600 cells mL⁻¹ (7% of the total community). Diversity and presence of *nifH* gene expression in environmental samples have been reported for environments around the world (8, 66). However, reported proportions of diazotrophic microbial community and abundance measurements are limited. Unicellular cyanobacteria expressing *nifH* genes can represent up to 10% of the microbial community composition (65), but this number does
not account for potential heterotrophic nitrogen fixers. Our method now makes the
generation of such numbers possible, allowing enumeration of all active cells in the
biological nitrogen fixing community to become routine

Expression and rates: Carbon fixation. From the photosynthesis versus irradiance
curves prepared using the cultures of *Synechococcus* sp., *T. hemii* and *T. oceanica* grown
over a diel cycle (Fig. 2), we calculated P\textsubscript{max}, the maximum \(^14\text{C}\) uptake at saturating photon
flux intensities and differential gene expression using TSA-FISH (Fig. 4). Highest P\textsubscript{max} in
*Synechococcus* sp. was observed after the peak in *rbcL* transcription (Fig. 2A). *T. heimii*
transcribed *rbcL* to a greater extent throughout the night compared to during the day (Fig.
2B). P\textsubscript{max} rates were 10 times lower during the day (~4 mg C m\(^{-3}\) h\(^{-1}\)) than highest rates at
night (~45 mg C m\(^{-3}\) h\(^{-1}\)); peaks during the dark period are not unusual (29). *T. oceanica* *rbcL*
expression and P\textsubscript{max} were tightly coupled throughout the light and dark cycles (Fig. 2C) with
maximum values measured during the day and lowest values measured during the night (9).
Greatest diatom P\textsubscript{max} (~220 mg C m\(^{-3}\) h\(^{-1}\)) was five-times higher than those measured in *T.
heimii* but significantly lower than in those measured in *Synechococcus* sp..

Marine primary producers range from prokaryotes to eukaryotes and span a size
range covering many orders of magnitude (30). In order for samples covering a diverse
range of cell types and sizes to be compared, the size normalized fluorescence (SNF) was
calculated for each of the cultures, removing fluorescent variability inherently associated
with cell size, using equation 1 below:

\[
\text{SNF} = \frac{\text{Fluorosample}}{\text{Fluorobeads}} \times \frac{\text{FSsample}}{\text{FSbeads}} \times \text{positive cells mL}^{-1}
\]  

Where Fluoro\textsubscript{sample} and Fluoro\textsubscript{beads} are the voltage values of red fluorescence for the
respective particles and FS is the forward scatter of the particles (indicative of size).
Multiple regression models were then used to determine correlation between \( rbcL \) expression (\# positive cells mL\(^{-1}\)) and \( P_{\text{max}} \) (maximum \(^{14}\)C uptake) plus and minus a 3, 6, 9, and 12 hour time shift. We found \( rbcL \) expression lead \( P_{\text{max}} \) across the diel light cycle by 6 hours (\( r = 0.823, p < 0.001; \) Figure 2); previous studies have shown strong correlations (6, 28, 40, 46) and that mRNA expression can lead \( P_{\text{max}} \) by 3-9 hours (29). For \( Synechococcus \) sp. there was maximum correlation between mRNA expression and \( P_{\text{max}} \) when \( P_{\text{max}} \) was shifted 6 h earlier (\( F = 3.139, p = 0.093 \) n = 18). Due to the consistency in cell size of \( Synechococcus \) sp., the size normalized fluorescence measure may confound the relationship between mRNA expression and photosynthesis. That is, when \(^{14}\)C uptake and SNF were compared, there was no relationship, but when \( rbcL \) positive cell concentration is compared to \( P_{\text{max}} \) (\( F = 4.911, p = 0.04, n = 18 \)) the relationship was significant. By contrast, regression models show maximum correlation between mRNA expression (SNF) and \( P_{\text{max}} \) when \( P_{\text{max}} \) is shifted 6 h earlier (\( F = 18.994, p < 0.001 \) n = 38) for \( T. \) heimii (Fig. 2B). \( T. \) oceanica regression models show maximum correlation between mRNA expression (SNF) and \( P_{\text{max}} \) when there is no shift (\( F = 2.195, p = 0.0147 \) n = 39).

Measurement of rates using traditional methods (\( P_{\text{max}} \)) and mRNA FISH showed significant correlation (Fig. 4), allowing determination of C fixation rates from size normalized fluorescence using the equation of the regression line, equation (2):

\[
P_f = (3 \times 10^{-8}) \times (\text{SNF/ul})^{0.82}
\]

The strength of the relationship (\( r = 0.823, p < 0.001; \) Fig. 4) shows that it is possible to not only enumerate the cells mL\(^{-1}\) actively transcribing \( rbcL \), accounting for differential expression between different components of the community (prokaryotes vs eukaryotes) but to also estimate C-fixation.
Comparison of results of $P_f$ (generated using TSA FISH) and C fixation rates previously reported for the south eastern Indian Ocean are very similar. Estimated C fixation rates ($P_f$) from samples collected in the eastern Indian Ocean range from 0.2-8 mgC m$^{-3}$d$^{-1}$ with a mean of 1.95 mgC m$^{-3}$d$^{-1}$, corresponding to rates previously reported for this region (24, 57). One station (which showed evidence of two size differentiated groups of diazotrophs, Fig. S6) had anomalously high values, ranging from 4.5 to 8.5 mgC m$^{-3}$d$^{-1}$ through the diel cycle. When this station is removed the values for the other 7 stations have a maximum value of 2.7 and an average of 1.36 mgC m$^{-3}$d$^{-1}$.

**Expression and rates: Nitrogen Fixation.** Highest positive $nifH$ community expression, at stations sampled in the Indian ocean, was 10,500 cells mL$^{-1}$ (19% of the total community) with an average of 600 cells mL$^{-1}$ (7% of the total community) whilst highest biological nitrogen fixation (BNF) rates were 16.38 nmol L$^{-1}$ h$^{-1}$ (mean 3.85 nmol L$^{-1}$ h$^{-1}$). The pattern observed in the eastern Indian Ocean (Fig. 3) mimics that of free-living single celled diazotrophs (*Cyanothece* sp.; 3). This includes a peak in $nifH$ expression (left y-axis) and N$_2$ fixation (right y-axis) at night and lowest values for both during the day. A regression model shows significant correlation between mRNA positive cells mL$^{-1}$ measured at the start of incubations with measured $^{15}$N$_2$ uptake ($F = 6.194, p = 0.026, n = 16$).

Nitrogen fixing organisms are less diverse and span a smaller size range (8), though some are symbiotic with larger algae (5, 64). Changes in fluorescence intensity of a sample may then be due more to positive target concentration than to differential expression (see equation 3). Therefore, concentration of active $nifH$ cells per ml was calculated and showed significantly correlation (without any time shifts) with $^{15}$N$_2$ uptake ($r = 0.554 p = 0.026$; Fig. 5) when examined for stations D6, D27, D48 and CTD16 (Fig. S1) from the south eastern Indian Ocean. The resultant regression (center line, Figure 5) can be used to estimate
biological nitrogen fixation BNF from the number of \text{nifH} positive cells in a sample, equation 3:

\[
\text{BNF} = (0.05)*\text{(positive cells/mL)}^{0.5} \quad (3)
\]

Using the above equation to estimate (BNF), rates range from 0.5-5 nmol l\(^{-1}\) h\(^{-1}\) with an average value of 2.76 nmol l\(^{-1}\) h\(^{-1}\), similar to those measured in this region (mean 3.5 nmol l\(^{-1}\) h\(^{-1}\); 54).

**Discussion**

Cultured and environmental microbial populations actively transcribing \text{rbcL} and \text{nifH} were successfully hybridized using the method described and optimized in this study. Simultaneous measurements using traditional, direct measurements of isotopic C and N incorporation were significantly correlated with fluorescently labeled cells (\(r^2 = 0.68\) and 0.45 respectively as shown in Figure 4 and 5 respectively). Our method yields not only number of cells actively fixing C and N but can also be used to estimate rates, providing a new approach for elucidating the connectivity of the C and N cycles simultaneously and on the same sample.

Using the various controls described in optimizing this method, the probes and optimization outlined herein represents a reliable tool to quantify the active primary production and biological nitrogen fixation communities in environmental samples. These findings indicate the \text{rbcL} and \text{nifH} oligonucleotide probes are specific enough to detect the targeted processes (primary production and nitrogen fixation) in oligotrophic waters but they are not overly specific as to lead to non-specific binding (e.g., in coastal waters where nitrogen fixation does not occur or in \text{E. coli} cultures that do not express either of these
genes). Under the optimized conditions described, the increase in fluorescence above background allowed differentiation of mRNA targets.

With the rapid advance of molecular work, including sequencing and RNA quantification it would be advantageous for researchers utilizing this method to check the probe sequences against the vastly increasing diversity outlined in the repositories available (ours were aligned Jan 2010). Care was taken in compiling multiple sequences from single and multiple species covering the diverse evolutionary histories of microbes targeted with these probes. Further, the probes and method (formamide concentration) were optimized to include variation in target sequences (using variable bases and lower hybridization temperatures). The parity in percentages of FISH tagged microbes and measured percentages using flow cytometry (targeting groups using pigments and SYBR green) suggests that the majority of the population targeted was hybridized using the probes outlined here in.

The importance of incorporation of diel periodicity into studies has been demonstrated previously in cultured phytoplankton including representatives of diatoms, dinoflagellates, chlorophytes, chrysophytes and cyanobacteria (21) and environmental samples (22). This variation is similarly important when employing this method due to the lag between expression and fixation for both carbon and nitrogen. Diel variability, expressed as the amplitude of the cycle, range from 4-6 in environmental samples (21, 22, 63). Similar amplitudes in periodicity are demonstrated here: *Synechococcus* sp. 2-6 (mean 4), *T. hemii* 3-29 (mean 13), *T. oceanica* 2-15 (mean 6.5). Typical protocols for primary production measurements (i.e. incubations measuring uptake of isotopically labeled C) suggest incubating through the light cycle (31). Integrated values of C fixation (P_{max} the maximum 14C uptake at saturating photon flux intensities) for the entire diel cycle in this study are nearly
two times higher than values integrated for the light period alone; ~46% less for

*Synechococcus* sp., ~60% less for *T. hemii*, ~44% less for *T. oceanica* similar to values previously

reported (22, 63). Both the diel variability and potential underestimation of C fixation are

greatest for *T. hemii*. Interestingly, the variability is lowest for *Synechococcus* sp..

Correspondingly, rate measurements are overestimated when measurements only occur in

the daylight for *Synechococcus* sp. and *T. oceanica*. *T. hemii* rates are underestimated, as the

peak in carbon fixation occurs at night. Phytoplankton community composition likely plays

an important role in variability in the diel cycle, as demonstrated by the different diel

patterns presented here.

The work described here also presents insight into the diel cycles of both carbon and

nitrogen fixation. Diatoms exhibit classical transcription and C fixation; both peak during the

day. However, cyanobacteria can fix large amounts of C at night. At first this seems counter

intuitive for photosynthesis, but actual fixation of C through action by RuBisCO is part of the

dark reactions, which do not require light. Further, the dinoflagellates appear to have a

different strategy altogether, with a large spike in C-fixation just prior to the start of the

light cycle. These results indicate that 1) primary producer community composition is

important in determination of net primary production and 2) a substantial portion of C

fixation may occur at night. *Synechococcus* sp. and *Prochlorococcus* sp. dominate the open

ocean environments (45). Thus, current estimates (15) and methods (1) of net primary

production which extrapolate primary production contingent upon daylight hours may be

underestimating the large amount of C fixation occurring in the dark.

There is no significant time lag between *nifH* transcription and BNF (Fig. 3). BNF and

*nifH* expression peaked at night and were minimal during the daylight hours; temporally

separating C and N fixation in natural populations. Such temporal separation of C and N
fixation has been observed in laboratory samples (3). Observation of this pattern in environmental samples is particularly interesting because it points to the possibility that the N₂ fixation of the single celled diazotrophic community (a large percentage of the microbial population in the south eastern Indian Ocean) in environmental samples is either being temporally separated on an organismal level, or that the O₂ generation (resultant of C fixation) on a community level inhibits N₂ fixation at a community level. Either scenario demonstrates the need to measure the interconnectivity of these two cycles on a fine temporal scale using the same methods.

The probes and optimization parameters outlined in this study provide an exciting new tool for elucidating the connectivity of the C and N cycles via simultaneous quantification of the fraction of the microbial community actively fixing C and N. Carbon fixation and biological nitrogen fixation are partially controlled by transcription of *rbcL* and *nifH* genes (9, 28, 40, 46). Therefore, mRNA TSA-FISH may offer new insights into the number of organisms actively performing these functions, as well as estimates of these two processes. Instead of highly stringent, species specific conditions, we also show it is possible to detect and quantify a more evolutionarily diverse, yet functionally similar, community. The power of quantifying functional gene expression *in situ* combines community enumeration and estimation of important rate measurements. Flow cytometry allows more rapid quantification of positive cells than traditional epifluorescence microscopy, as well as analysis of a larger number of events, thereby providing rapid, robust datasets. By combining detection of mRNA, using TSA-FISH, and enumeration via flow cytometry, the method optimized and outlined in this study offers robust, high sensitivity analyses of active microbial populations.
Acknowledgements

The authors would like to thank John Norman (Beckman Coulter) who was extremely helpful in setting up and trouble-shooting the flow cytometer. Critiques from George A. Jackson, Anya Schulze and Gilbert T. Rowe on early versions of this manuscript as well as comments of three reviewers and the editor improved the final version of the manuscript. We would also like to thank Mona Hochman of the Seafood Safety Lab for providing negative control cultures. The authors acknowledge the facilities, scientific and technical assistance of the National Imaging Facility at the Centre for Microscopy, Characterization & Analysis, The University of Western Australia, a facility funded by the University, State and Commonwealth Governments. ASM was supported by the NSF East Asia and Pacific Summer Institute Fellowship in conjunction with the Australian Academy of Science (Grant #1209422) during the initial phase of this work and later by a Texas A&M University Dissertation Fellowship. AKS was supported by a Texas Sea Grant award to AQ. The Australian phase of this work was supported by an ARC Discovery Grant to AMW and co-workers. Indian Ocean ship time was awarded by the Australian Marine National Facility to Helen Phillips (University of Tasmania) and Chris Wilcox (CSIRO).
References


27. Jeffrey SW, Mantoura RFC, Wright SW. 1997. Phytopankton pigments in
Tabita FR, Paul JH. 2007. Phytoplankton carbon fixation gene (RuBisCO) transcripts
and air-sea CO₂ flux in the Mississippi River plume. The International Society for
Microbial Ecology Journal 1:517-531.
29. John DE, Lopez-Diaz JM, Cabrera A, Santiago NA. 2012. A day in the life in the
dynamic marine environment: how nutrients shape diel patterns of phytoplankton
photosynthesis and carbon fixation gene expression in the Mississippi and Orinoco
Sons, Inc.
Global Ocean Flux Study (JGOFS) Core Measurements.
32. Kolber ZS, Falkowski PG. 1993. Use of Active Fluorescence to Estimate
fluorescence using fast repetition rate techniques: defining methodology and
34. Lee JHW, Katano T, Chang M, Han M-S. 2012. Application of tyramide signal
amplification-fluorescence in situ hybridization and flow cytometry to detection of


Table 1. Steps and methods involved in optimization of FISH.

<table>
<thead>
<tr>
<th>Process</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixation</td>
<td>48, 51</td>
</tr>
<tr>
<td>2% PFA 30 min RT</td>
<td></td>
</tr>
<tr>
<td>2-4% PFA 1-24 h</td>
<td>49</td>
</tr>
<tr>
<td>5-10% Formalin</td>
<td>62</td>
</tr>
<tr>
<td>Overnight</td>
<td></td>
</tr>
<tr>
<td>4% PFA 12-24 hrs</td>
<td>This Study</td>
</tr>
<tr>
<td>4°C</td>
<td></td>
</tr>
<tr>
<td>Storage</td>
<td>Collect and resuspend cells/pellet in molecular grade 100% EtOH and store at -80°C</td>
</tr>
<tr>
<td>Permeabilization</td>
<td>48, 51</td>
</tr>
<tr>
<td>[5mg mL⁻¹] lysozyme 30 min RT</td>
<td></td>
</tr>
<tr>
<td>96% EtOH</td>
<td>34</td>
</tr>
<tr>
<td>7% Tween20</td>
<td></td>
</tr>
<tr>
<td>0.1% DEPC 12 min RT</td>
<td>48</td>
</tr>
<tr>
<td>Or 0.5% SDS 10-15 min</td>
<td></td>
</tr>
<tr>
<td>[5mg mL⁻¹] lysozyme 30 min RT</td>
<td>This study</td>
</tr>
<tr>
<td>1% SDS 1 hr RT</td>
<td></td>
</tr>
<tr>
<td>Quenching</td>
<td>48, 51</td>
</tr>
<tr>
<td>0.1% DEPC 12 min RT</td>
<td></td>
</tr>
<tr>
<td>1-3% H₂O₂ 60 min RT</td>
<td>kit</td>
</tr>
<tr>
<td>0.1% DEPC 12 min RT</td>
<td>34</td>
</tr>
<tr>
<td>0.01% H₂O₂ 10 min RT</td>
<td>This study</td>
</tr>
<tr>
<td>Probe concentration</td>
<td>48, 51</td>
</tr>
<tr>
<td>250 ng µL⁻¹ 24 hrs</td>
<td></td>
</tr>
<tr>
<td>25 ng µL⁻¹ 2.5 hrs</td>
<td>34</td>
</tr>
<tr>
<td>50 ng µL⁻¹</td>
<td>52</td>
</tr>
<tr>
<td>Probe vs cell conc.</td>
<td></td>
</tr>
<tr>
<td>Fluorophore concentration</td>
<td>48, 51</td>
</tr>
<tr>
<td>0.25-0.5 µg mL⁻¹ 5 min RT</td>
<td></td>
</tr>
<tr>
<td>1:50</td>
<td>34</td>
</tr>
<tr>
<td>1:100 5-10 min RT</td>
<td>kit</td>
</tr>
<tr>
<td>1:100 30 min RT</td>
<td>This study</td>
</tr>
</tbody>
</table>
Figure 1. Hybridization of phytoplankton cultures using universal *rbcL* probe. Seven cultures were hybridized with *rbcL* and negative probes: *Amphora coffeaeformis* (AC), *Dunaliella tertiolecta* (DT), *Emiliania huxleyi* (EH), *Synnechococcus sp.* (SS), *Thoracosphaera heimii* (TH), *Thalassiosira oceanica* (TO), *Escherichia coli* (EC). Samples were run using 25% formamide for the *rbcL* (striped bars) probe and 40% formamide for the Negative probe (grey bars) representing roughly equivalent specificities. Error bars (+2 SE) represent triplicate samples (except *D. tertiolecta*, *T. oceanica* and *E. coli*, n = 2). White bars represent the intrinsic fluorescence of the samples (no probe added; controls). Some pigments remained, particularly in the diatoms. * represents samples with significant differences between positive and negative probes.

Figure 2. Diel cycle estimates of primary production measured using FISH of *rbcL* (P; black circles) or using 14C P<sub>max</sub> (white boxes) from triplicate laboratory samples. A = *Synechococcus* sp. B = *Thoracosphaera heimii* C = *Thalassiosira oceanica*. Symbols represent median values for replicate samples, top bar is the 75<sup>th</sup> percentile, and the bottom bar is the 25<sup>th</sup> percentile. Note the different strategies, *rbcL* mRNA expression leads C uptake in *Synechococcus* cultures, follows closely with *T. oceanica* and occurs at night in *T. heimii*.

Figure 3. Environmental diel *nifH* expression and N fixation. Note the strong diel cycle of increased expression and N<sub>2</sub> fixation during dark hours, and decreased expression and measured fixation during day light hours. Circles represent concentration of cells expressing *nifH* (cells mL<sup>-1</sup>), triangles represent N<sub>2</sub> fixation (nmol L<sup>-1</sup> h<sup>-1</sup>).
Figure 4: Regression of size normalized fluorescence and C fixation. C fixation follows \textit{rbcL} expression by 6 hours, therefore to determine the relationship between C fixation and \textit{rbcL} expression has been shifted 6 hours. Measurements of the different cultures are represented using different symbols (circles = \textit{Synechococcus} sp.; triangles = \textit{T. oceanica}; squares = \textit{T. hemii}). The regression line (center line) shows a strong correlation between size normalized fluorescence and \( P_{\text{max}} \) and can be used to calculate C fixation from size normalized fluorescence.

Figure 5: Regression of \textit{nifH} expression versus biological nitrogen fixation in the eastern Indian ocean. The different 24 hour stations are represented by the different shapes (CTD16 triangles, D6 square, D27 diamonds, D48 circles; station locations can be found on Fig. S1). The regression line (center line) shows a strong correlation between positive target cells and N\(_2\) fixation measured using traditional incubation methods and can be used to calculate \( N_2 \) fixation from positively tagged cells mL\(^{-1}\).