Environmental Microbiology (2006) **8**(10), 1721–1735 doi:10.1111/j.1462-2920.2006.01108.x

Application of a *nifH* oligonucleotide microarray for profiling diversity of N_2-fixing microorganisms in marine microbial mats

Pia H. Moisander,[1]* Lily Shiue,[2] Grieg F. Steward,[3] Bethany D. Jenkins,[4] Brad M. Bebout[5] and Jonathan P. Zehr[1]

[1]*Department of Ocean Sciences, University of California Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA.*

[2]*Center for Biomolecular Science and Engineering, University of California Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA.*

[3]*Department of Oceanography, University of Hawaii at Manoa, 1000 Pope Road, Honolulu, HI 96822, USA.*

[4]*Department of Cell and Molecular Biology, University of Rhode Island, 316 Morrill Hall, 45 Lower College Road, Kingston, RI 02881, USA.*

[5]*Exobiology Branch, NASA Ames Research Center, Moffett Field, CA 94035, USA.*

Summary

Diazotrophic community structure in microbial mats from Guerrero Negro (GN), Baja California, Mexico, was studied using polymerase chain reaction amplification of the *nifH* gene and a newly developed *nifH* oligonucleotide microarray. Ninety-six oligonucleotide probes designed for *nifH* sequences from cultivated isolates and the environment were printed on glass microarrays. Phylogenetic analysis showed that the probes represented all of the main *nifH* clusters. Specificity was tested by (i) evaluation of cross hybridization using individual targets, and (ii) comparison of the observed hybridization signals and those predicted from the sequences cloned from microbial mats. Signal intensity had a positive relationship with target concentration and the percentage identity between probe and target. Under moderate stringency and high target concentration, specificity of the probes varied from 77% to 100% with the individual targets tested. At the end of a 7-month long nutrient manipulation experiment in GN microbial mats, no expression of nitrogen fixation under nitrogen loading was detected, although a diverse community of diazotrophs was detected. The diversity in diazotrophic population present was higher than in the population expressing the *nifH* gene, and there were taxa specific differences in response to nutrients. The *nifH* microarray is a powerful tool for diazotroph community analysis in the marine environment.

Received 28 December, 2005; accepted 29 March, 2006. *For correspondence. E-mail pmoisander@pmc.ucsc.edu; Tel. (+1) 831 459 5846; Fax (+1) 831 459 4882.

Introduction

Microbial mats are thought to be analogues of Earth's earliest environments and sites where much of early microbial evolution occurred (Des Marais and Walter, 1999). In these environments, microbially mediated elemental cycling and biogeochemical processes are tightly coupled at small spatial and temporal scales (Hoehler *et al*., 2001). As a result, complex microbial communities are present (Risatti *et al*., 1994; Nübel *et al*., 1999). To a large degree, biogeochemical processes such as nitrification, sulfate reduction and photosynthesis, are performed by different groups of microorganisms. However, the genetic potential to fix nitrogen gas (N_2) into biologically available ammonium is widespread among diverse microbial groups found in mats, based on investigations of sequence information from the nitrogenase iron protein encoding gene *nifH* (Zehr *et al*., 1995; Steppe and Paerl, 2002, Omoregie *et al*., 2004a,b). It is not well understood how *nifH* diversity in mats is controlled by environmental factors, however, it is known that the overall community composition in microbial mats may change, for example, when mats are transferred from field to artificial conditions (Abed and Garcia-Pichel, 2001). It is also not understood what controls the proportion of the diazotrophic community that actively expresses the nitrogenase genes. It has recently been reported that nitrogenase gene diversity can be high in a number of environments, even though relatively few of the *nifH* phylotypes are expressed (Omoregie *et al*., 2004b; Zehr *et al*., 2006).

Characterization of community responses to environmental changes in functional gene diversity and gene expression has been hampered by the absence of methods that can adequately probe for gene diversity in

complex communities, such as microbial mats. The goal of this study was to develop a method that would allow high throughput profiling of many diazotroph community members simultaneously. We developed a functional gene microarray technique to examine N_2-fixing microbial community structure and function in a marine microbial mat.

Functional gene macro- and microarrays have been introduced to microbial ecology relatively recently and have been applied in a few ecosystem studies (e.g. Guschin *et al*., 1997; Small *et al*., 2001; Wu *et al*., 2001; Wilson *et al*., 2002; Bodrossy *et al*., 2003; Taroncher-Oldenburg *et al*., 2003; Jenkins *et al*., 2004; Rhee *et al*., 2004; Steward *et al*., 2004). The array we describe in this article is the first one specifically designed to target microbial mat diazotrophs, and includes a larger selection of *nifH* probes than prior arrays reported in the literature.

Polymerase chain reaction (PCR) amplification of target genes from DNA has been successfully combined with hybridization to microarrays for detection of diverse microbes from various environments, including soil, and pathogens in clinical samples and shellfish (e.g. Small *et al*., 2001; Wang *et al*., 2002; Mitterer *et al*., 2004; Panicker *et al*., 2004). When differences in species composition or gene expression need to be compared among many samples, microarrays can help to alleviate the bottleneck of traditional cloning and sequencing approaches. In the traditional approach, extensive sequencing would have to be performed on libraries prepared from every sample to allow meaningful comparisons among them. For many natural environments, including microbial mats, the complexity is so high that creating even one comprehensive sequence library can be daunting (Hughes *et al*., 2001; Kemp and Aller, 2004). Microarray approaches are attractive because once the community has been described and probes developed, new environmental samples representing part of this community can be screened against this matrix of probes repeatedly with minimal effort. As sequencing efforts produce more information on the systems under investigation, the probe set can continue to be supplemented as required.

In this article, we describe the design of a functional gene microarray targeting the *nifH* gene and the use of this array to profile diversity and expression of the *nifH* gene in marine microbial mats collected from Guerrero Negro (GN), Baja California Sur, Mexico, with and without nutrient loading.

Results and discussion

Ninety-six *nifH* sequences were selected from a database including several thousand sequences from GenBank (Table 1, supplemental Table S1). The criteria for selection of *nifH* sequences to be used as probes was based on including *nifH* sequences from: (i) GN mats (Omoregie *et al*., 2004a,b), because this was the initial application for this array, (ii) a wide range of cultivated microorganisms to represent different phylogenetic groups and, (iii) estuarine, coastal and open ocean environments to enhance the applicability of the array to other marine environments. The 60-mer probe region was the same in all probes and started directly downstream from the nifH1 primer, used for *nifH* amplification (Zehr and Turner, 2001). This region includes both conserved and non-conserved regions, allowing for phylogenetic discrimination. In typical *nifH* phylogenetic analysis, conservation of amino acids is considered most relevant in determining evolutionary relationships. To investigate how closely the discrimination based on DNA in the microarray probe region (60 bp) corresponds to phylogenetic relationships in a protein tree based on the *c*. 321 bp partial *nifH* region commonly used in *nifH* phylogenetic analyses, neighbour-joining trees were constructed and compared based on the two sequence regions (Fig. 1). Comparisons showed that the main *nifH* clusters (Chien and Zinder, 1994; Zehr *et al*., 2003) were conserved in the trees. Some differences in tree topology were observed, but most of the differences were within the major clusters. The relatedness among individual organisms determined by amino acid sequences may be slightly different from the phylogenetic affiliations based on microarray probe sequences (Fig. 1), however, the primary purpose and utility of a functional gene microarray is to detect differences in community composition and to target specific microorganisms, not to study evolutionary relationships per se.

The same probe region was used for all sequences in order to preserve phylogenetic information in the designed probes. Therefore, the GC content among probes varied to some degree (Table 1). Mean GC% and T_m were 56.6% ($\pm$9.5) and 68°C ($\pm$4) respectively. Most of the variability in GC% and T_m were from sequences from cultivated organisms originating from various environments. The groups of probes for GN and Chesapeake Bay had more consistent GC% and T_m than the probes overall, at 60.2% ($\pm$6.3) and 70°C ($\pm$3) respectively. Although the effects of variability in probe GC content are well known, this issue has been difficult to overcome in array design (e.g. El Fantroussi *et al*., 2003). One alternative is to modify the probe length to fix the T_m (Nicolaisen *et al*., 2005), but it is also known that two probes with identical T_m that are different in length may have different hybridization behaviour due to their length (e.g. Bodrossy *et al*., 2003).

Array performance

An internal control probe set was an essential component of the microarray design (Table 2). With one-colour

Table 1. Probe sequences of *nifH* included in the microarray.

Organism or collection site	GenBank accession no.	Environment	Group	GC%	T_m C°
Azospirillum brasilense ATCC 29145	M64344	Plant roots, Brazil	α-*Proteobacteria*	70	74
Sinorhizobium meliloti NRRL L-45	AY221811	*Medicago sativa* (alfalfa)	α-*Proteobacteria*	61	70
Xanthobacter flavus NRRL B-14838	AY221812	Soil, Russia	α-*Proteobacteria*	66	72
Anabaena cylindrica UTEX 629, PCC 7122	AY221813	Pond water, England	*Cyanobacteria*	50	66
Anabaena variabilis ATCC 29413 (*Nostoc* sp. PCC 7937)	U89346	Fresh water, Mississippi, USA	*Cyanobacteria*	51	66
Cyanothece sp. ATCC 51142	AF003336	Intertidal area, Texas, USA	*Cyanobacteria*	46	64
Gloeothece sp. ATCC 27152	L15554	Not known	*Cyanobacteria*	45	64
Nostoc commune UTEX 584	L23514	Soil extract, Scotland	*Cyanobacteria*	51	66
Marine stromatolite cyanobacterium (C1C5)	AF227947	Microbial mat, Bahamas	*Cyanobacteria*	50	66
Symploca sp. PCC 8002	AY221816	Marine mud, North Wales	*Cyanobacteria*	48	65
Crocosphaera watsonii WH8501 (*Synechocystis* sp.)	AF300829	North Pacific water column	*Cyanobacteria*	40	62
Tolypothrix sp. PCC 7101	AY221817	Soil, Borneo	*Cyanobacteria*	46	64
Trichodesmium thiebautii	U23507	Unknown	*Cyanobacteria*	40	62
Xenococcus sp. PCC 7305, ATCC 29373	U73135	Marine aquarium, USA	*Cyanobacteria*	43	63
Desulfobacter curvatus DSM 3379 , ATCC 43919	AF065620	Marine mud, Italy	δ-*Proteobacteria*	45	64
Desulfobacter latus DSM 3381	AY221822	Marine mud, Italy	δ-*Proteobacteria*	60	70
Desulfotomaculum nigrificans DSM 574	AY221823	Unknown	*Firmicutes*	46	64
Desulfovibrio salexigens	AF227926	Unknown	δ-*Proteobacteria*	58	69
Arcobacter nitrofigilis (*Campylobacter nitrofigilis*) DSM 7299	AY221825	*Spartina alterniflora* roots	ε-*Proteobacteria*	30	58
Paenibacillus durus (*P. or Bacillus azotofixans*) DSM 5976	AY221826	Wheat roots	*Firmicutes*	55	68
Azotobacter vinelandii (*nifH2* alternative ngenase Fe protein)	X13519	Unknown	γ-*Proteobacteria*	66	72
Azotobacter vinelandii (*nifH1*)	M20568	Unknown	γ-*Proteobacteria*	58	69
Marine mat bacterium (Bird Shoal sw4)	AF046829	Microbial mat, NC, USA	Uncultivated	41	62
Vibrio diazotrophicus DSM 2604, ATCC 33466	AY221828	Sea urchin gastrointestinal tract	γ-*Proteobacteria*	48	65
Treponema pectinovorum	AF325798	Human mouth	Spirochaetes	60	70
Methanococcus voltae	X03777	Unknown	Archae	31	59
Methanococcus vannielii DSM 1224	AY221830	Marine mud, San Francisco Bay	Archae	36	60
Chlorobium limicola DSM 245	AY221831	Hot spring, USA	Chlorobia	61	70
CB, North Bay, 8 m, 4 April 2001, clone CB894H10	AY223944	Estuarine water column	Uncultivated	60	70
CB, North Bay, 8 m, 4 April 2001, clone CB894H4	AY223946	Estuarine water column	Uncultivated	55	68
CB, North Bay, 8 m, 4 April 2001, clone CB894H7	AY223949	Estuarine water column	Uncultivated	66	72
CB, North Bay, 8 m, 4 April 2001, clone CB894H8	AY223950	Estuarine water column	Uncultivated	60	70
CB, North Bay, bottom, 4 April 2001, clone CB895H10	AY223951	Estuarine water column	Uncultivated	60	70
CB, North Bay, bottom, 4 April 2001, clone CB895H2	AY223953	Estuarine water column	Uncultivated	55	68
CB, North Bay, bottom, 4 April 2001, clone CB895H3	AY223954	Estuarine water column	Uncultivated	68	73
CB, North Bay, bottom, 4 April 2001, clone CB895H7	AY223956	Estuarine water column	Uncultivated	65	72
CB, Mid-Bay, surface, 5 April 2001, clone CB907H2	AY223968	Estuarine water column	Uncultivated	66	72
CB, Mid-Bay, surface, 5 April 2001, clone CB907H4	AY223976	Estuarine water column	Uncultivated	65	72
CB, Mid-Bay, surface, 5 April 2001, clone CB907H8	DQ269130	Estuarine water column	Uncultivated	66	72
CB, Mid-Bay, 11.6 m, 5 April 2001, clone CB908H2	DQ269131	Estuarine water column	Uncultivated	63	71
CB, Mid-Bay, bottom, 5 April 2001, clone CB909H4	DQ269132	Estuarine water column	Uncultivated	63	71

Table 1. *cont.*

Organism or collection site	GenBank accession no.	Environment	Group	GC%	T_m C°
CB, Mid-Bay, bottom, 5 April 2001, clone CB909H7	AY224003	Estuarine water column	Uncultivated	68	73
CB, Mid-Bay, bottom, 5 April 2001, clone CB909H8	AY224004	Estuarine water column	Uncultivated	60	70
CB, South Bay, surface, 6 April 2001, clone CB910H10	AY224006	Estuarine water column	Uncultivated	53	67
CB, South Bay, 4 m, 6 April 2001, clone CB911H1	AY224014	Estuarine water column	Uncultivated	56	68
CB, South Bay, 4 m, 6 April 2001, clone CB911H3	AY224015	Estuarine water column	Uncultivated	66	72
CB, South Bay, 4 m, 6 April 2001, clone CB911H4	AY224016	Estuarine water column	Uncultivated	50	66
CB, South Bay, bottom, 6 April 2001, clone CB912H5	AY224023	Estuarine water column	Uncultivated	61	70
CB, South Bay, bottom, 6 April 2001, clone CB912H8	AY224024	Estuarine water column	Uncultivated	58	69
CB, South Bay, bottom, 6 April 2001, clone CB912H9	AY224025	Estuarine water column	Uncultivated	66	72
CR, 8 m, 1 July 2000, clone CB914H3	AY224028	Estuarine water column	Uncultivated	60	70
CR, 8 m, 1 July 2000, clone CB914H6	AY224030	Estuarine water column	Uncultivated	60	70
CR, 8 m, 1 July 2000, clone CB914H9	AY224033	Estuarine water column	Uncultivated	55	68
CR, 1 m, 1 July 2000, clone CB916H1	AY224034	Estuarine water column	Uncultivated	60	65
CR, 1 m, 1 July 2000, clone CB921H10	AY224038	Estuarine water column	Uncultivated	63	71
CR, 1 m, 1 July 2000, clone CB921H1	AY224039	Estuarine water column	Uncultivated	58	69
CR, 1 m, 1 July 2000, clone CB921H4	AY224041	Estuarine water column	Uncultivated	63	71
CR, 1 m, 1 July 2000, clone CB921H7	AY224042	Estuarine water column	Uncultivated	68	73
CR, 1 m, 1 July 2000, clone CB921H8	AY224043	Estuarine water column	Uncultivated	55	68
CR, 1 m, 1 July 2000, clone CB921H9	AY224044	Estuarine water column	Uncultivated	68	73
CR, surface, 3 April 2001, clone CB891H2	DQ269129	Estuarine water column	Uncultivated	55	68
CR, surface, 3 April 2001, clone CB891H8	AY223911	Estuarine water column	Uncultivated	68	73
GN, *Microcoleus* mat, top 5 mm, June 2001, clone GN1061A16	DQ269137	Microbial mat	Uncultivated	68	73
GN, *Microcoleus* mat, top 5 mm, June 2001, clone GN1061A24	AY232338	Microbial mat	Uncultivated	65	72
GN, *Lyngbya* mat, top 5 mm, June 2001, clone GN1062A1	DQ269136	Microbial mat	Uncultivated	56	68
GN, *Lyngbya* mat, top 5 mm, June 2001, clone GN1062A02	AY244717	Microbial mat	Uncultivated	68	73
GN, *Lyngbya* mat, top 5 mm, June 2001, clone GN1064A14	DQ269139	Microbial mat	Uncultivated	60	70
GN, *Lyngbya* mat, top 5 mm, June 2001, clone GN1062A16	AY232350	Microbial mat	Uncultivated	53	67
GN, *Lyngbya* mat, top 5 mm, June 2001, clone GN1062A17	DQ269138	Microbial mat	Uncultivated	66	72
GN, *Lyngbya* mat, top 5 mm, June 2001, clone GN1062A19	DQ269141	Microbial mat	Uncultivated	66	72
GN, *Lyngbya* mat, top 5 mm, June 2001, clone GN1062A20	DQ269142	Microbial mat	Uncultivated	53	67
GN, *Lyngbya* mat, top 5 mm, June 2001, clone GN1062A911	DQ269140	Microbial mat	Uncultivated	55	68
GN, *Lyngbya* mat, top 5 mm, June 2001, clone GN1063A01	AY232362	Microbial mat	Uncultivated	60	70
GN, *Lyngbya* mat, top 5 mm, June 2001, clone GN1063A08	AY232364	Microbial mat	Uncultivated	60	70
GN, *Lyngbya* mat, top 5 mm June 2001, clone GN1063A15	AY232366	Microbial mat	Uncultivated	50	66
GN, *Lyngbya* mat, top 5 mm, June 2001, clone GN1063A17	AY232368	Microbial mat	Uncultivated	66	72
GN, *Lyngbya* mat, top 5 mm, June 2001, clone GN1063A22	AY232372	Microbial mat	Uncultivated	71	74
GN, *Lyngbya* mat, top 5 mm, June 2001, clone GN1063A23	AY232373	Microbial mat	Uncultivated	68	73
GN, *Microcoleus* mat, top 5 mm, June 2001, clone GN1064A02	AY232381	Microbial mat	Uncultivated	63	71
GN, *Microcoleus* mat, top 5 mm, June 2001, clone GN1064A17	AY232383	Microbial mat	Uncultivated	53	67
GN, *Microcoleus* mat, top 5 mm, June 2001, clone GN1064A18	AY232384	Microbial mat	Uncultivated	53	67
GN, *Microcoleus* mat, top 5 mm, June 2001, clone GN1064A23	AY232387	Microbial mat	Uncultivated	63	71
GN, *Microcoleus* mat, top 5 mm, July 2001, clone GN821A2	DQ269134	Microbial mat	Uncultivated	51	66

Table 1. *cont.*

Organism or collection site	GenBank accession no.	Environment	Group	GC%	T_m C°
GN, *Microcoleus* mat, top 5 mm, July 2001, clone GN821A09	AY299641	Microbial mat	Uncultivated	38	61
GN, *Microcoleus* mat, top 5 mm, July 2001, clone GN821A11	AY299643	Microbial mat	Uncultivated	55	68
GN, *Microcoleus* mat, top 5 mm, July 2001, clone GN821A12	DQ269135	Microbial mat	Uncultivated	61	70
GN, *Microcoleus* mat, mid 10–15 mm, July 2001, clone GN822A08	AY244723	Microbial mat	Uncultivated	55	68
GN, *Microcoleus* mat lower 10–15 mm, clone GN823A08	AY244735	Microbial mat	Uncultivated	50	66
HOT, clone HT1200	AF059637	N. Pacific Ocean water column	Uncultivated	38	61
HOT, clone HT1900	DQ269147	N. Pacific Ocean water column	Uncultivated	41	62
HOT, clone HT1901	AF299426	N. Pacific Ocean water column	Uncultivated	65	72
HOT, clone HT1915	DQ269144	N. Pacific Ocean water column	Uncultivated	63	71
HOT, mRNA clone HT1918	DQ269145	N. Pacific Ocean water column	Uncultivated	45	64
HOT, mRNA clone HT1919	DQ269146	N. Pacific Ocean water column	Uncultivated	40	62
HOT, 25 m marine isolate, clone HTBACTJ29A	DQ269143	N. Pacific Ocean water column	Uncultivated	70	74
HOT, marine isolate HOT J7, clone CC1114H1	DQ269133	N. Pacific Ocean water column	Uncultivated	60	70

Original environments are indicated for both the cultivated organisms and environmental samples. Group designations are based on the NCBI database. ATCC, American Type Culture Collection, USA; DMSZ, German National Resource Center for Biological Material, Germany; NRRL, Agricultural Research Service Culture Collection, USA; PCC, Pasteur Culture Collection, France; UTEX, Culture Collection of Algae at the University of Texas at Austin, USA. CR, Choptank River; CB, Chesapeake Bay; GN, Guerrero Negro; HOT, Hawaiian Ocean Time Series station Aloha. T_m at 30% formamide and 300 mM NaCl was calculated using the MELT software (Jo P. Sanders, http://www.xs4all.nl/~sahjps/melt.html).

Table 2. Control genes included in the microarray.

No.	GenBank accession no.	Gene (human)	Probe	Gene bp	Probe T_m	Probe GC%
1	BC007659	Diaphorase NADPH/NADPH (cytochrome b-5 reductase)	TGGGCACACTCCAGCAGACGCCCGAATTCAAATCCTG GAAGGATGGAAGAAACGCCTGGA	1121	68	55
2	BC008972	Uncharacterized bone marrow protein	TACTCCCAGACTCCGTACGCCTTCACGTCCTCCTCCAT GCTGAGGCGCAATACACCGCTT	1118	69	58
3	BC002599	Corticotropin releasing hormone from brain	GGAAATGGCCAGGGCCGAGCAGTTAGCACAGCAAGCT CACAGCAACAGGAAACTCATGGA	825	68	55
4	BC065005	Mitochondrial ribosomal protein	GCACTGCATGCCTGGCCCCCGAATGACCCCCAGATTCG ACTCCAACAACGTGGTCCTCAT	623	70	70
5	BC004236	Ubiquitin carrier protein	GGGCTCTCTTCCTCCTTCCACCGTGACCCCAACCTCTC CTGTCCCCTCCCTCCAACTCTG	1031	71	63
6	BC015569	ADP-ribosylation-like factor 9 interacting protein	CCATCCGCCCCCAGCTCTGTTTGTGATGTACCCCTCCT CCTGTGTGCTTTCTTCCCCAGC	1014	70	60

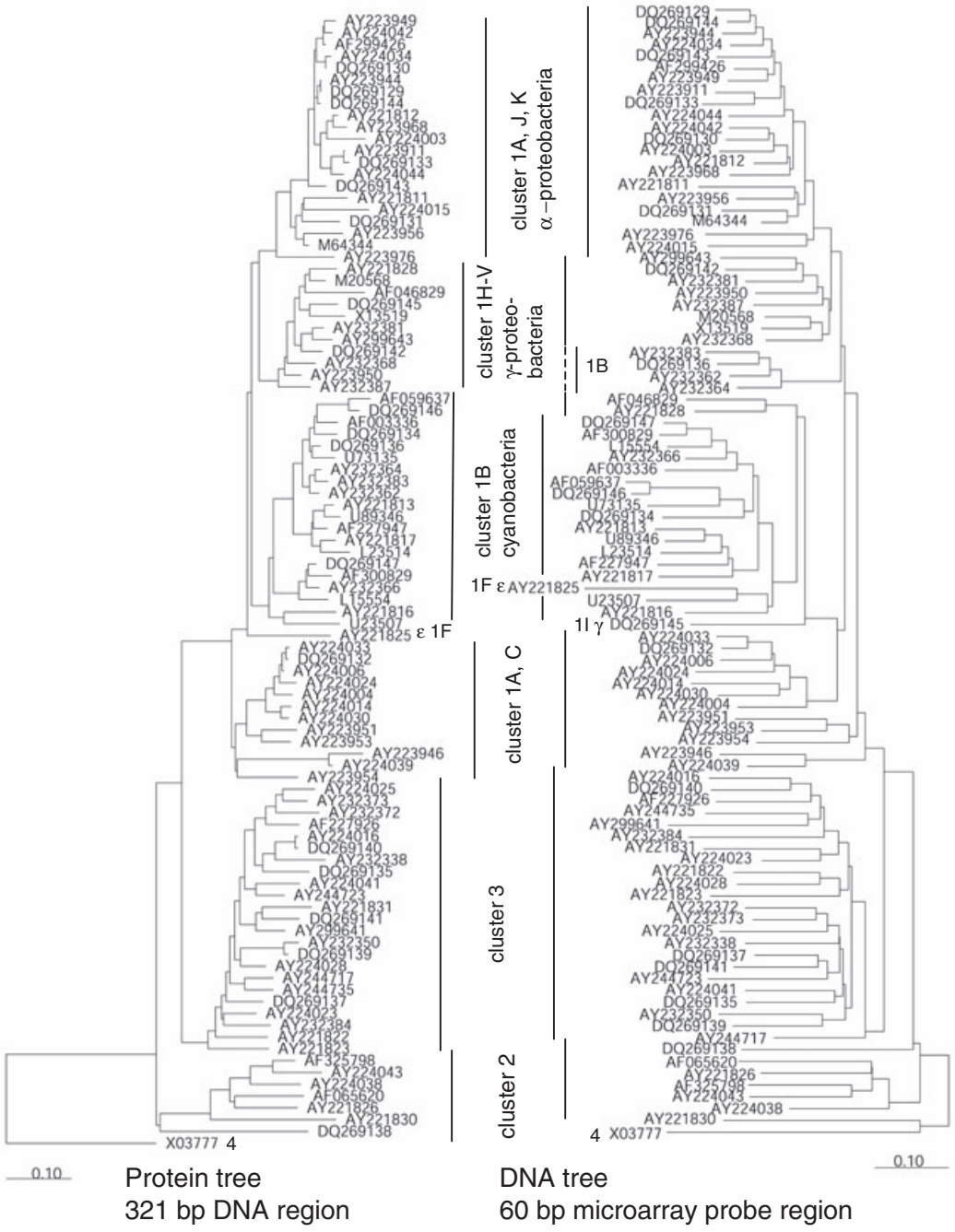

Fig. 1. Comparison of neighbour-joining trees constructed with the (A) translated 321 bp *nifH* region and (B) the microarray probe region (60 bp, DNA tree). Both regions start directly downstream from the inner *nifH* forward primer. The *nifH* clusters are shown following Zehr and colleagues (2003).

applications such as ours careful design of internal controls is necessary, to make comparisons between slides. The internal control set was useful for several reasons: (i) it allowed us to normalize hybridization signal and thereby to compare signal intensities between slides, (ii) it served as a measure of background binding (target hybridization to any DNA on the slide), and (iii) the controls served as 'landing lights' for array and spot finding during data analysis. Using our threshold criteria (see *Experimental procedures*) the signal intensities in negative controls were consistently zero, suggesting that background binding to DNA was negligible. Therefore, we used the local spot background intensities to determine background.

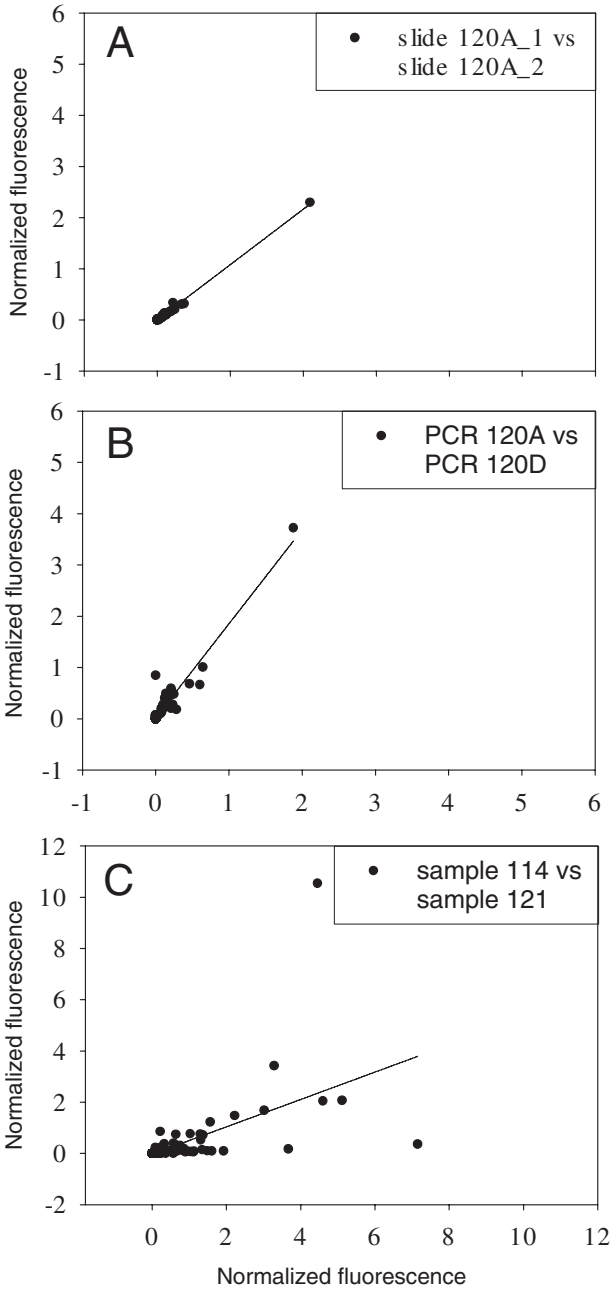

Fig. 2. Examples of experimental variability. Correlation of fluorescence signal between (A) two slides hybridized with the same PCR product, (B) two independent PCR reactions from the same microbial mat DNA extract, and (C) two independently processed replicate mat samples.

We evaluated the variability among hybridizations by hybridizing the same biotinylated PCR product four times (Fig. 2A). The data from replicate slides had a high degree of correlation, however, it could not be quantified due to the skewed data distribution (large number of zeros and few high values), even after transformation. The variability between slides increased when individual PCR reactions from the same DNA extract were compared (Fig. 2B), but the correlation is still high. Variability due to patchiness of the mat samples was tested by extracting replicate microbial mat samples and carrying out independent PCR reactions and hybridizations. The variability between replicate mat samples was clearly higher than variability among slides or PCR reactions, therefore biological variability is the main driver of observed differences between slides (Fig. 2C).

The *nifH* probe sequences on the array had an average sequence identity of 70% (SD 0.1%), minimum 21% and maximum 95%. The average probe T_m was 70°C, and a hybridization temperature of T_m 20°C (50°C) was used in initial tests. Preliminary tests showed that at 50°C, target sequences were detected with little or no cross hybridization, and probes originating from GN had positive signals when hybridized to samples amplified from GN microbial mats (data not shown). At the same time, low or no signal was present from other probes on the array. Consequently, most hybridizations were carried out at 50°C for 18 h, with washing conditions as recommended by the slide kit manufacturer, and the degree of cross hybridization (array specificity) was studied under these conditions (Figs 3 and 4). Under increasing target concentrations (tested using the *Anabaena cylindrica* clone sequence AY221813), signal intensity was close to linear between 1 and 12 ng of target DNA in the hybridization mix but appeared to saturate between 12 and 16 ng (Fig. 4). In some cases, in a hybridization of targets that each were identical to one probe on the array, low non-specific signal was present from a few of the probes that were less than 100% identical (Figs 3 and 4). In cases where cross hybridization was observed, the identity required for detection was not dependent on the GC content of the probes. The non-specific signal intensity was usually higher as probe similarity with the target increased, however, this relationship was not entirely consistent. Cross hybridization was defined as signal that was 7% or more of the normalized signal from the 100% identity target probe. With this cut-off, the identity of 77% or higher was required for hybridization to occur under the stringency used. Therefore, in hybridizations with environmental samples, if all signal lower than 7% of the highest signal is removed from the analysis, the array can detect with approximately 77% specificity or higher. If the low level signal is removed from the analysis, some information of low abundance targets may be lost, however, this method significantly improves the array specificity. This method also assumes that at least one target in the hybridization has close to 100% identity with one probe on the array, underscoring the importance of using the array only in environments that are well represented in the probe selection. Data interpretation is also complicated by additive effects from

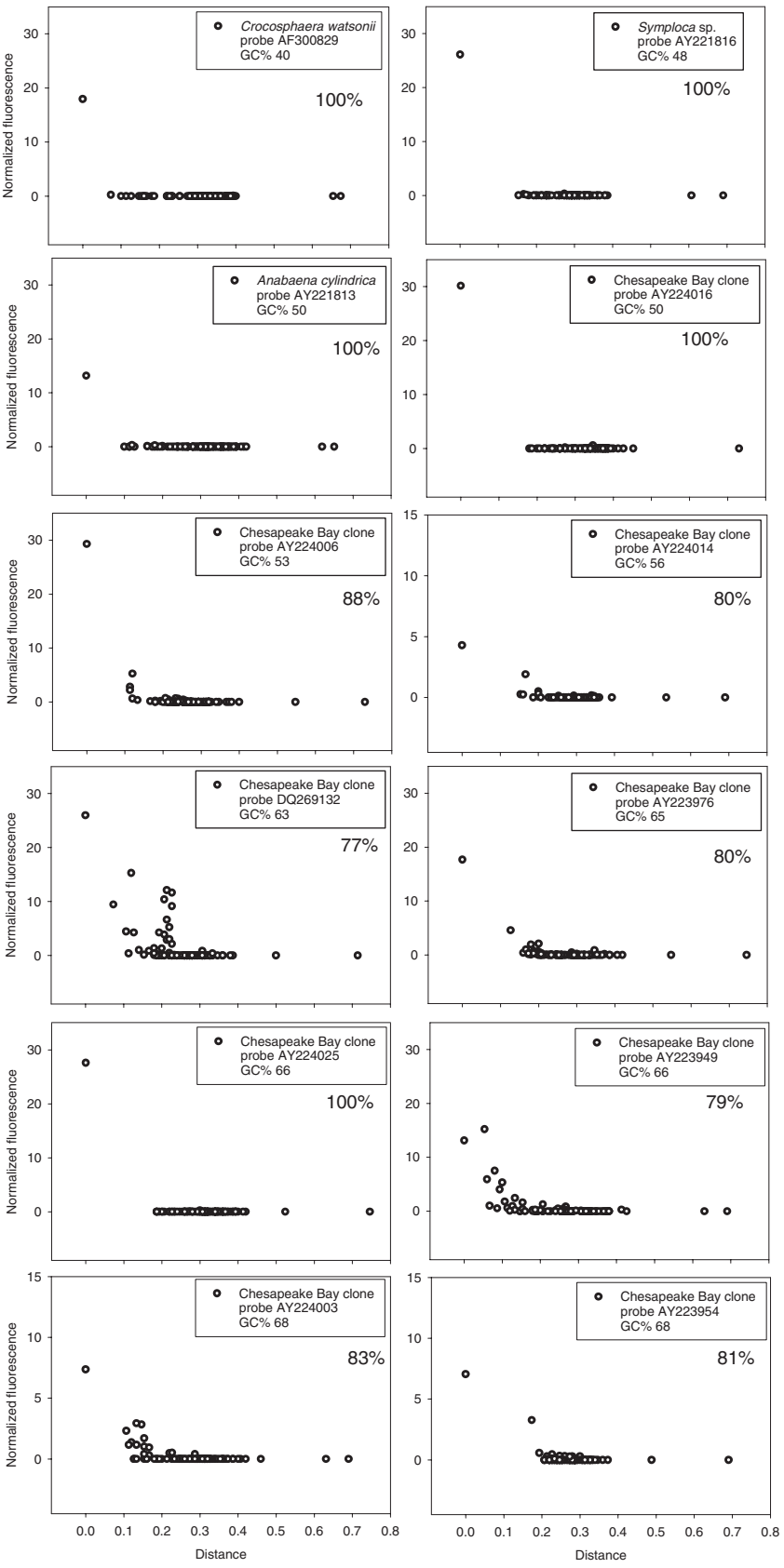

Fig. 3. Specificity tests carried out using individual PCR products as target sequences matching eight array probes. GC% for each probe is shown in the legends. Distance is proportion of nucleotides that differ between the probe (immobilized spot on the array) and target (PCR product in the hybridization mix). Each dot represents the average of eight replicate spots (median fluorescence with local background subtracted and normalized with internal controls) of each probe. Specificity for each target is shown. Specificity is presented as minimum probe identity (%) required for signal of 7% or more of the normalized target signal.

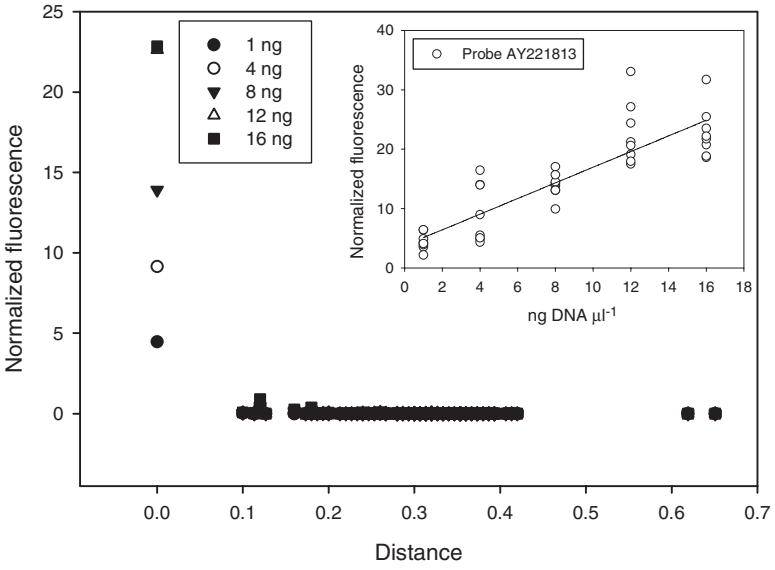

Fig. 4. Specificity test carried out for varied concentrations of target matching the *Anabaena cylindrica* probe AY221813. The inset shows the relationship between total amount of DNA (ng) in the array hybridization mix and normalized fluorescence intensity of individual spots in the probe AY221813 ($R^2 = 0.94$, $n = 40$). To include all spots in the regression, median fluorescence intensity in each spot was exceptionally normalized (usually average was calculated for replicate spots first, then the average was normalized).

multiple targets cross-hybridizing, which is a problem shared by other functional gene arrays. Although hybridizations with individual targets indicated probe specificity minimum of 77%, targets from environmental samples are likely to be detected with a higher specificity than that, because target abundance in environmental samples is lower than in these tests, and the signal intensity and degree of cross hybridization increase with target abundance. With more target, non-specific signal occurred with less identity to a probe, and with increased DNA concentrations, the number of probes showing non-specific signal increased (data not shown). The cross hybridization tests were carried out under conditions where a single target PCR product was initially at a concentration of 6 ng μl^{-1} (prior to all labelling steps) while in environmental samples that are mixtures of many targets, 5 ng μl^{-1} PCR product starting concentration were used. Targets in hybridizations with environmental samples are therefore present at considerably lower concentrations than in the specificity tests with individual targets, leading to higher specificity in natural samples than in the test results shown in Fig. 3.

With probes as long as 60 nucleotides, significant differences in GC content may skew the hybridization signal and complicate interpretation. In natural environmental samples, diverse targets with different abundances are present simultaneously, competing for the hybridization sites with dynamics that are difficult to replicate in artificial mixtures. As a second approach to validate the array response specificity, array probe sequences and sequences from clone libraries obtained from microbial mats were compared. The signal intensity of each probe was correlated with the distance to the sequence of the most similar clone (Fig. 5). In most cases, signal intensity increased with increasing sequence identity (decreasing distance). This suggested the probes detected targets in mixed environmental samples in a comparable manner as in hybridization mixtures containing only one target identical to one probe. The comparison also suggested that probes with identity as low as 68% may sometimes hybridize. The absolute lower limit of sequence identity required for positive signal or degree of quantification cannot be determined from these comparisons, however, because target abundances are not known and it is likely the clone libraries were not saturated. It is estimated that

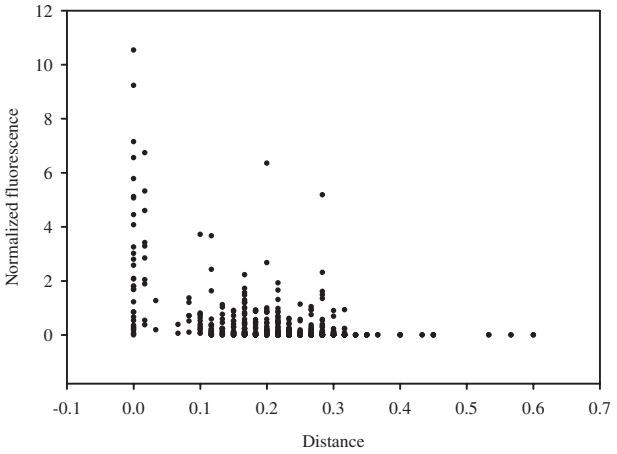

Fig. 5. Normalized array fluorescence in each probe plotted against distance of each probe to most similar clone sequenced from each nutrient treatment. Pooled data are shown from the three nutrient treatments (N, P, no nutrients), with two replicate samples per treatment. Separate clone libraries were included for each nutrient treatment. Each dot represents average of eight replicate spots on the array (median fluorescence in each spot with local background subtracted and normalized with positive controls) of each probe.

environmental microbial clone libraries often need to reach several hundred to approach the actual diversity of the system (Hughes *et al*., 2001; Kemp and Aller, 2004).

Based on tests with target sequences that were identical to probes on the array, specificity was comparable with that reported for other functional gene microarrays and macroarrays (Wu *et al*., 2001; Taroncher-Oldenburg *et al*., 2003; Steward *et al*., 2004). Specificity could most likely be improved by increasing hybridization stringency. However, with increased stringency (and specificity), information from the community is more easily lost, if specific probes are not included on the array for all target sequences present in the sample. This is a very likely scenario when working with environmental samples, in particular with systems such as microbial mats in which diversity is extremely high and may not be fully characterized. From the perspective of environmental microbiology where sequencing efforts continuously detect novel diversity, some degree of cross hybridization is probably desirable. In particular, if a microarray is applied to a gene whose background database is much smaller than that for *nifH*, it is difficult to predict how much diversity the array will fail to detect. Additionally, it was evident from the results that the sequence similarity is not an absolute indicator of the degree of potential cross hybridization. The ratio of AT to GC and positions of mismatches are likely to have an effect on differences in cross hybridization among probes. We continue to investigate the array specificity limits with a microarray that has the same chemistry and probe design as the 96-probe array described in this study but a larger probe selection (768 probes) (Moisander *et al*., unpublished data).

The final number of gene copies following amplification may not be directly proportional to initial target abundance in the community, therefore relative signal intensities should be interpreted with caution. However, while better approaches remain to be developed, PCR is a means to increase signal intensity to detectable levels for profiling functional gene diversities from environmental samples, and allows microarrays to be developed, at a minimum, for qualitative detection of differences among communities. This is similar to what any other PCR-based community analyses can provide, including T-RFLP (terminal restriction fragment length polymorphism), TGGE/DGGE (thermal/denaturing gradient gel electrophoresis), or ARISA (automated ribosomal intergenic spacer analysis). In addition, these other community profiling techniques require parallel sequencing in order to obtain information about the community composition. The power of the microarray approach lies in the fact that limited phylogenetic information about the community composition is obtained directly, without a requirement for cloning and sequencing, and that a sample can be simultaneously screened with up to thousands of probes.

Effect of nutrients on microbial mat community composition and function

The *nifH* microarray was used to investigate diazotroph community composition and function of a *Lyngbya* sp. cyanobacterium dominated microbial mats from GN. Prior investigations have shown that N_2 fixation in these mats is occurring predominantly at night (Omoregie *et al*., 2004b), therefore we compared the composition of *nifH* DNA from the mats with *nifH* transcripts (mRNA) at midnight. Comparisons of array hybridizations of DNA and RNA [amplified by PCR and reverse transcription (RT)-PCR, respectively] revealed dramatic differences (Fig. 6). A great diversity of *nifH* sequences was detected in DNA samples, but only a small subset was actively expressing *nifH*. Similar trends from GN mats were previously reported (Omoregie *et al*., 2004b) by cloning and sequencing. The major trends in the data were consistent among independently processed mat sample replicates for both DNA and RNA in each nutrient treatment. Several probes that had moderate signal when hybridized with targets amplified from DNA did not detect any signal from targets amplified from RNA. Only a few probes showed high signals in targets amplified from RNA, indicating expression of *nifH*. Seven probes (DQ269134, DQ269136, DQ269137, AY232364, AY232366, AY232372, AY232373), all of which represented sequences from the GN mats, were among the ones that detected expression. Probe DQ232366 detected high expression in each of the independent replicates of both 'no nutrients' and P treatments. *Cyanothece* sp. was the closest relative to this sequence in GenBank based on the *nifH* amino acid sequence (95% identity in the full amplified *nifH* region). A second commonly expressed cyanobacterial probe DQ269136 detected high expression in the P treatment, and low levels in the 'no nutrients' treatment, while its abundance based on DNA appeared reduced in the N treatment. Closest relative to this probe based on amino acid sequence was *Leptolyngbya* sp. (97% identity). *Leptolyngbya* sp. was the closest relative to probe AY232364 as well (95% identity). Two of the sequences showing high expression in both P and 'no nutrients' treatment had closest known relatives in *deltaproteobacteria*, with 86–87% sequence identities to *Desulfovibrio salexigens* (probe AY232372) and *Desulfovibrio vulgaris* (probe AY232373). These sequences remained present in the mat at high abundances under N loading, in non-N_2 fixing conditions, suggesting these diazotrophs can efficiently rely on alternative sources of N. One probe that detected expression (DQ269134) was identical to *nifH* from *Halothece* sp. (cyanobacterium). The closest

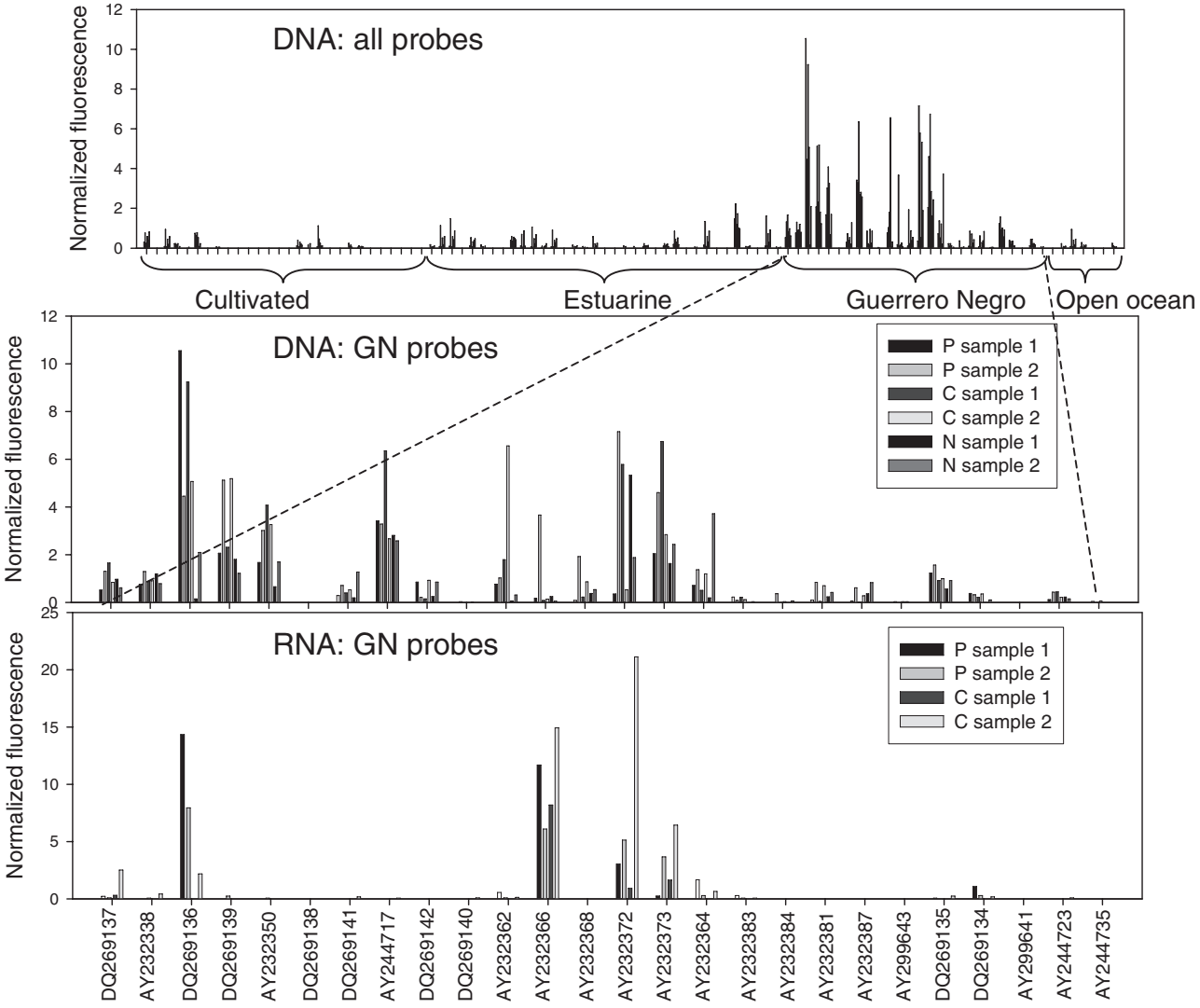

Fig. 6. Hybridization signal strengths for each probe on the array when hybridized to targets prepared from a GN microbial mat. Both DNA and RNA were extracted from two independent mat samples from P, N and 'no nutrients' (C) treatments each, and all steps through hybridizations were carried out independently for each sample. No transcription was detected in the N treatment. The upper panel shows DNA data from all probes, and the panels below show DNA and RNA data from the GN probes only.

known relative to probe DQ269137 was *Prosthecochloris aestuarii* (Chlorobia, draft genome) with 87% identity at the amino acid level. The relative importance of different diazotrophic groups to N_2 fixation in these microbial mats has continued to be unresolved. Our microarray data indicate that, based on the relative signal intensities, transcript abundance in these microbial mats is comparable in cyanobacterial and deltaproteobacterial (or cluster 3) groups, suggesting both groups significantly contribute to N_2 fixation. Sulfate reducers (in cluster 3) were shown to be actively fixing N_2 in other microbial mats (Steppe and Paerl, 2002).

Although in some cases there were large differences in signal intensities between replicate mat samples, major trends were consistent between the replicates in both DNA and RNA arrays. The differences between replicate mat samples are likely due to mat patchiness, but in addition there is variability in several analytical steps that include extraction, PCR, biotinylation and hybridization.

Interestingly, in the N amended treatment, *nifH* expression was not detected by the array, yet numerous probes showed positive signal and verified presence of a diverse assemblage of *nifH* containing organisms. Combined N can rapidly repress nitrogenase activity in cultivated organisms (e.g. Postgate, 1987). Therefore, persistence of diazotrophs during the experiment is remarkable, considering the mats had been receiving combined N additions for several months, and suggests diazotrophs can effectively compete or maintain presence in mats under conditions of high N. Taken together, our results suggest

that in spite of high *nifH* diversity in the GN mats, only a few of the organisms that possess the *nifH* gene are actively expressing *nifH* at a given time, even in instances when availability of fixed N is potentially limiting rates of primary production or growth. While the probe selection may not completely cover the diversity in the system, the array was useful in identifying differences. Overall, these results indicate that the presence of N_2 fixation genes in microbial mats may be a relic from past environments, and/or that some of them may have been retained in microbial mats for purposes other than N_2 fixation.

In summary, we have demonstrated the usefulness of the *nifH* array as a tool for diazotroph community analysis. The *nifH* microarray described in this article can distinguish microorganisms belonging to different *nifH* clusters. Based on the microarray results, active N_2 fixers in GN *Lyngbya* spp. microbial mats form only a small proportion of all diazotrophs present. The results revealed that both cyanobacteria and organisms belonging to cluster 3 significantly contribute to N_2 fixation in the mats. With further development, the *nifH* microarray method may become even more powerful approach in microbial community fingerprinting. Although a high throughput method, current functional gene microarrays have limitations in specificity and are most informative when used in combination with methods that allow more specific target detection, such as quantitative PCR. The number of probes on the array is currently a limitation for the detection power in previously uncharacterized environments, however, with increases in probe diversity, the array can become a more general tool applicable in various marine environments.

Experimental procedures

Array printing

The microarrays were constructed with 5′-Acrydite modified 60-bp sense oligonucleotide probes (Integrated DNA Technologies, Iowa). The oligonucleotides were aliquoted into duplicate wells of the printing plate at 21 μM in aqueous universal EZrays buffer (Matrix Technologies, NH, USA), with 0.005% Sarkosyl. The EZ rays universal slides were treated before and after spotting according to the manufacturer's instructions. Briefly, to activate the surface thiol groups prior to spotting, slides were treated with an activator solution for 15 min, then rinsed and dried in a centrifuge. After spotting, the residual thiol groups were quenched by treating the slides with a quench solution for 30 min, then washed twice with deionized water and dried in a centrifuge. Each well of the oligonucleotide was printed onto each slide four times, resulting in eight replicate spots for each probe. The arrays were printed at the UCSC microarray facility using a custom made spotting robot with a 16-pin set-up and ArrayMaker software (Joseph DeRisi, University of California San Francisco). The temperature and relative humidity during the print run averaged 24°C and 30% respectively. After spotting, slides were stored at room temperature in the dark. A few slides from each spotting batch were stained with 65 μl of a 1:200 dilution of OliGreen (Molecular Probes, Eugene, OR) in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.5) for 45 min under a cover slip, rinsed twice with TE for 10 s, then dried in a filtered air stream, in order to evaluate spot quality. After the slides were stained with OliGreen, they were scanned at 532 nm with the GenePix 4000B scanner (Axon Instruments).

Normalization

For between-slide comparisons, positive and negative control probes were developed using a set of human genes. Several criteria were used when choosing control genes from the NCBI database. To simplify amplification, we required sequences to be approximately 1 kb in length. An oligonucleotide (60-mer) was designed for the potential control genes using a web based tool by Exiqon (http://oligo.lnatools.com/expression/), and repeat mask (http://www.repeatmasker.org/) was used to check for repeated sequences. Only oligonucleotides with a GC content and T_m within the range of the *nifH* probes on the array were included as potential control probes. The selected oligonucleotide probe sequences were required to have no significant self annealing or secondary structure. Using these criteria, 60-mer oligonucleotide probes were designed for six human genes from 620 to 1100 bp long (Table 2). The potential control probe and full gene sequences were blasted against the NCBI database to search for potential regions with similarity with *nifH*. The selected control sequences had no BLAST hits to *nifH* genes. The final six probes selected as controls had 52% or less similarity with each other, therefore no cross hybridization among controls was expected.

The genes for which the successful control probes were designed were purchased from the American Type Culture Collection clone collection. Plasmid DNA from *Escherichia coli* with the gene insert was purified using a Qiagen (Valencia, CA) Miniprep kit, and genes were amplified using ready-made primers for the T7 and SP6 or T3 and T7 promoter sites (IDT, Iowa City, IA), depending on the vector. Plasmid DNA Miniprep was diluted at 1:10–1:50 to serve as the PCR template, and 1 μl was added per 50 μl of reaction. Each 50 μl of PCR reaction had 5 μl of 10× buffer, 2.5 mM MgCl, 200 μM dNTPs and 2 U *Taq* polymerase (Promega, Madison, WI). Amplification conditions were 5 min 94°C initial denaturation, then 30 cycles of 30 s 94°C, 30 s 42°C and 30 s 72°C, with a final extension of 3 min at 72°C. Polymerase chain reaction products were separated on a 1.2% TAE agarose gel, excised and cleaned using the Qiagen (Valencia, CA) gel extraction kit. The products were quantified fluorometrically with a plate reader using PicoGreen detection (Molecular Probes, Eugene, OR). No binding to *nifH* probes was seen in hybridizations with the positive controls (data not shown). Control probes 1, 2 and 4 were used as positives and probes 3 and 6 served as negatives (see *Data analysis*) (Table 2). Hybridization of the PCR product for control 5 did not produce consistent results and control 5 was excluded from further analyses.

Microbial mat experiments

Microbial communities in natural microbial mats dominated by the cyanobacterium *Lyngbya* spp. were studied using the

microarray. A 7-month long nutrient addition experiment was carried out at the NASA Ames greenhouse facility, Moffett Field, CA (Bebout *et al*., 2002). To initiate the experiment, pieces of desiccated *Lyngbya* sp. dominated microbial mats were cut from a natural marsh near GN, Baja California Sur, Mexico (27°43.402′N, 113°55.322′W), on 25 April 2004 and transported to NASA Ames. The mats were placed under natural irradiance in greenhouse flumes with circulating seawater collected nearby the collection site and salinity was adjusted to 40 ppt using distilled water. Three nutrient treatments, with nutrients added weekly starting 24 June 2004, were included: control (no nutrient additions), phosphorus (P) and nitrogen (N). N was added as KNO_3 and NH_4Cl (100 μM final concentration) and P was added as KH_2PO_4 (15 μM final concentration). Each treatment consisted of two flumes with three mat pieces per flume. In order to simulate the natural conditions of periodic desiccation, the mats in the greenhouse were wetted only one to two times per week. Mat DNA and RNA samples were collected at a 4- to 6-week interval for a period of 7 months for determination of microbial community structure and function. Only the data collected at the end of the experiment (at 7 months) are discussed here. Mat samples were cored using a stainless steel custom made 10.3-mm diameter cylinder, cleaned with 70% EtOH between samples. The mat piece was cut into quadrants using a 70% EtOH-cleaned scalpel, then put in a cryovial, and immediately snap frozen in liquid nitrogen. The cores were stored at −80°C until extraction.

Target extraction, amplification and sequencing

DNA from the microbial mat samples was extracted using the method by Tillett and Neilan (2000) with modifications. A 50- to 100-mg mat piece was placed in a tube with 0.2-g sterile glass beads (0.1-g 100 μm and 0.1-g 200 μm diameter beads) and 750 μl of xanthogenate buffer. The tubes were agitated in a Fast Prep machine (Qbiogene, Morgan Irvine, CA) twice for 20 s at setting 5.5 during a 30-min to 1-h incubation at 70°C. After the incubation the tubes were kept on ice for 30 min, then centrifuged for 10 min at 12 000 *g*. The supernatant was pipetted to clean tubes and precipitated with 750 μl of isopropanol for 10 min. After centrifugation for 10 min, the supernatant was discarded and the pellet washed once with 1 ml of 70% EtOH. The pellet was air dried and finally resuspended in 100 μl of TE buffer at 70°C. The extract was purified using the Qiagen (Valencia, CA) DNeasy Plant kit following the manufacturer's recommendations. RNA was extracted using the RNAwiz reagent (Ambion, Austin, TX) following the manufacturer's protocol. The extraction was completed with cleanup using the Qiagen (Valencia, CA) RNeasy kit with on-column DNA digestion carried out twice for 1 h, and extracts were stored at −80°C. cDNA was synthesized using the Invitrogen (Carlsbad, CA) Superscript III reverse transcriptase and the inner reverse *nifH* primer (Zehr and Turner, 2001), then followed up with regular *nifH* PCR amplification. For each RT reaction, a set of controls were included to which no RT was added, to check for potential DNA contamination. Amplification of *nifH* from the mat DNA and cDNA was carried out as follows. Prior to PCR, extracted DNA was quantified fluorometrically using PicoGreen (Molecular Probes, Eugene, OR). Two μl of DNA extract at 0.2 ng μl^{-1} was added to each PCR reaction to serve as the PCR template. In RT-PCR, 2 μl of cDNA was used as a template in the PCR reaction. For amplification, a nested set of *nifH* primers was used (Zehr and Turner, 2001). For the first nested reaction, final concentrations of 2.5 mM $MgCl_2$, 200 μM dNTPs (each nucleotide), 1 μM primers nifH3 (reverse) and nifH4 (forward), and 2 U *Taq* Polymerase (Promega, Madison, WI) were included in the reaction, with the final volume adjusted to 50 μl with 5 kDa filtered nuclease free water. Amplification was carried out using 2 min at 94°C initial denaturation, then 25 cycles with 30 s at 94°C, 30 s at 57° and 1 min at 72°C, with final extension at 72°C for 7 min. One μl from the first round reaction was used as a template in the second round nested reaction. The PCR mix and conditions in the first and second round reaction were identical, except the inner nested primers nifH1 (forward) and nifH2 (reverse) (Zehr and Turner, 2001) were used in the second round reaction. The PCR products of approximately 359 bp were separated on a 1.2% TAE agarose electrophoresis gel, and bands were then excised and purified using the Qiagen (Valencia, CA) gel purification kit. The purified PCR product was used in microarray hybridizations (see *Target labelling and hybridization*) and sequencing. For sequencing, the *nifH* fragments were cloned into pGEM-T vector (Promega, Madison, WI). Plasmid Miniprep cleaning was carried out using the Montage Miniprep kit (Millipore, Billerica, MA) in a 96-well plate format. Sequencing was performed with the ABI3100 (UC Santa Cruz) or ABI 3730 (UC Berkeley) using 25–27 sequencing cycles of the following: 96°C for 10 s, 50°C for 5 s and 60°C for 4 min using BigDye v.3.1.

Target labelling and hybridization

Target for array hybridization was prepared using psoralen-biotin labelling with a Bright Star® kit (Ambion, Austin, TX) following the manufacturer's instructions and modifications of previously published protocols for macroarrays (Jenkins *et al*., 2004; Steward *et al*., 2004). Ten μl of target at 3–6 ng μl^{-1} (unless otherwise specified), 5 μl of internal, positive control mix, and 5 μl of TE buffer were combined and denatured for 12 min at 100°C, then quickly chilled on ice. For each set of arrays to be compared, the same target concentration was used in biotinylation and hybridization reactions. Psoralen-biotin (2 μl) was added, and the mixture was incubated on ice under long wavelength UV light for 1 h. The mix was diluted with 78 μl of TE buffer, extracted twice with 200 μl of butanol, and stored at −20°C. Immediately prior to hybridization, 13.2 μl of target was heat denatured at 97°C for 5 min, then spun down and mixed with hybridization buffer (at hybridization temperature) to final concentrations of 2× SSPE (20 mM sodium phosphate, 300 mM sodium chloride, 2 mM EDTA), 1% SDS (sodium dodecyl sulfate) and 30% formamide. Slides and cover slips were dusted with filtered air, and total volume of 24 μl of hybridization mix was evenly applied to the array underneath the cover slip (LifterSlip, Erie Scientific, Portsmouth, NH). Pieces of Whatman filter paper were placed inside the microarray hybridization chamber (Ambion, Austin, TX) and saturated with the hybridization buffer or water. The chambers were sealed tightly and placed in a circulating water bath at the desired temperature for 18 h. After hybridization, unbound target was removed by washing

twice at room temperature for 30 s with 4× SSPE with 0.005% Tween 20, then twice with 20 mM sodium phosphate buffer with 0.005% Tween 20 at pH 8.0, and finally once with 20 mM sodium phosphate buffer at pH 8.0. The slides were dried in a filtered air stream. For staining, streptavidin-Alexa Fluor 555 (Molecular Probes, Eugene, OR) at a concentration of 10 μg ml^{-1} in hybridization buffer was applied under a cover slip, and the slides were stained in a humidified chamber in the dark for 1 h. After staining, slides were washed twice for 10 s with 20 mM sodium phosphate buffer with 0.005% Tween 20, once with 20 mM sodium phosphate buffer, and dried. The slides were scanned at 10-μm pixel size resolution at 532 nm with a GenePix 4000B slide scanner (Axon Instruments, Sunnyvale, CA).

Data analysis

Microarray data analysis was carried out with GenePix Pro 6.0 software. The threshold for detection was determined as follows. Spots were initially considered positive if they had: (i) a signal-to-noise ratio ≥ 3, and (ii) $\geq 70\%$ of pixels with greater than two standard deviations higher than background. Furthermore, if four or more spots of total of eight replicates in each slide showed a positive signal after these threshold scripts were run, the probe was determined to be positive. When these threshold criteria were used, the negative controls were consistently zero. This suggested that there was no significant difference between the baseline signal of spots (signal that any spotted DNA would have in response to any target) and signal in the slide area between spots. Therefore, local background was considered a suitable measure for background signal, and the negative controls were not used in the analysis.

Average background corrected signal intensity for each probe (F_{probe}) or control ($F_{control}$) was calculated as follows:

$$F_{probe} \text{ or } F_{control} = [(F_{S1} - F_B) + (F_{S2} - F_B) + \ldots + (F_{Sn} - F_B)]/n$$

where F_{S1} is raw fluorescence in spot replicate 1 (median pixel intensity in spot), F_B is local background fluorescence (local median intensity), and n is number of replicate spots included.

Normalized probe signal intensity F_{PN} was calculated as follows:

$$F_{PN} = F_{probe}/F_{control}$$

Phylogenetic analyses

For neighbour-joining phylogenetic trees for the microarray sequences, sequences were aligned in ARB software using Pfam seed alignment (Bateman *et al.*, 2002) and HMMER software (http://hmmer.wustl.edu/). Minor manual adjustments were made to alignment of two sequences. Kimura and Jukes-Cantor corrections were used to create protein and DNA trees respectively. For the protein tree, a 321-bp DNA region was included, and for the microarray probe DNA tree, the full 60 bp DNA probe region was included. Both regions started directly downstream from the nifH1 primer.

Specificity

To investigate array specificity, hybridizations were carried out with twelve different cloned *nifH* gene fragments containing sequences that are 100% identical to probes on the array. Polymerase chain reaction was carried out from archived PCR products or plasmid Minipreps using *nifH* primers nifH3 and nifH4, and the PCR product concentration was adjusted to be identical in all target clones prior to biotinylation. Separate hybridization reactions at identical target concentration were carried out for each *nifH* clone at 50°C for 18 h. Additionally, the effect of DNA concentration on specificity (cross hybridization) and degree of linearity in signal intensity was tested using one *nifH* sequence (*A. cylindrica* sequence AY221813). As a second approach to address specificity, array data were compared with a sequencing effort from microbial mat samples from the nutrient addition experiment. Separate clone libraries were generated for each treatment, and each included 46–53 clones. GenBank accession numbers for the mat clone sequences are DQ337961–DQ338107.

Acknowledgements

We thank A. Morrison and M. Hogan for technical assistance. During part of this work PM held a National Research Council Research Associateship at NASA Ames Research Center. Additional funding was provided by NASA-UCSC University Affiliated Research Center Aligned Research Program, NASA Exobiology, and the Gordon and Betty Moore Foundation.

References

Abed, R.M.M., and Garcia-Pichel, F. (2001) Long-term compositional changes after transplant in a microbial mat cyanobacterial community revealed using a polyphasic approach. *Environ Microbiol* **3:** 53–62.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., *et al.* (2002) The Pfam protein families database. *Nucl Acids Res* **30:** 276–280.

Bebout, B.M., Carpenter, S.P., Des Marais, D.J., Discipulo, M., Embaye, T., Garcia-Pichel, F., *et al.* (2002) Long-term manipulations of intact microbial mat communities in a greenhouse collaboratory: simulating Earth's present and past field environments. *Astrobiology* **2:** 383–402.

Bodrossy, L., Stralis-Pavese, N., Murrell, J.C., Radajewski, S., Weilharter, A., and Sessitsch, A. (2003) Development and validation of a diagnostic microbial microarray for methanotrophs. *Environ Microbiol* **5:** 566–582.

Chien, Y.-T., and Zinder, S.H. (1994) Cloning, DNA sequencing, and characterization of a *nifD*-homologous gene grom the archaeon *Methanosarcina barkeri* 227 which resembles *nifD1* from the eubacterium *Clostridium pasteurianum*. *J Bacteriol* **176:** 6590–6598.

Des Marais, D.J., and Walter, M.R. (1999) Astrobiology: exploring the origins, evolution, and distribution of life in the universe. *Ann Rev Syst* **30:** 397–420.

El Fantroussi, S., Urakawa, H., Berhard, A.E., Kelly, J.J., Noble, P.A., Smidt, H., *et al.* (2003) Direct profiling of environmental microbial populations by thermal dissociation analysis of native rRNAs hybridized to oligonucleotide microarrays. *Appl Environ Microbiol* **69:** 2377–2382.

Guschin, D.Y., Mobarry, B.K., Proudnikov, D., Stahl, D.A., Rittman, B.E., and Mirzabekov, A.D. (1997) Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology. *Appl Environ Microbiol* **63:** 2397–2402.

Hoehler, T.M., Bebout, B.M., and Des Marais, D.J. (2001) The role of microbial mats in the production of reduced gases on the early Earth. *Nature* **6844:** 324–327.

Hughes, J.B., Hellmann, J.J., Ricketts, T.R., and Bohannan, B.J.M. (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol* **67:** 4399–4406.

Jenkins, B.D., Steward, G.F., Short, S.M., Ward, B.B., and Zehr, J.P. (2004) Fingerprinting diazotroph communities in the Chesapeake Bay by using a DNA macroarray. *Appl Environ Microbiol* **70:** 1767–1776.

Kemp, P.F., and Aller, J.Y. (2004) Estimating prokaryotic diversity: when are 16S rDNA libraries large enough? *Limnol Oceanogr: Methods* **2:** 114–125.

Mitterer, G., Huber, M., Leidinger, E., Kirisits, C., Lubitz, W., Mueller, M.W., and Schmidt, W.M. (2004) Microarray-based identification of bacteria in clinical samples by solid-phase PCR amplification of 23S ribosomal DNA sequences. *J Clin Microbiol* **42:** 1048–1057.

Nicolaisen, M., Justesen, A.F., Thrane, U., Skouboe, P., and Holmstrøm, K. (2005) An oligonucleotide microarray for the identification and differentiation of trichothecene producing and non-producing *Fusarium* species occurring on cereal grain. *J Microbiol Meth* **62:** 57–69.

Nübel, U., Garcia-Pichel, F., Kühl, M., and Muyzer, G. (1999) Quantifying microbial diversity: morphotypes, 16S rRNA genes, and carotenoids of oxygenic phototrophs in microbial mats. *Appl Environ Microbiol* **65:** 422–430.

Omoregie, E.O., Crumbliss, L.L., Bebout, B.M., and Zehr, J.P. (2004a) Comparison of diazotroph community structure in *Lyngbya* sp. & *Microcoleus chthonoplastes* dominated microbial mats from Guerrero Negro, Baja, Mexico. *FEMS Microbiol Ecol* **47:** 305–318.

Omoregie, E.O., Crumbliss, L.L., Bebout, B.M., and Zehr, J.P. (2004b) Determination of nitrogen-fixing phylotypes in *Lyngbya* sp. & *Microcoleus chthonoplastes* cyanobacterial mats from Guerrero Negro, Baja, Mexico. *Appl Environ Microbiol* **70:** 2119–2128.

Panicker, G., Call, D.R., Krug, M.J., and Bej, A.K. (2004) Detection of pathogenic *Vibrio* spp. in shellfish by using multiplex PCR and DNA microarrays. *Appl Environ Microbiol* **70:** 7436–7444.

Postgate, J.R. (1987) *Nitrogen Fixation*. London, UK: Edward Arnold.

Rhee, S.-K., Liy, X., Chong, S.C., Wan, X., and Zhou, J. (2004) Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl Environ Microbiol* **70:** 4303–4317.

Risatti, J.B., Capman, W.C., and Stahl, D.A. (1994) Community structure of a microbial mat: the phylogenetic dimension. *Proc Natl Acad Sci* **91:** 10173–10177.

Small, J., Call, D.R., Brockman, F.J., Straub, T.M., and Chandler, D.P. (2001) Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. *Appl Environ Microbiol* **67:** 4708–4716.

Steppe, T.F., and Paerl, H.W. (2002) Potential N-2 fixation by sulfate reducing bacteria in a marine intertidal microbial mat. *Aquat Microb Ecol* **28:** 1–12.

Steward, G.F., Jenkins, B.D., Ward, B.B., and Zehr, J.P. (2004) Development and testing of a DNA macroarray to assess nitrogenase (*nifH*) gene diversity. *Appl Environ Microbiol* **70:** 1455–1465.

Taroncher-Oldenburg, G., Griner, E.M., Francis, C.A., and Ward, B.B. (2003) Oligonucleotide microarray for the study of functional gene diversity in the nitrogen cycle in the environment. *Appl Environ Microbiol* **69:** 1159–1171.

Tillett, D., and Neilan, B.A. (2000) Xanthogenate nucleic acid isolation from cultured and environmental cyanobacteria. *J Phycol* **36:** 251–258.

Wang, R.-F., Beggs, M.L., Robertson, L.H., and Cerniglia, C.E. (2002) Design and evaluation of oligonucleotide-microarray method for the detection of human intestinal bacteria in fecal samples. *FEMS Microbiol Lett* **213:** 175–182.

Wilson, K.H., Wilson, W.J., Radosevich, J.L., DeSantis, T.Z., Viswanathan, V.S., Kuczmarski, T.A., and Andersen, G.L. (2002) High-density microarray of small-subunit ribosomal DNA probes. *Appl Environ Microbiol* **68:** 2535–2541.

Wu, L., Thompson, D.K., Li, G., Hurt, R.A., Tiedje, J.M., and Zhou, J. (2001) Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Appl Environ Microbiol* **67:** 5780–5790.

Zehr, J.P., Mellon, M., Braun, S., Litaker, W., Steppe, T., and Paerl, H.W. (1995) Diversity of heterotrophic nitrogen fixation genes in a marine cyanobacterial mat. *Appl Environ Microbiol* **61:** 2527–2532.

Zehr, J.P., and Turner, P.J. (2001) Nitrogen fixation: nitrogenase genes and gene expression. *Meth Microbiology* **30:** 271–286.

Zehr, J.P., Jenkins, B.D., Short, S.M., and Steward, G.F. (2003) Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environ Microbiol* **5:** 539–554.

Zehr, J.P., Church, M.J., and Moisander, P.H. (2006) Diversity, distribution and biogeochemical significance of nitrogen-fixing microorganisms in anoxic and suboxic ocean environments. In *Past and Present Water Column Anoxia.* NATO Science Series IV. Earth and Environmental Sciences, Vol. 64. Neritin, L.N. (ed.). Dordrecht, The Netherlands: Springer, pp. 337–371.

Supplementary material

The following supplementary material is available for this article online:

Table S1. Probe sequences in the *nifH* microarray.

This material is available as part of the online article from http://www.blackwell-synergy.com

