Applied and Environmental Microbiology, Mar. 2004, p. 1455–1465 Vol. 70, No. 3
0099-2240/04/$08.00+0 DOI: 10.1128/AEM.70.3.1455–1465.2004
Copyright © 2004, American Society for Microbiology. All Rights Reserved.

Development and Testing of a DNA Macroarray To Assess Nitrogenase (*nifH*) Gene Diversity

Grieg F. Steward,[1] Bethany D. Jenkins,[2] Bess B. Ward,[3] and Jonathan P. Zehr[2]*

Department of Oceanography, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Honolulu, Hawaii 96822[1]; Ocean Sciences Department and Institute of Marine Sciences, University of California, Santa Cruz, Santa Cruz, California 95064[2]; and Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544[3]

Received 26 August 2003/Accepted 1 December 2003

A DNA macroarray was developed and evaluated for its potential to distinguish variants of the dinitrogenase reductase (*nifH*) gene. Diverse *nifH* gene fragments amplified from a clone library were spotted onto nylon membranes. Amplified, biotinylated *nifH* fragments from individual clones or a natural picoplankton community were hybridized to the array and detected by chemiluminescence. A hybridization test with six individual targets mixed in equal proportions resulted in comparable relative signal intensities for the corresponding probes (standard deviation, 14%). When the targets were mixed in unequal concentrations, there was a predictable, but nonlinear, relationship between target concentration and relative signal intensity. Results implied a detection limit of roughly 13 pg of target ml^{-1}, a half-saturation of signal at 0.26 ng ml^{-1}, and a dynamic range of about 2 orders of magnitude. The threshold for cross-hybridization varied between 78 and 88% sequence identity. Hybridization patterns were reproducible with significant correlations between signal intensities of duplicate probes ($r = 0.98$, $P < 0.0001$, $n = 88$). A mixed *nifH* target amplified from a natural Chesapeake Bay water sample hybridized strongly to 6 of 88 total probes and weakly to 17 additional probes. The natural community results were well simulated ($r = 0.941$, $P < 0.0001$, $n = 88$) by hybridizing a defined mixture of six individual targets corresponding to the strongly hybridizing probes. Our results indicate that macroarray hybridization can be a highly reproducible, semiquantitative method for assessing the diversity of functional genes represented in mixed pools of PCR products amplified from the environment.

The biochemical transformations of nitrogen have a major influence over biological productivity on Earth. The many pathways of the nitrogen cycle (nitrification, denitrification, ammonification, and nitrogen fixation) form a web of redox reactions catalyzed by specific groups of microorganisms. Biological nitrogen fixation, the enzyme-catalyzed reduction of dinitrogen (N_2) to ammonium, is essential for maintaining fertility in many ecosystems (28) by making nitrogen available from the large gaseous atmospheric reservoir of N_2, which is not directly accessible by eukaryotes or by many prokaryotes. Nitrogenase, the enzyme responsible for catalysis, is found in phylogenetically diverse groups of prokaryotes. In recent years, the molecular sequences of nitrogenase genes have been used to investigate the diversity of diazotrophs in marine (36), freshwater (13), estuarine (1), salt marsh (3, 11), hypersaline (24), and terrestrial (29) environments, including specialized habitats such as termite guts (9, 16). These studies have shown that there are diverse nitrogenase genes in natural environments and suggest that communities of N_2-fixing microorganisms differ markedly among habitats (33).

Investigations of nitrogenase genes in the environment initially used PCR amplification with subsequent cloning and sequencing to catalog sequence diversity (17, 27, 34, 35). Analysis of clone libraries by exhaustive sequencing can be time-consuming and expensive, which severely restricts the number of samples that can be analyzed. Given the extensive nitrogenase sequence diversity being discovered in each new environment investigated (33), it is clear that quantitative comparisons among large numbers of samples will be necessary to understand the spatial and temporal variability of complex communities of diazotrophs and how this variability relates to ecosystem function.

To improve analytical throughput, a number of fingerprinting methods, such as denaturing gradient gel electrophoresis (DGGE) (14) and terminal restriction fragment length polymorphism (TRFLP) (10), have been adapted to the study of nitrogenase gene diversity (15, 18). These methods have shown promise, but the fingerprints provide either no (DGGE) or sometimes ambiguous (TRFLP) information about the phylogenetic composition of a community. In DGGE, different fragments are discriminated only by relative duplex stability, and in TRFLP, discrimination relies on a small subset of the information content of a sequence, namely, the location of specific restriction sites. An appealing feature of hybridization-based methods, such as DNA arrays, is that discrimination among sequences is primarily a function of their overall similarity to probes of known sequence. Initial applications of microarrays to the investigation of microbial diversity targeted 16S rRNA (7, 12, 23, 30), but microarrays are also being developed for assessing functional gene diversity (26, 31) and expression (6).

The goal of this study was to design simple, inexpensive DNA macroarrays that would facilitate the examination of nitrogenase gene diversity in the environment, in particular (*nifH*) genes in aquatic water samples. The arrays were designed to reflect the diversity of the *nifH* gene sequences we

* Corresponding author. Mailing address: Ocean Sciences Department and Institute of Marine Sciences, University of California, Santa Cruz, Santa Cruz, CA 95064. Phone: (831) 459-4009. Fax: (831) 459-4882. E-mail: zehrj@cats.ucsc.edu.

TABLE 1. Station names, locations, and sampling depths for water collected in the Choptank River and Chesapeake Bay for DNA extraction

Station[a] (date [yr/mo/day])	Location (latitude, longitude)	Sampling depth (m)
CT100 (2000/7/11)	38° 48′ N, 75° 55′ W	1
CT200 (2000/7/11)	38° 37′ N, 76° 08′ W	1
		7.9
CT100 (2001/4/3)	38° 48′ N, 75° 55′ W	1
CB100 (2001/4/4)	39° 21′N, 76° 11′ W	1.8
		7.7
		9.7
CB200 (2001/4/5)	38° 34′ N, 76° 27′ W	1.8
		11.2
		17.6
CB300 (2001/4/6)	37° 18′ N, 76° 09′ W	1.9
		8.3
		11.3

[a] Choptank River and Chesapeake Bay stations are indicated by CT and CB prefixes, respectively.

obtained from clone libraries generated by PCR amplification of genomic DNA samples obtained from the Choptank River and the Chesapeake Bay. We also included *nifH* genes from diverse reference microorganisms of known phylogenetic affiliation. In this report, we present an evaluation of the performance of the array with targets derived from individual sequences, defined mixtures of sequences, and a natural community. Our results indicate that the macroarray procedure is a highly reproducible and at least semiquantitative method for assessing the genetic diversity within pools of mixed PCR products.

MATERIALS AND METHODS

DNA extraction. (i) Culture collection isolates. Cultured diazotrophs were obtained from the Pasteur Culture Collection of Cyanobacteria (PCC, Paris, France), the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ, Braunschweig, Germany), the Agriculture Research Services culture collection at the National Center for Agricultural Utilization Research (NRRL, Peoria, Ill.), and the Culture Collection of Algae at the University of Texas at Austin. Whenever feasible, microorganisms were grown in the appropriate medium prior to extraction of DNA. For microorganisms that we were not equipped to cultivate, the entire inoculum received from the supplier was extracted. For liquid cultures, an aliquot of 3 to 8 ml was centrifuged, and the resulting cell pellets were washed once with STE buffer (20% sucrose, 50 mM Tris-HCl, 50 mM EDTA), resuspended in 250 μl of STE buffer containing 5 mg of lysozyme ml^{-1}, and incubated at 25°C for 30 min. Proteinase K (final concentration, 2 mg ml^{-1}) and sodium dodecyl sulfate (SDS) (final concentration, 1%) were added to the lysate, and the resulting solutions were incubated for an additional 30 min at 65°C. Acid-washed, 0.1-mm-diameter beads were added (0.25 g per sample) with 250 μl of PCI (phenol, chloroform, isoamyl alcohol [25:24:1]). Tubes were shaken in a FastPrep bead-beating instrument (Q-Biogene, Carlsbad, Calif.) two times for 10 s each time at 6 m s^{-1} and then centrifuged to separate the phases. The aqueous phase was transferred to a new tube, and the organic phase was back extracted with 100 μl of TE (10 mM Tris-HCl [pH 8.0], 1 mM EDTA), and the second aqueous phase was pooled with the first aqueous phase. Samples were extracted once more with PCI and then extracted with chloroform. DNA was precipitated in 0.2 M NaCl and 2 volumes of ethanol overnight at −20°C (22). Samples were centrifuged for 30 min at 15,000 × *g* at 4°C. Pellets were washed with 70% ethanol and resuspended in 50 μl of TE.

(ii) Chesapeake Bay environmental samples. Water samples collected from the Choptank River in August 2000 and from the Chesapeake Bay and Choptank River in April 2001 (Table 1) were filtered through 2-μm Sterivex capsules (Millipore, Billerica, Mass.) using a peristaltic pump. The volume filtered per capsule ranged from 280 to 2,000 ml of water. Residual water was evacuated from the capsule by applying air pressure with a hand-operated syringe. Filter capsules were immediately placed in liquid nitrogen, then shipped on dry ice to the laboratory, and stored at −80°C prior to extraction. STE buffer (1.8 ml) containing 5 mg of lysozyme ml^{-1} was added to the filter. Samples were incubated at 25°C for 1 h, proteinase K (final concentration, 2 mg ml^{-1}) and SDS (final concentration, 1%) were added to the samples, and the samples were incubated at 60°C for 1 h. Lysate was aspirated from the filter capsule with a syringe and transferred to a Corex tube (Corning, Acton, Mass.). Filter capsules were rinsed once with 1 ml of TE, and the rinse was pooled with the corresponding sample in the Corex tube. RNase cocktail (Ambion, Austin, Texas) (final concentration, 2.5 U ml^{-1}) was added, and samples were incubated for 10 min at room temperature. Ammonium acetate was added to a final concentration of 2 M, and samples were centrifuged at 14,800 × *g* for 6 min to precipitate proteins. Supernatant was transferred to a new tube, and ethanol (2 volumes) was added. Nucleic acids were precipitated at −20°C for 30 min and pelleted by centrifugation at 12,000 × *g* for 35 min at 4°C. Pellets were washed with 70% ethanol and resuspended in 500 μl of TE. Samples were extracted first with an equal volume of PCI and then with an equal volume of chloroform and precipitated in ethanol and 0.3 M sodium acetate. Samples were resuspended in 10 μl of TE, 180 μl of Qiagen (Valencia, Calif.) ATL buffer, and 200 μl of Qiagen AL buffer and further purified using the DNeasy minikit (Qiagen) following the manufacturer's instructions.

PCR amplification of *nifH* genes from genomic DNA extracts. A fragment of the *nifH* gene (318 to 333 bp) was amplified by nested PCR with degenerate primers as described previously (32, 37), but with a new set of outer primers (24). Each 50-μl first-round PCR mixture contained 5 μl of 10× ExTaq buffer (Takara, Shuzo, Japan), 4 μl of 2.5 mM deoxynucleoside triphosphates, 37.5 μl of water, 0.5 μl of 100 μM CDHPnif53F (5′-TGAGACAGATAGCTATYTAYGGHAA-3′), 0.5 μl of 100 μM CDHPnif723R (5′-GATGTTCGCGCGGCACGAADTRNATSA-3′), 0.5 μl of ExTaq DNA polymerase (5 U μl^{-1}), and 2 μl of template DNA. Thermal cycling consisted of the following steps: (i) an initial denaturation at 94°C for 5 min; (ii) 30 cycles, with 1 cycle consisting of denaturation at 94°C for 1 min, annealing at 50°C for 1 min, and extension at 72°C for 1 min; and (iii) a final 7-min extension at 72°C. For the second round of the nested amplification, one microliter of this reaction mixture was used as the template in a 50-μl reaction mixture containing the same reagent mixture described above, but with 38.5 μl of water and the inner primers nifH2 (5′-TGYGAYCCNAARGCNGA-3′) and nifH1 (5′-ADNGCCATCATYTCNCC-3′) (34). The thermal cycling protocol for the nested reactions was the same as above except that the annealing temperature was raised to 57°C. PCR products from the nested reaction were gel purified. A single dominant band of the expected size was observed in most cases using these conditions. Exceptions were amplifications for two isolates (*Enterobacter aerogenes* and *Methanothermus fervidus*) that yielded many nonspecific products of various sizes. Gel-purified fragments were quantified by fluorometry using the PicoGreen double-stranded DNA quantitation kit (Molecular Probes, Eugene, Oreg.) and a TD700 filter-based fluorometer (Turner Designs, Sunnyvale, Calif.) or a Cary spectrofluorometer (Varian, Palo Alto, Calif.).

Cloning and sequencing of amplified *nifH* gene fragments. PCR products were cloned into a pGEM T-system II vector (Promega, Madison, Wis.) following the manufacturer's instructions. Plasmids were purified using a mini-spin plasmid purification kit (Qiagen), sequenced using BigDye version 3 cycle sequencing chemistry (Applied Biosystems, Foster City, Calif.) with T7 and Sp6 primers, and analyzed with ABI PRISM 310 or 3100 genetic analyzer.

Selection of clones for array construction. All of the sequences obtained in this study were combined into a database with other nitrogenase sequences retrieved from GenBank and used to build a distance tree using the neighbor-joining algorithm in order to identify representative clones for macroarray construction (Fig. 1). Sixty-five clones from the library of Chesapeake Bay and Choptank River clones were identified that represented the range of diversity found in those environments. Clones were chosen such that, on average, each shared 87% sequence identity with the next nearest selected clone (range, 73 to 96%). An additional three clones from a previous investigation in the Pacific Ocean (38) were also selected. A subset of the selected clones, primarily clones from cultivated isolates, was used to prepare an initial array (hereafter referred to as the test array) with 21 probes and 6 control DNA spots. A second larger array called the Chesapeake Bay version 1 (CBv1) array, was prepared with the 65 selected Chesapeake environmental clones, 2 Pacific Ocean environmental clones, clones from 21 cultivated microorganisms, and 8 control DNA spots (Table 2).

PCR amplification of cloned *nifH* gene fragments for array probes. For the test array, cloned nitrogenase gene fragments were amplified by nested PCR from plasmid minipreps for spotting as array elements (i.e., probes). First-round PCR mixtures contained the same constituents as described above for *nifH*

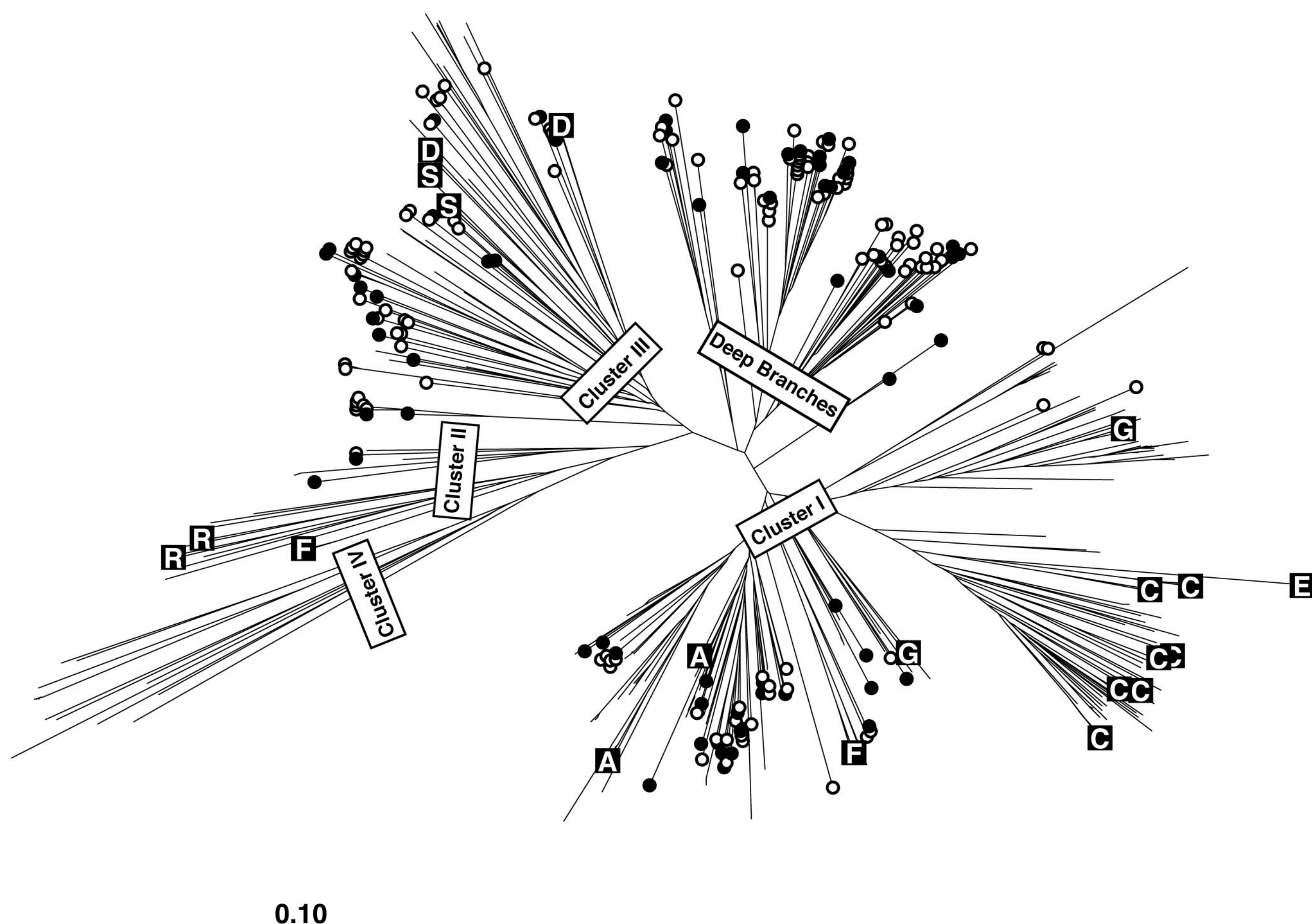

FIG. 1. Neighbor-joining tree constructed from an uncorrected distance matrix among 447 *nifH* sequence fragments generated as part of this study (228 fragments) or retrieved from GenBank (219 fragments). For GenBank sequences, only those sequences from cultivated bacteria that spanned at least the region corresponding to our fragment were included. Sequences belonging to commonly recognized major clusters (clusters I to IV [5]) and some additional deep branches are labeled. Branches with open or closed circles contain sequences derived from Chesapeake Bay samples. Closed circles indicate those clones represented as probes on the array. Branches marked with a letter in a black box are sequences obtained from cultivated bacteria as part of this study that are also represented as probes on the array. The boxed letters indicate the phylogenetic designation of the bacterium from which the sequence was derived (A, D, E, and G indicate members of the alpha, delta, epsilon, and gamma subdivisions of the *Proteobacteria*, respectively; C, cyanobacteria; F, *Firmicutes*; S, green sulfur bacteria, R, *Archaea*).

amplification but with Sp6 and T7 (0.5 μM each) as the primers and 1 to 2 μl of miniprep plasmid DNA (ca. 9 to ≤200 ng) as the template. Thermal cycling conditions consisted of an initial denaturation at 96°C for 2 min, followed by 30 cycles, with 1 cycle consisting of denaturation at 96°C for 30 s, annealing at 50°C for 30 s, and extension at 72°C for 1 min. To increase the amount of product, a second-round, nested amplification was performed using nifH1 and nifH2 primers as described above. For the CBv1 array, a modified cycling program provided sufficient product with only a single round of amplification using the Sp6 and T7 primers. The modified program consisted of the following: (i) an initial denaturation at 94°C for 5 min; (ii) 30 cycles, with 1 cycle consisting of denaturation at 94°C, annealing at 42°C, and extension at 72°C (30 s at each temperature); and (iii) a final extension at 72°C for 3 min. Amplified DNA was gel purified using a QIAquick gel extraction kit (Qiagen). Concentration of the gel-purified DNA was determined using the PicoGreen assay as described above.

Array construction. Concentrations of the PCR-amplified, gel-purified *nifH* fragments from selected individual clones were adjusted to equal concentrations for spotting as probes on macroarrays. The concentration used for the test array was 5 ng μl^{-1}. Subsequent tests showed no difference in signal with lower concentrations of probe (data not shown), so probes for the CBv1 array were normalized to 2.5 ng μl^{-1}. Biotinylated phage lambda DNA was directly spotted as a dilution series onto arrays to serve as a positive control for signal detection. The control DNA was prepared by serial dilution of biotinylated lambda DNA (see biotinylation procedure below) in a diluent of 1× TE containing unlabeled lambda DNA such that the total DNA in each standard was always 5 ng (test array) or 2.5 ng (CBv1 array). DNA probes and controls were transferred to the wells of a 96-well, flat-bottom plate and replica spotted onto SuPerCharge nylon membranes (Schleicher & Schuell, Keene, N.H.) using a hand-operated, 96-pin tool with slot pins designed to deliver 1 μl per pin (V&P Scientific, San Diego, Calif.). The pin tool was pretreated with surfactant per the manufacturer's instructions to improve the reliability of spot delivery. The spotted DNA was denatured by incubating the membrane face up for 10 min on sheets of Whatman paper soaked with denaturation solution (3 M NaCl and 0.4 N NaOH). Membranes were neutralized by dipping them in two baths of 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and a final bath of 1× SSC. Membranes were exposed face down on a UV transilluminator for 2 min (irradiation time determined empirically; data not shown). Air-dried membranes were stored at room temperature between sheets of filter paper.

Biotinylation of target and control DNA. Target DNA for hybridization to arrays was prepared by amplifying, gel purifying, and quantifying the *nifH* gene fragment from isolates or environmental genomic DNA as described above. Target DNA and phage lambda control DNA were biotinylated using the BrightStar psoralen-biotin kit (Ambion) according to the manufacturer's instructions. Briefly, psoralen-biotin reagent was added to heat-denatured DNA (20 to 50 ng μl^{-1} in 1× TE), and the mixture was irradiated with a long-wavelength (365-

TABLE 2. Sources of the probes printed on the CBv1 *nifH* macroarray[a]

Cluster or branch	Array position[b]	Isolate or clone[c]	GenBank accession no.	Phylogenetic affiliation
Cluster I	A2	*Arcobacter nitrofigilis* (DSM 7299)	AY221825	ε *Proteobacteria*[d]
	B2	*Synechocystis* sp. (WH001)	AY221818	Cyanobacteria
	C2	*Synechocystis* sp. (WH8501)	AY221821	Cyanobacteria
	D2	Station ALOHA clone HT1903	AF299422	
	E2	Station ALOHA clone HT1902	AF299420	
	F2	*Anabaena cylindrica* (UTEX 629)	AY221813	Cyanobacteria
	G2	*Tolypothrix* sp. (PCC 7101)	AY221817	Cyanobacteria
	H2	*Nostoc muscorum* (UTEX 486)	AY221814	Cyanobacteria
	A3	*Oscillatoria sancta* (PCC 7515)	AY221815	Cyanobacteria
	B3	*Symploca* sp. (PCC 8002)	AY221816	Cyanobacteria
	C3	*Vibrio diazotrophicus* (DSM 2604)	AY221828	γ *Proteobacteria*
	D3	*Paenibacillus azotofixans* (DSM 5976)	AY221826	*Firmicutes*
	E3	CB894H2	AY223945	
	F3	CB909H9	AY224005	
	G3	*Klebsiella oxytoca* (NRRL B-199)	AY221827	γ *Proteobacteria*
	H3	CB909H5	AY224002	
	A4	*Azotobacter chroococcum* (NRRL B-14637)	AY351672	γ *Proteobacteria*
	B4	CB894H8	AY223950	
	C4	*Frankia* sp. (DSM 43829)	AY351671	Actinobacteria
	D4	CB911H3	AY224015	
	E4	CB891H1	AY223907	
	F4	CB891H4	AY223912	
	G4	CB894H10	AY223944	
	H4	CB914H1	AY224026	
	A5	CB916H1	AY224034	
	B5	CB895H8	AY223957	
	C5	CB891H8	AY223911	
	D5	CB907H2	AY223968	
	E5	CB907H9	AY223997	
	F5	*Xanthobacter flavus* (NRRL B-14838)	AY221812	α *Proteobacteria*
	G5	CB891H3	AY223908	
	H5	CB921H3	AY224040	
	A6	CB921H7	AY224042	
	B6	CB909H7	AY224003	
	C6	CB895H7	AY223956	
	D6	CB914H5	AY224029	
	E6	*Sinorhizobium meliloti* (NRRL L-45)	AY221811	α *Proteobacteria*
	F6	CB908H5	AY223998	
Deep branches				
	G6	CB916H3	AY224036	
	H6	CB916H7	AY224037	
	A7	CB910H3	AY224008	
	B7	CB910H8	AY224012	
	C7	CB910H4	AY224009	
	D7	CB910H6	AY224011	
	E7	CB907H5	AY223979	
	F7	CB912H4	AY224020	
	G7	CB910H10	AY224006	
	H7	CB911H10	AY224013	
	A8	CB909H8	AY224004	
	B8	CB911H1	AY224014	
	C8	CB914H6	AY224030	
	D8	CB907H10	AY223960	
	E8	CB907H4	AY223976	
	F8	CB894H4	AY223946	
	G8	CB895H1	AY223952	
	H8	CB921H1	AY224039	
	A9	CB891H5	AY223910	
	B9	CB894H6	AY223948	
	C9	CB895H2	AY223953	
	D9	CB895H4	AY223955	
	E9	CB914H7	AY224031	
	F9	CB895H9	AY223958	
	G9	CB895H10	AY223951	
	H9	CB894H5	AY223947	
	A10	CB895H3	AY223954	
Cluster II	B10	*Methanothermobacter thermoautotrophicus* (DSM 1850)	AY221829	*Archaea*
	C10	*Methanococcus vannielli* (DSM 1224)	AY221830	*Archaea*
	D10	CB921H10	AY224038	
	E10	CB921H8	AY224043	

Continued on facing page

TABLE 2—*Continued*

Cluster or branch	Array position[b]	Isolate or clone[c]	GenBank accession no.	Phylogenetic affiliation
Cluster III	F10	CB911H9	AY224019	
	G10	CB912H2	AY224021	
	H10	CB911H7	AY224018	
	A11	CB910H5	AY224010	
	B11	CB914H8	AY224032	
	C11	CB909H1	AY224000	
	D11	CB914H3	AY224028	
	E11	*Desulfobacter latus* (DSM 3381)	AY221822	δ *Proteobacteria*
	F11	*Desulfotomaculum nigrificans* (DSM 574)	AY221823	δ *Proteobacteria*
	G11	CB912H9	AY224025	
	H11	CB921H4	AY224041	
	A12	CB916H2	AY224035	
	B12	CB907H6	AY223986	
	C12	CB912H5	AY224023	
	D12	*Chlorobium limicola* (DSM 245)	AY221831	Green sulfur bacteria
	E12	CB909H2	AY224001	
	F12	*Pelodictyon luteolum* (DSM 273)	AY221832	Green sulfur bacteria
	G12	CB910H2	AY224007	
	H12	CB912H1	AY224020	

[a] Sources of the probes printed on the CBv1 *nifH* macroarray listed in order of array position along with GenBank sequence accession numbers and the phylogenetic affiliation of cultivated representatives.
[b] Array positions as shown in Fig. 3.
[c] Isolates are listed by their binomial designation followed by the specific culture collection designation in parentheses. All environmental clones from this study are shown with a CB prefix.
[d] ε *Proteobacteria*, epsilon subdivision of the *Proteobacteria.*

nm-wavelength) UV lamp for 45 min on ice while shielded from room light. The mixture was diluted to 100 μl with 1× TE, extracted twice with water-saturated butanol, and stored at −20°C.

Array hybridization and signal detection. Membranes were incubated in hybridization buffer (1 mM EDTA, 6% SDS, 0.25 M sodium phosphate [pH 7.2], 0 to 40% formamide) at 65°C (test array) or 60°C (CBv1 array) for 1 to 2 h. The buffer was drained and replaced with 2 to 5 ml of hybridization buffer containing labeled, heat-denatured (99°C, 10 min) target at a final concentration of 25 ng ml^{-1} (test array) or 10 ng ml^{-1} (CBv1 array) (ca. 0.1 nM). Hybridization was performed in heat-sealed plastic pouches overnight (8 to 12 h) at 65°C (test array) or 60°C (CBv1 array). Hybridization buffer was drained, and the blots were transferred to trays and washed. The blots were washed twice for 5 min each time in 2× SSC plus 1% SDS at 65°C, twice for 15 min each time in 0.1× SSC plus 1% SDS at 65°C, and twice for 5 min each time in 1× SSC at room temperature. Bound biotinylated targets and biotinylated control DNA were detected using the Southern-*Star* chemiluminescence detection system (Applied Biosystems). Blots were incubated twice for 5 min each time in blocking buffer (1× phosphate-buffered saline, 0.2% I-Block reagent, 0.5% SDS) and then incubated for 20 to 30 min with a streptavidin-alkaline phosphatase conjugate (AvidX-AP; Applied Biosystems) diluted 1:5,000 (test array) or 1:2,000 (CBv1 array) in blocking buffer. Blots were washed four times in washing buffer (Applied Biosystems) and twice in assay buffer (Applied Biosystems), drained, and placed face up on a plastic sheet. Membranes were covered with a thin layer of CDP-*Star* chemiluminescence reagent (Applied Biosystems) and incubated for 5 to 30 min. Excess reagent was drained, and membranes were sandwiched between plastic sheets. Signal was recorded by exposure of the membranes to X-ray film (*Biomax* Light; Kodak, Rochester, N.Y.). Film images were scanned, and signal intensities for spots were extracted from the scanned images using ImageQuant software package (Amersham Biosciences, Sunnyvale, Calif.).

Nucleotide sequence accession numbers. Sequences of *nifH* fragments from cultivated isolates were submitted to GenBank and assigned accession numbers AY351671, AY351672, and AY221811 to AY221832. Sequences from the Chesapeake Bay and Choptank River *nifH* clone libraries were submitted to GenBank and assigned accession numbers AY223907 to AY224045.

RESULTS

Cross-hybridization experiments with the test array. Single-target hybridization to the test array with the *nifH* fragment from the cyanobacterium *Nostoc muscorum* showed that target cross-hybridization varied as a function of percent similarity to the probe and formamide concentration (Fig. 2). At 10% formamide, the target cross-hybridized with probes with >67% sequence identity. Cross-hybridization occurred primarily with other cyanobacterial probes, but some cross-hybridization was detected with a probe from a member of the gamma subdivision of *Proteobacteria* (gamma proteobacterium) (*Vibrio diazotrophicus*). At 20% formamide, the similarity required for cross-hybridization increased to ≥71%. At 40% formamide, self-hybridization was still strong, but hybridization to the next most similar probe (*Anabaena cylindrica*, 84% identity) was not detected even with prolonged exposures of up to 2 min (data not shown). Three additional hybridizations were performed with single targets prepared from the alpha proteobacterium *Sinorhizobium meliloti*, the cyanobacterium *Synechocystis* (*Crocosphaera*) sp. strain WH8501 (38), and the green sulfur bacterium *Chlorobium limicola*. For each experiment, the relative signal intensity for each probe was quantified and plotted against percent sequence identity shared with the target (Fig. 2). Considering the results for all four targets at the highest stringency (40% formamide) together, no cross-hybridization was detected for any probe sharing ≤79% identity with the target. Conversely, there were no cases where a probe sharing ≥86% identity with the target failed to cross-hybridize.

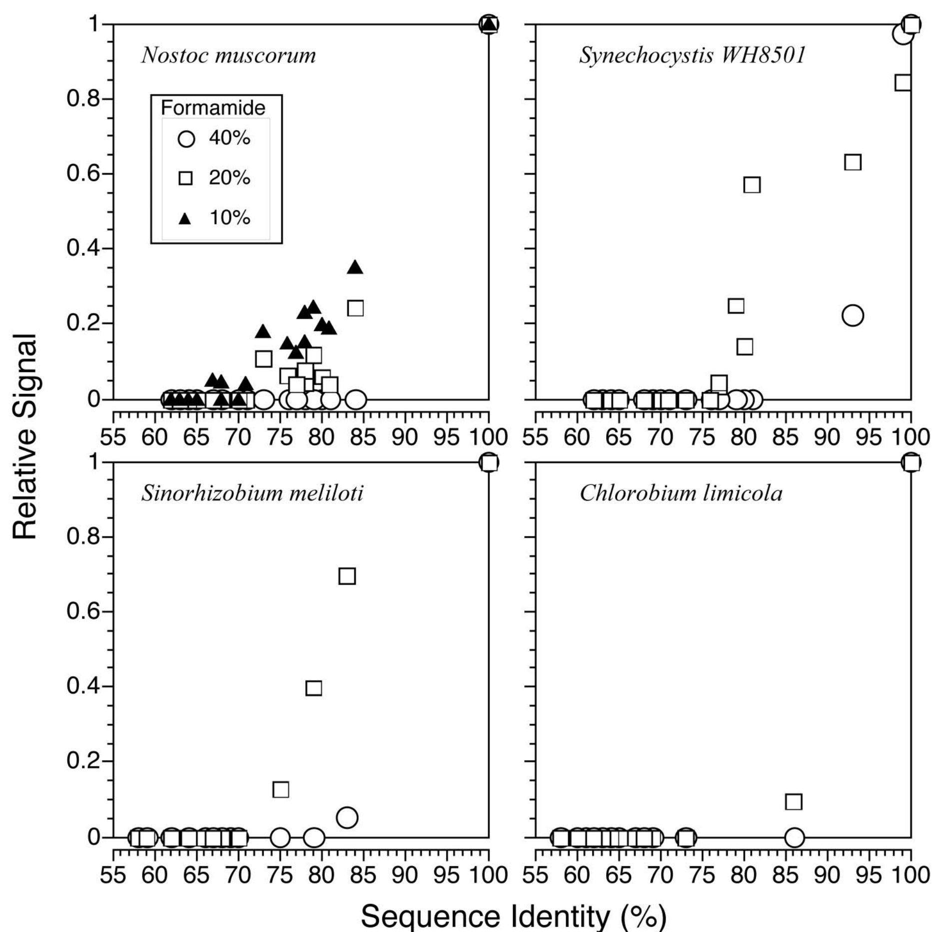

FIG. 2. Cross-hybridization experiments using four targets hybridized individually at different stringencies to test arrays composed of 21 diverse *nifH* probes. Relative hybridization signal is plotted against percent sequence identity between probe and target. The source organisms for the targets are shown in each graph. Symbols indicating formamide concentrations are the same for all panels. The 10% formamide treatment was performed only with the *Nostoc muscorum* target.

Experiments with the CBv1 array. A cross-hybridization test was performed with a single target (*S. meliloti*) on a copy of the Chesapeake Bay macroarray printed with duplicate spots (data not shown). Self-hybridization was detected to the duplicated *S. meliloti* probe spots, but no cross-hybridization to the next closest probe represented on the array (*Xanthobacter flavus*, 83% identity) was detected.

A mixed *nifH* amplification product from a DNA sample collected from Chesapeake Bay Station CB200 (38° 34′ N, 76° 27′ W) at a depth of 11.2 m was hybridized to the CBv1 array (Fig. 3A and B). Probe CB914H3 (position D11) in cluster III yielded the highest signal, and five additional probes (four deep branching and one in cluster I) had signals >25% of that maximum. The most similar sequences to these six probes (excluding other clones generated in this study) were determined by BLAST analysis (2) using the GenBank database (Table 3). The probe with the highest signal showed 91% sequence identity with a clone derived from the Neuse River estuary. Two of the four positive, deeply branching probes (CB916H3 and CB910H3) were most similar (89 and 88%) to a clone derived from bacteria associated with decaying sea grass, and a third probe was most similar (89%) to a clone derived from soil. The fourth probe was most similar (96%) to an alpha proteobacterium (*Bradyrhizobium* sp.), but the match extended over less than two-thirds of the fragment (194 of 324 bp). The sixth positive probe from cluster I was most similar (94%) to an uncultivated bacterium associated with a sweet potato plant but was also similarly related (94%) to *nifH* genes from an alpha proteobacterium (*Bradyrhizobium* sp.).

An independent hybridization of the *nifH* amplified target from the sample collected from CB200 at a depth of 11.6 m was performed using a second array that was printed with duplicate, slightly offset spots (autoradiograph not shown). Relative signal intensities from corresponding probes displayed significant ($P < 0.0001$, $n = 88$) positive correlations whether considering duplicate spots on one array (Fig. 4A, $r = 0.984$) or corresponding probes on duplicate arrays hybridized independently (Fig. 4B, $r = 0.975$).

Two defined target mixtures (10 ng ml^{-1}) were prepared by mixing amplification products from six clones that yielded the highest hybridization signals in the natural community hybridizations (above). For target mixture A (ratio of 1:1:1:1:1:1 [Fig. 3C]), the signal from each perfect-match probe was normalized to the mean value. The observed relative signal ratios ranged

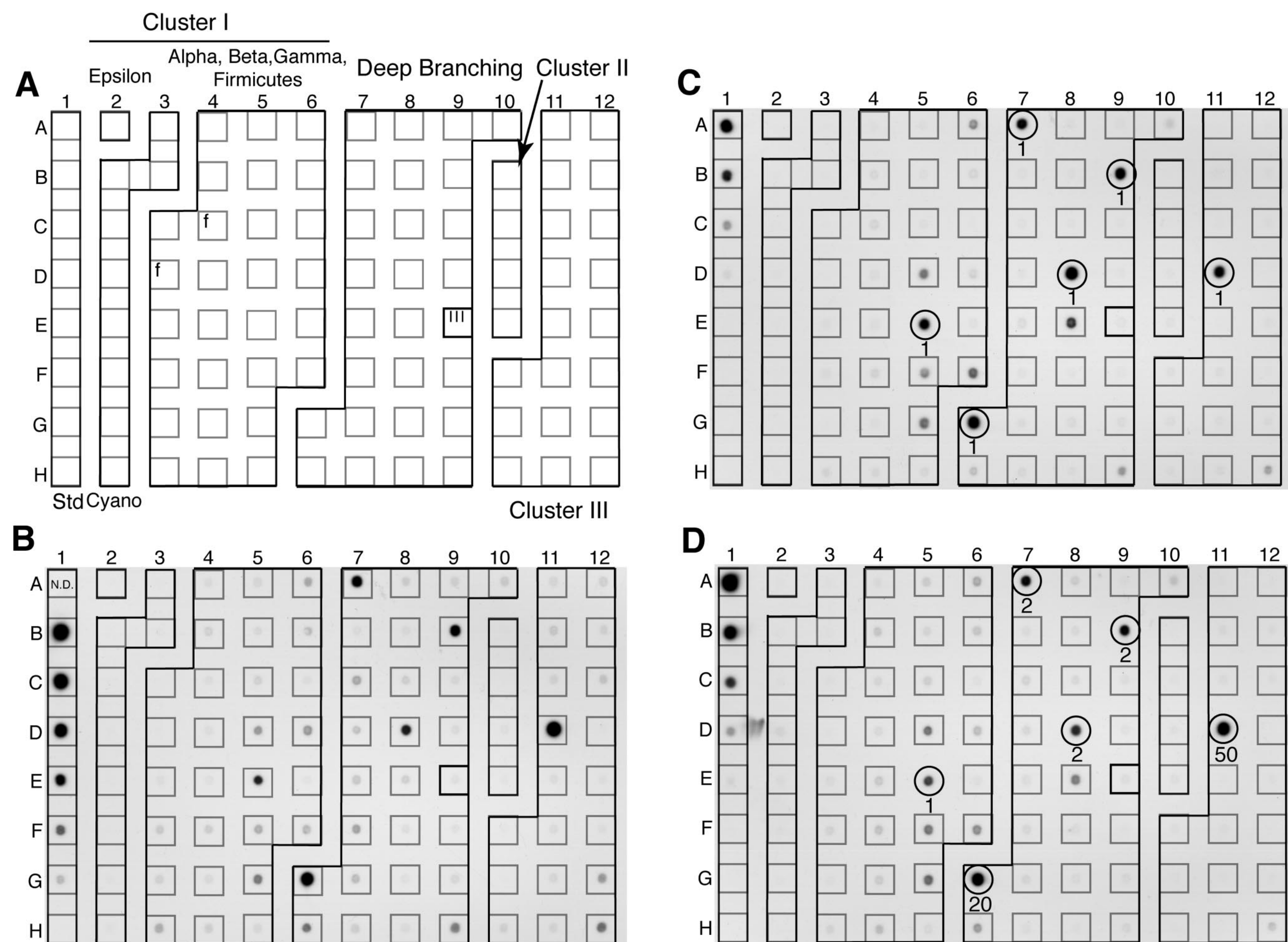

FIG. 3. Map illustrating the arrangement of probe types on the CBv1 *nifH* array (A) and autoradiographs showing chemiluminescence signal resulting from hybridizations with natural (B) or simulated (C and D) community targets. Cluster I includes members of the epsilon, alpha, beta, and gamma subdivisions of the *Proteobacteria* and cyanobacteria (Cyano). In panel A, the two sequences from *Firmicutes* (f) and a misplaced cluster III probe (III) are indicated. The template grid is laid over each autoradiograph to aid in identification of corresponding probes among arrays. The natural community target was prepared by direct PCR amplification from a community genomic DNA sample from central Chesapeake Bay. The simulated community targets were prepared by mixing amplification products from six individual *nifH* clones. Clones were chosen to correspond with the probes yielding the highest signal in the natural community hybridization. Probes that are perfect matches to the six targets are circled, and the relative concentrations of the targets within each experiment are indicated under the circled probe. Column 1 in each case contains a dilution series of biotinylated lambda DNA in a background of unlabeled lambda DNA as a positive control. The amounts of biotinylated DNA in the standards (Std) are (from top to bottom) no data (spot misprint), 0.83, 0.28, 0.093, 0.031, 0.010, 0.0034, and 0 ng (panel B) or 2.5, 0.62, 0.16, 0.039, 0.0098, 0.0024, 0.00061, and 0 ng (panels C and D).

TABLE 3. Summary information for the six probes yielding the highest signals when the CBv1 array was hybridized with a target derived from a natural Chesapeake Bay picoplankton community

Clone (probe location[a])	Signal (% of peak)	Closest match (GenBank accession no.)[b]	% Similarity[c]
CB914H3 (D11)	100	Uncultivated; clone NRS1C607 from Neuse River Estuary (AF518561)	91 (323/327)
CB916H3 (G6)	74	Uncultivated; clone SIS2-7 from decaying sea grass (AF389736)	89 (321/327)
CB894H6 (B9)	50	Uncultivated; clone g1-HW4 from soil (AY196408)	89 (327/327)
CB910H3 (A7)	48	Uncultivated; clone SIS2-7 from decaying sea grass (AF389736)	88 (320/327)
CB907H10 (D8)	30	*Bradyrhizobium* sp. strain IRBG 230 (AB079617)	96 (194/324)
CB907H9 (E5)	26	Uncultivated; clone nifH51 from sweet potato (AY159593)	94 (324/324)

[a] Array location (as in Fig. 3).
[b] The most similar sequence in GenBank as determined by the BLAST algorithm (2).
[c] Similarities between the probe and the top-ranked BLAST match are presented as a percentage followed, in parentheses, by the ratio of the length of the matching region to the total length of the submitted fragment.

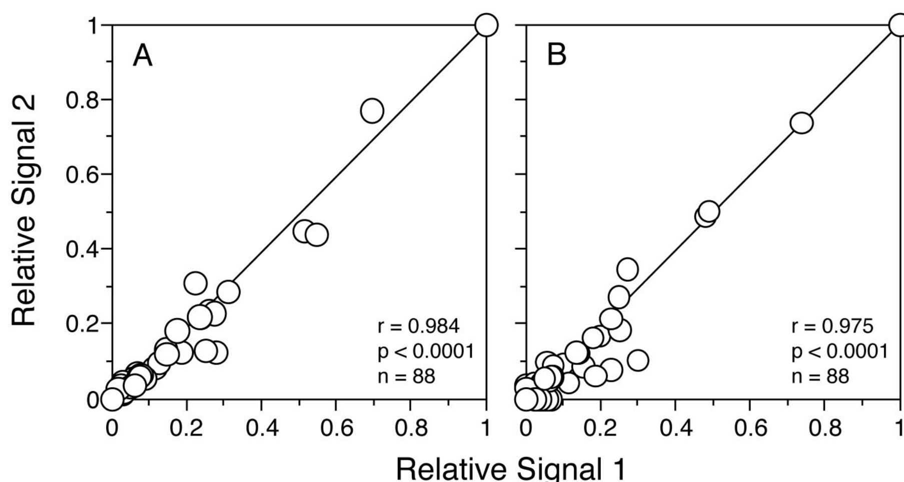

FIG. 4. Correlation of relative signal intensities between duplicate probes spotted on a single array (A) and between corresponding probes on duplicate arrays hybridized independently. The same natural community target was used in each hybridization. The plotted lines represent a 1:1 relationship.

from 0.8 to 1.54, with a mean and standard deviation of 1 ± 0.18 or an error of 14%, and the results were similar for duplicate experiments (Fig. 5). Signal intensities for all probes with a signal detected in at least one of the experiments were significantly correlated ($r = 0.96$, $P < 0.0001$, $n = 15$). These reproducible differences in relative signal intensity among the six probes were not attributable to differences in G+C content (data not shown). A single experiment with target mixture B (ratio of 50:20:2:2:2:1 [Fig. 3D]) indicated that the relationship between relative signal intensity for the perfect-match probes and the relative target concentration was nonlinear, with signal saturating at higher target concentrations (Fig. 6). Cross-hybridization to non-perfect-match probes in these two experiments varied as a function of percent identity to the next most similar target (Fig. 7). No signal was detected for any probe sharing <78% identity with one of the targets ($n = 64$). Conversely, a signal was detected for all probes with >88% identity to at least one of the six targets ($n = 15$).

The target ratios in mixture B were chosen to roughly mimic the relative signal intensity ratios obtained with the natural community hybridization, at least in rank order. Quantitative comparison of relative probe intensities resulting from the hybridizations with natural (Fig. 3B) and artificial (Fig. 3D)

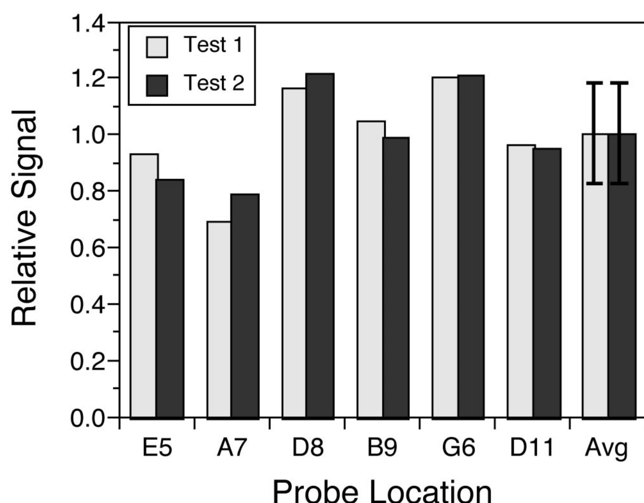

FIG. 5. Relative signal intensity for the six perfect-match probes corresponding to six targets mixed in equal ratios and hybridized to the CBv1 array. Signal intensities were normalized to the average intensity. Error bars for the average (Avg) intensity are the standard deviation among the six probes.

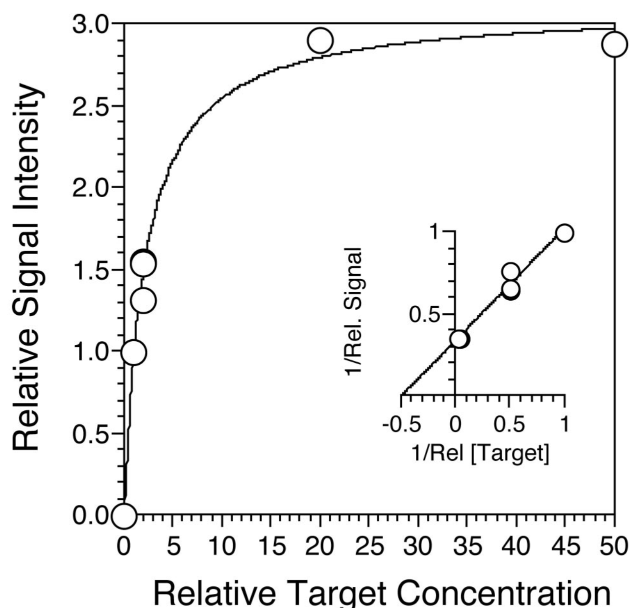

FIG. 6. Relative signal intensity for perfect-match probes as a function of relative concentration of the corresponding six targets composing an artificial community. The insert shows data presented as a double-reciprocal plot with a line fit by linear regression (Rel, relative).

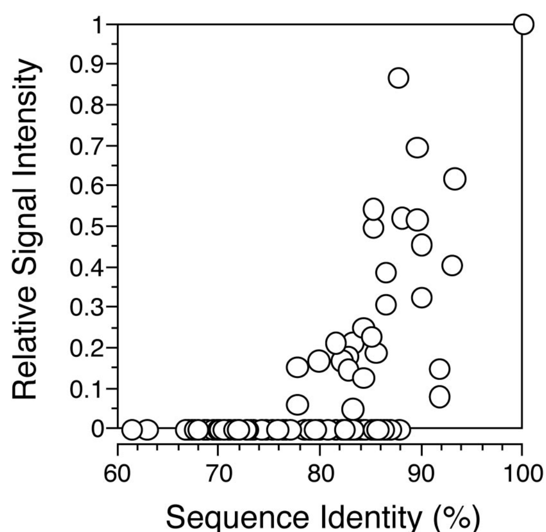

FIG. 7. Cross-hybridization data for all probes in two artificial community hybridization experiments. Within each experiment, signal from each cross-hybridizing probe was normalized to the signal from the most similar probe having a perfect match to one of the targets. Relative signal is plotted as a function of the percent identity between probe and the most similar target.

communities revealed a significant ($P < 0.0001$, $n = 88$) positive correlation (Fig. 8, $r = 0.941$).

DISCUSSION

Performance of the array. Our results suggest that the DNA macroarray procedure which we have developed for profiling mixtures of amplified *nifH* gene fragments is highly reproducible, and to some extent, quantitative. A high degree of reproducibility was illustrated by the significant correlations observed for intra- and interarray comparisons for simple mixtures of six targets and for a complex target mixture derived from a natural community. We note that these experiments demonstrate only the reproducibility of the macroarray procedure itself (array printing and hybridization and signal detection). As our procedure employs PCR-amplified targets, reproducibility for field applications will ultimately be influenced by all of the potential biases in the PCR (4, 8, 19, 21, 25). With some care taken in the amplification procedure, bias can be minimized (19). In practice, good reproducibility has been observed for arrays hybridized with targets amplified independently from the same DNA sample (8a).

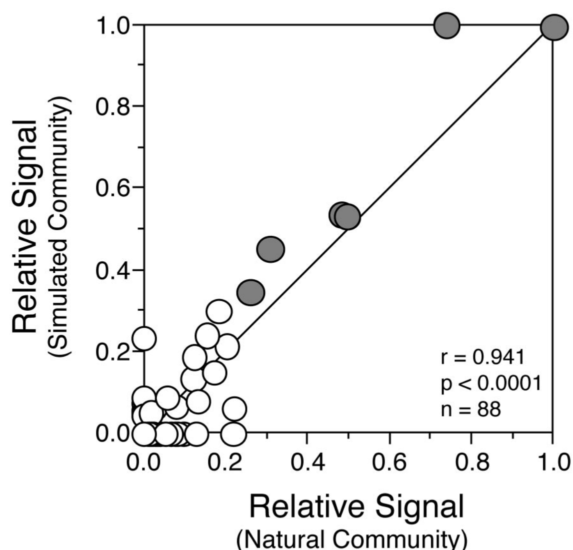

FIG. 8. Correlation between relative signal intensities on arrays hybridized with a natural community target and a simulated community composed of six individual targets. Signals from the probes that are perfect matches to the six targets in the simulated community are indicated by the shaded circles. Within each experiment, signals were normalized to peak probe intensity. The plotted line represents a 1:1 relationship.

Experiments with targets at known concentrations showed that the signal intensity of perfect-match probes was a roughly predictable function of target concentration. The relative signals from six targets that were mixed in equal concentrations were similar. The variability among probe signals was reproduced in two independent experiments, suggesting that there may be inherent differences in the hybridization kinetics of the six probes. The differences were not explained by variations in G+C content, however, and may simply reflect errors in the determination of the individual target concentrations in the stock solutions. Results obtained with unequal target ratios indicated that the relationship between target concentration and relative signal intensity is predictable, but nonlinear, with signal saturation occurring at the highest concentrations (ca. 2 ng/ml). Analysis of the double-reciprocal plot of the data in Fig. 6 and conversion into absolute concentrations indicates that the half-saturation of signal occurred at an absolute target concentration of approximately 0.26 ng ml^{-1}. The lowest known concentration of target tested was 0.13 ng ml^{-1}, which was readily detected. However, extrapolation from the relationship between concentration and signal in Fig. 6 to the lowest signals reliably detected suggests that the detection limit is an order of magnitude lower. The dynamic range for the assay thus appears to be roughly 2 orders of magnitude.

A dilution series of biotinylated lambda DNA was included to compensate for nonlinearity in the response of the X-ray film in recording the chemiluminescence signal. The standard curves showed that in some cases the response was linear over the entire standard range, but in some cases saturation was apparent at higher standards (data not shown). This standard curve controls only for detection system response (any nonlinearity of the chemiluminescence reaction and film response) and not for nonlinearity in hybridization kinetics. In practice, and perhaps for the aforementioned reason, normalization to the standard curve provided little or no improvement in interpreting the hybridization results, and probe signals were instead normalized within an experiment to the average or lowest perfect-match probe signal or, for natural samples, simply to peak probe signal. The standard curve did prove useful, however, as a positive control for the signal detection system and for gauging appropriate exposure times.

Analyses of the cross-hybridization of targets to non-perfect-match probes indicate that, for the stringency conditions that we routinely employed, the threshold for hybridization of target to probe lies between 78 and 88% sequence identity. This

is similar to that reported for oligonucleotide (26) and clone-based arrays (31). The cross-hybridization threshold was readily controlled by adjusting stringency with formamide concentration. Analysis of changes in hybridization pattern in response to deliberate manipulation of stringency may be useful for extracting more information from a sample, since perfect-match signal declines less than signals from cross-hybridization (Fig. 2) (26, 31). Comparison of hybridization signals at several stringencies could thus reveal the presence of sequences having different degrees of similarity to each probe. This phenomenon is currently being exploited with technology for monitoring real-time denaturation patterns on microarrays (7). Signal from cross-hybridizing probes above the threshold is positively correlated with percent identity to the target, but significant variability was evident both for single-target (Fig. 2) and mixed-target (Fig. 7) experiments. The scatter in the relationship between percent similarity and relative cross-hybridization intensity suggests considerable probe-to-probe variability. Possible contributing factors include variations in G+C content and variations in the distribution of mismatches, both of which will affect duplex stability. This was illustrated in a recent study using an oligonucleotide functional gene array where prediction of cross-hybridization was improved if the free energy of binding was considered in addition to overall sequence similarity (26).

The cross-hybridization threshold achievable with the DNA array has implications for array design and interpretation of hybridization patterns. Using higher levels of stringency improves resolution among closely related sequence types but simultaneously increases the number of probes required to detect the same range of genetic diversity. We had intended to design the CBv1 array to minimize the number of probes needed to assure that every sequence in the Chesapeake Bay *nifH* sequence library could hybridize (or cross-hybridize) to at least one probe. However, subsequent analysis of *nifH* diversity turned up new sequences, many with much less than 85% similarity to any probe on the array. This highlights one of the difficulties one may encounter in the development of clone-based arrays, particularly in environments with high diversity. Exhaustive assessment of the diversity must be made to ensure a truly representative array, but this can be difficult to achieve. Oligonucleotide-based arrays are not constrained by clone library diversity, since any desired probe sequence can be synthesized. High-density arrays covering all known (and even as yet unknown) variants of a particular gene region can be created to make potentially universally applicable arrays. Some attempts in this direction have been made for examining 16S rRNA gene diversity (30). The results have been promising, but the major drawback to oligonucleotide arrays at present is their expense. Like clone-based arrays, the results are still only semiquantitative (26, 31).

Potential for application of the array in the field. Our results highlight the reproducibility and potential for quantification of the macroarray approach. Interpreting the quantitative information obtained from complex environmental samples can be confounded by a number of nonlinear factors that influence hybridization. Even with only a single, perfectly matched probe-target pair, the relationship between concentration of target in solution and amount of target hybridized to a membrane-bound probe is nonlinear due to the underlying kinetics (e.g., Fig. 6). Competition for binding sites among perfectly matched and slightly mismatched targets will contribute to nonlinear relationships between target concentration and probe signal, as will sequence-dependent variables such as G+C content and the distribution of mismatches.

Our experiments were not designed to explicitly investigate these finer scale issues, but systematic differences in melting temperature among the probes on the CBv1 array seems likely, since the G+C content of the probes varies from 37 to 68%. We did not observe a significant correlation ($r = 0.32$, $P = 0.57$, $n = 6$) between G+C content and the relative signal intensity for the six perfect match probes in mixture A, but the power of this test was limited because of the few probes analyzed and the limited range of G+C content represented (56 to 65%). Comparison of distance matrices using only the 5′ or 3′ halves of the *nifH* sequences generated for this study indicates that there is significant variability in the distribution of mismatches (data not shown). In the most extreme case, CB907H9 and CB907H10 share 73% identity over their 5′ halves compared to 97% for their 3′ halves. Such extreme disparity could be indicative of chimeras (20) but may also simply reflect differences in constraints on evolution over different regions of the molecule. A number of these complications can be circumvented. The influence of G+C content on binding kinetics, for example, might be minimized by the addition of tetramethylammonium chloride to the hybridization buffer (22). Uneven distribution of mismatches (and perhaps differences in G+C content) may be minimized by careful selection of the fragment location used for the probe. This is easier to achieve with oligonucleotide arrays, since clone-based arrays are constrained by the availability of conserved sequences flanking the selected region to allow PCR amplification.

As a semiquantitative tool, macroarray analyses may serve as an alternative or a supplement to other fingerprinting methods, such as DGGE or analysis of TRFLP. An advantage of the macroarray approach over these other methods is that hybridization to a particular probe provides a more direct, robust indication of phylogenetic affinity based on overall sequence similarity. A disadvantage is that a large amount of work can be required at the outset to establish a clone library, characterize the diversity, and design a useful array. Any diversity not represented on the array will go undetected. Once a comprehensive array is developed for a particular environment, however, analysis of samples appears to be relatively straightforward. Our ability to closely replicate the hybridization or cross-hybridization pattern of a natural sample suggests that the relative probe signals obtained from the natural sample provided at least a semiquantitative representation of the identity and relative abundance of targets in the PCR mixture. In this case, the natural target mixture was dominated by six clones representing three divergent clusters.

The reliable performance of the macroarray and its lower cost compared to microarrays suggest that this is a viable method for conducting spatial or temporal surveys of functional gene diversity in the environment. The usefulness of the macroarray in such applications has been demonstrated in a parallel study (8a). By substituting templates derived from RNA rather than DNA, the macroarray procedure we have developed here may also be used to study patterns of nitroge-

nase gene expression among complex communities in the environment.

ACKNOWLEDGMENTS

M. A. Voytek and J. L. Collier are gratefully acknowledged for assistance with the collection of field samples. We thank G. A. Jackson for creating and maintaining the Biocomplexity project web site (http://snow.tamu.edu/).

This research was supported in part by NSF Biocomplexity grants to J.P.Z. (OCE 9981437) and B.B.W. (OCE 9981482).

REFERENCES

1. **Affourtit, J., J. P. Zehr, and H. W. Paerl.** 2001. Distribution of nitrogen-fixing microorganisms along the Neuse River Estuary, North Carolina. Microb. Ecol. **41:**114–123.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Bagwell, C. E., and C. R. Lovell.** 2000. Persistence of selected *Spartina alterniflora* rhizoplane diazotrophs exposed to natural and manipulated environmental variability. Appl. Environ. Microbiol. **66:**4625–4633.
4. **Chandler, D. P., J. K. Frederickson, and F. J. Brockman.** 1997. Effect of PCR template concentration on the composition and distribution of total community 16S rDNA libraries. Mol. Ecol. **6:**475–483.
5. **Chien, Y. T., and S. H. Zinder.** 1996. Cloning, functional organization, transcript studies, and phylogenetic analysis of the complete nitrogenase structural genes (*nifHDK2*) and associated genes in the archaeon *Methanosarcina barkeri* 227. J. Bacteriol. **178:**143–148.
6. **Dennis, P., E. A. Edwards, S. N. Liss, and R. Fulthorpe.** 2003. Monitoring gene expression in mixed microbial communities by using DNA microarrays. Appl. Environ. Microbiol. **69:**769–778.
7. **El Fantroussi, S., H. Urakawa, A. E. Bernhard, J. J. Kelly, P. A. Noble, H. Smidt, G. M. Yershov, and D. A. Stahl.** 2003. Direct profiling of environmental microbial populations by thermal dissociation analysis of native rRNAs hybridized to oligonucleotide microarrays. Appl. Environ. Microbiol. **69:**2377–2382.
8. **Farrelly, V., F. A. Rainey, and E. Stackebrandt.** 1995. Effect of genome size and *rrn* gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. Appl. Environ. Microbiol. **61:**2798–2801.

8a. **Jenkins, B. D., G. F. Steward, S. M. Short, B. B. Ward, and J. P. Zehr.** 2004. Fingerprinting diazotroph communities in the Chesapeake Bay by using a DNA macroarray. Appl. Environ. Microbiol. **70:1767–1776.**

9. **Lilburn, T. C., K. S. Kim, N. E. Ostrom, K. R. Byzek, J. R. Leadbetter, and J. A. Breznak.** 2001. Nitrogen fixation by symbiotic and free-living spirochetes. Science **292:**2495–2498.
10. **Liu, W.-T., T. L. Marsh, H. Cheng, and L. J. Forney.** 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Appl. Environ. Microbiol. **63:**4516–4522.
11. **Lovell, C. R., Y. M. Piceno, J. M. Quattro, and C. E. Bagwell.** 2000. Molecular analysis of diazotroph diversity in the rhizosphere of the smooth cordgrass, *Spartina alterniflora*. Appl. Environ. Microbiol. **66:**3814–3822.
12. **Loy, A., A. Lehner, N. Lee, J. Adamczyk, H. Meier, J. Ernst, K.-H. Schleifer, and M. Wagner.** 2002. Oligonucleotide microarray for 16S rRNA gene-based detection of all lineages of sulfate-reducing prokaryotes in the environment. Appl. Environ. Microbiol. **68:**5064–5081.
13. **MacGregor, B. J., B. Van Mooy, B. J. Baker, M. Mellon, P. H. Moisander, H. W. Paerl, J. Zehr, D. Hollander, and D. A. Stahl.** 2001. Microbiological, molecular biological and stable isotopic evidence for nitrogen fixation in the open waters of Lake Michigan. Environ. Microbiol. **3:**205–219.
14. **Muyzer, G., E. C. De Wall, and A. G. Uitterlinden.** 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. **59:**695–700.
15. **Noda, S., M. Ohkuma, R. Usami, K. Horikoshi, and T. Kudo.** 1999. Culture-independent characterization of a gene responsible for nitrogen fixation in the symbiotic microbial community in the gut of the termite *Neotermes koshunensis*. Appl. Environ. Microbiol. **65:**4935–4942.
16. **Ohkuma, M., S. Noda, and T. Kudo.** 1999. Phylogenetic diversity of nitrogen fixation genes in the symbiotic microbial community in the gut of diverse termites. Appl. Environ. Microbiol. **65:**4926–4934.
17. **Ohkuma, M., S. Noda, R. Usami, K. Horikoshi, and T. Kudo.** 1996. Diversity of nitrogen fixation genes in the symbiotic intestinal microflora of the termite *Reticulitermes speratus*. Appl. Environ. Microbiol. **62:**2747–2752.
18. **Poly, F., L. J. Monrozier, and R. Bally.** 2001. Improvement in the RFLP procedure for studying the diversity of *nifH* genes in communities of nitrogen fixers in soil. Res. Microbiol. **152:**95–103.
19. **Polz, M. F., and C. M. Cavanaugh.** 1998. Bias in template-to-product ratios in multitemplate PCR. Appl. Environ. Microbiol. **64:**3724–3730.
20. **Qiu, X. Y., L. Y. Wu, H. S. Huang, P. E. McDonel, A. V. Palumbo, J. M. Tiedje, and J. Z. Zhou.** 2001. Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. Appl. Environ. Microbiol. **67:**880–887.
21. **Reysenbach, A.-L., L. J. Giver, G. S. Wickham, and N. Pace.** 1992. Differential amplification of rRNA genes by polymerase chain reaction. Appl. Environ. Microbiol. **58:**3417–3418.
22. **Sambrook, J., and D. W. Russell.** 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
23. **Small, J., D. R. Call, F. J. Brockman, T. M. Straub, and D. P. Chandler.** 2001. Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. Appl. Environ. Microbiol. **67:**4708–4716.
24. **Steward, G. F., J. P. Zehr, R. Jellison, J. P. Montoya, and J. T. Hollibaugh.** Vertical distribution of nitrogen-fixing phylotypes in a meromictic, hypersaline lake. Microb. Ecol., in press.
25. **Suzuki, M. T., and S. J. Giovannoni.** 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. **62:**625–630.
26. **Taroncher-Oldenburg, G., E. M. Griner, C. A. Francis, and B. B. Ward.** 2003. Oligonucleotide microarray for the study of functional gene diversity in the nitrogen cycle in the environment. Appl. Environ. Microbiol. **69:**1159–1171.
27. **Ueda, T., Y. Suga, N. Yahiro, and T. Matsuguchi.** 1995. Genetic diversity of N_2-fixing bacteria associated with rice roots by molecular evolutionary analysis of a *nifD* library. Can. J. Microbiol. **41:**235–240.
28. **Vitousek, P. M., and R. W. Howarth.** 1991. Nitrogen limitation on land and in the sea: how can it occur? Biogeochemistry **13:**87–115.
29. **Widmer, F., B. T. Shaffer, L. A. Porteous, and R. J. Seidler.** 1999. Analysis of *nifH* gene pool complexity in soil and litter at a Douglas fir forest site in the Oregon Cascade Mountain Range. Appl. Environ. Microbiol. **65:**374–380.
30. **Wilson, K. H., W. J. Wilson, J. L. Radosevich, T. Z. DeSantis, V. S. Viwanathan, T. A. Kuczmarski, and G. L. Andersen.** 2002. High-density microarray of small-subunit ribosomal DNA probes. Appl. Environ. Microbiol. **68:**2535–2541.
31. **Wu, L. Y., D. K. Thompson, G. Li, R. A. Hurt, J. M. Tiedje, and J. Zhou.** 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. Appl. Environ. Microbiol. **67:**5780–5790.
32. **Zani, S., M. T. Mellon, J. L. Collier, and J. P. Zehr.** 2000. Expression of *nifH* genes in natural microbial assemblages in Lake George, New York, detected with reverse transcriptase PCR. Appl. Environ. Microbiol. **66:**3119–3124.
33. **Zehr, J. P., B. D. Jenkins, S. M. Short, and G. F. Steward.** 2003. Nitrogenase gene diversity and microbial community structure: a cross-system comparison. Environ. Microbiol. **5:**539–554.
34. **Zehr, J. P., and L. A. McReynolds.** 1989. Use of degenerate oligonucleotides for amplification of the *nifH* gene from the marine cyanobacterium *Trichodesmium thiebautii*. Appl. Environ. Microbiol. **55:**2522–2526.
35. **Zehr, J. P., M. Mellon, S. Braun, W. Litaker, T. Steppe, and H. W. Paerl.** 1995. Diversity of heterotrophic nitrogen fixation genes in a marine cyanobacterial mat. Appl. Environ. Microbiol. **61:**2527–2532.
36. **Zehr, J. P., M. T. Mellon, and S. Zani.** 1998. New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (*nifH*) genes. Appl. Environ. Microbiol. **64:**3444–3450.
37. **Zehr, J. P., and P. J. Turner.** 2001. Nitrogen fixation: nitrogenase genes and gene expression, p. 271–286. *In* J. H. Paul (ed.), Methods in marine microbiology. Academic Press, New York, N.Y.
38. **Zehr, J. P., J. B. Waterbury, P. J. Turner, J. P. Montoya, E. Omoregie, G. F. Steward, A. Hansen, and D. M. Karl.** 2001. Unicellular cyanobacteria fix N-2 in the subtropical North Pacific Ocean. Nature **412:**635–638.