4.

Chapter 4

Discrimination of shifts in a soil microbial community associated with TNT-contamination using a functional ANOVA of 16S rRNA hybridized to oligonucleotide microarrays

The organization of the chapters is as follows:

Chapter 1 emphasizes the characterization of microbial populations using molecular techniques based on phylogenetic markers.

Chapter 2 describes bacterial TNT transformation and shows that TNT denitration is promising for complete mineralization of the molecule.

Chapter 3 presents a functional ANOVA model to analyze dissociation curves from hybridizations to microarrays with thermal dissociation.

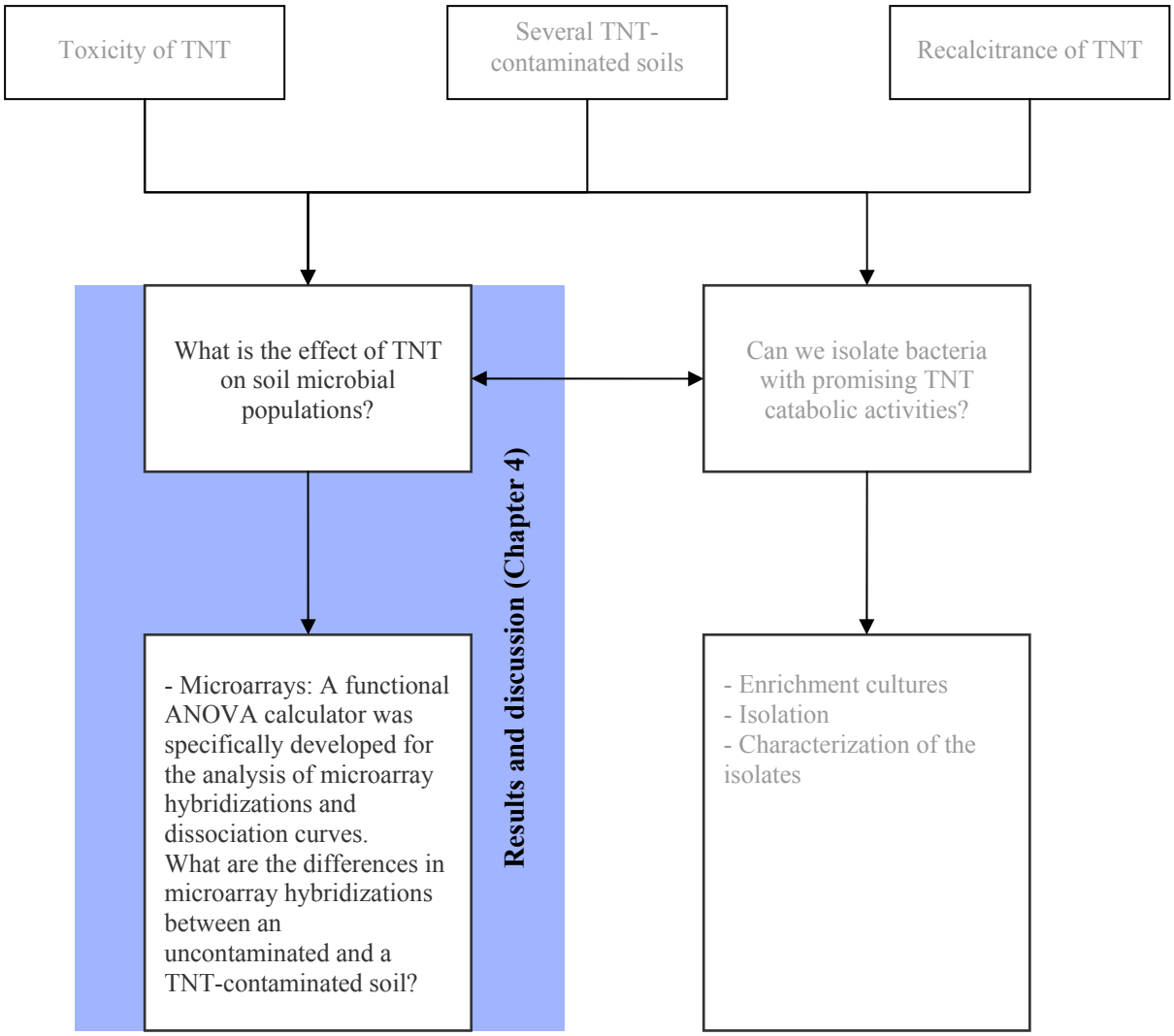
Chapter 4 compares hybridizations of in vitro transcribed 16S rRNA derived from an uncontaminated and a TNT-contaminated soil sample to a phylogenetic microarray with thermal dissociation analysis.

Chapter 5 compares extracted DNA and 16S PCR-DGGE fingerprints of uncontaminated and TNT-contaminated soil samples, and of a pristine soil artificially contaminated with TNT. Plate counts on agar plates inoculated with an uncontaminated and a TNT-contaminated soil sample are also evaluated.

Chapter 6 describes the isolation and characterization of a *Pseudomonas aeruginosa* strain that denitrates TNT in the absence of oxygen and presence of ferrihydrite.

Chapter 7 shows that a phenazine molecule produced by *Pseudomonas aeruginosa* denitrates TNT.

Chapter 8 discusses the results obtained from the characterization of bacterial populations of TNT-contaminated sites and the isolation of a TNT-degrading strain.


Toxicity of TNT
Several TNT-contaminated soils
Recalcitrance of TNT
What is the effect of TNT on soil microbial populations?
Can we isolate bacteria with promising TNT catabolic activities?
Results and discussion (Chapter 4)
- Microarrays: A functional ANOVA calculator was specifically developed for the analysis of microarray hybridizations and dissociation curves.
What are the differences in microarray hybridizations between an uncontaminated and a TNT-contaminated soil?
- Enrichment cultures
- Isolation
- Characterization of the isolates

Discrimination of shifts in a soil microbial community associated with TNT-contamination using a functional ANOVA of 16S rRNA hybridized to oligonucleotide microarrays

Laurent Eyers, [†] James C. Smoot, [‡] Laura M. Smoot, [‡] Céline Bugli, [§] Hidetoshi Urakawa, [ℓ] Zack McMurry, [‡] Slil Siripong, [⌂] Said El Fantroussi, [†] Philippe Lambert, [§] Spiros N. Agathos, [†] and David A. Stahl[*,‡]

Unit of Bioengineering and Institute of Statistics, University of Louvain, 1348 Louvain-la-Neuve, Belgium, Civil and Environmental Engineering, University of Washington, Seattle, Washington 98195, Ocean Research Institute, University of Tokyo, Tokyo 164-8639, Japan, and Civil and Environmental Engineering, Northwestern University, Evanston, Illinois 60208

[*]Corresponding author phone: (206) 685-3464; fax: (206) 685 9185; e-mail: dastahl@u.washington.edu. Corresponding author address: Civil and Environmental Engineering, University of Washington, 302 More Hall, Seattle, WA 98195.

[†] Unit of Bioengineering, University of Louvain.

[§] Institute of Statistics, University of Louvain.

[‡] University of Washington.

[ℓ] University of Tokyo.

[⌂] Northwestern University.

Environmental Science and Technology (2006) 40:5867-5873

4.1. Abstract

A functional ANOVA analysis of the thermal dissociation of RNA hybridized to DNA microarrays was used to improve discrimination between two soil microbial communities. Following hybridization of in vitro transcribed 16S rRNA derived from uncontaminated and 2,4,6-trinitrotoluene contaminated soils to an oligonucleotide microarray containing group- and species-specific perfect match (PM) probes and mismatch (MM) variants, thermal dissociation was used to analyze the nucleic acid bound to each PM-MM probe set. Functional ANOVA of the dissociation curves generally discriminated PM-MM probe sets when T_d values (temperature at 50% probe-target dissociation) could not. Maximum discrimination for many PM and MM probes often occurred at temperatures greater than the T_d. Comparison of signal intensities measured prior to dissociation analysis from hybridizations of the two soil samples revealed significant differences in domain-, group-, and species-specific probes. Functional ANOVA showed significantly different dissociation curves for 11 PM probes when hybridizations from the two soil samples were compared, even though initial signal intensities for 3 of the 11 did not vary.

4.2. Introduction

Microarrays containing DNA oligonucleotide probes that target rRNA molecules were developed to detect single species in environmental matrices (*1*) and to evaluate the bacterial composition of complex environmental systems (*2-6*). Ideally, 16S rRNA microarrays should provide a direct characterization of the bacteria present in the system and information on their relative abundances. When nucleic acids isolated from pure cultures are examined with DNA microarrays, unambiguous hybridization patterns are generally observed, even between closely related bacterial species (*7*). However, when microarrays are used to evaluate environmental samples, the relative abundance of target and closely related nontarget molecules in the environment is unknown. Therefore, it is difficult to determine the extent of cross-hybridization events using signal intensity values at a single wash temperature (*6*).

Previous studies using thermal dissociation analyses (dissociation curves) were successful at discriminating hybridization events of perfect match (PM) and single mismatch (MM) duplexes. El Fantroussi et al. (*6*) visually evaluated (i.e., without statistical analysis) the extent of the discrimination between dissociation curves of PM and MM duplexes. Liu et al. (*8*) used T_d values (temperature at which 50% of the probe-target duplexes remain intact) and obtained clear discrimination between PM and MM duplexes. Urakawa et al. (*9*) found that T_d values did not effectively discriminate PM from MM duplexes. Therefore, they used a discrimination index (DI) to determine an optimum discrimination temperature defined as the temperature corresponding to the maximum product of difference and ratio of the signal intensities of PM and MM duplexes. However, these differences in signal intensities were not demonstrated to be statistically significant. In order to establish an appropriate statistical framework for analysis, Bugli et al. (manuscript submitted) developed a functional ANOVA calculator to analyze each dissociation curve and identify statistically different dissociation curves.

Changes in microbial communities of polluted soil samples are the subject of numerous investigations. 2,4,6-Trinitrotoluene (TNT) is frequently found as a soil pollutant at TNT-production facilities and at sites used for ammunition testing and destruction. Despite its recalcitrance to microbial degradation and its toxicity for many microorganisms, few reports describe the effect of TNT on soil microbial populations, and there are currently no reports

providing extensive identification of bacterial populations in TNT-contaminated soils. Fuller and Manning (*10*) showed that Gram-positive bacteria are more sensitive to TNT than Gram-negative bacteria using heterotrophic plate counts and phospholipid fatty acid analyses. However, the identity of the bacteria was not resolved using these techniques. Eyers et al. (*11*) showed that the presence of TNT caused a dramatic alteration in soil populations using denaturing gradient gel electrophoresis of 16S rRNA gene amplicons, but sequencing of clones from the 16S rRNA gene amplicons was not performed in that study.

The general objective of this study was to further document the utility of DNA microarrays to resolve shifts in microbial community structure associated with TNT contamination. Specific objectives were to (i) compare hybridizations of in vitro transcribed 16S rRNA derived from uncontaminated soil and TNT-contaminated soil, and (ii) assess the effectiveness of functional ANOVA to discriminate PM and MM duplexes compared to traditional approaches based on signal intensity measured prior to dissociation analysis and T_d values. Functional ANOVA was used to compare thermal dissociation curves of PM probes hybridized with native RNA of *Pseudomonas putida* and in vitro transcribed 16S rRNA of *P. putida*, uncontaminated, and TNT-contaminated soil. To evaluate the specificity of each PM probe hybridized with RNA from soil samples, two single MM probes were designed for each PM probe. Dissociation curves generated from hybridizations of RNA from uncontaminated and TNT-contaminated soil samples were used to compare PM and MM probe performance and microbial composition of the two soil samples. The functional ANOVA was shown to resolve differences between most PM and MM probe sets, providing for much greater resolution of microbial communities than could be achieved using only differences in signal intensity or T_d.

4.3. Materials and methods

Sample collection and processing of nucleic acids. *Pseudomonas putida* sp. JLR 11 was used because of its known association with TNT degradation (*12*) and because it is a suitable reference target to validate probes Univ907, Univ1390, Eub338, Eub927, and Gam42a. The strain was provided by Dr. J. L. Ramos (CSIC, Estacion Experimental del Zaidin, Spain). Samples were collected from TNT-contaminated soil (23.0 ± 2.2 g TNT/kg soil) and uncontaminated soil (i.e., TNT was not detected (detection limit of 1 mg TNT/kg soil)) at a site in Bourges, France that was used for TNT destruction during the past 20 years. The uncontaminated soil, which was sampled a few meters from the contaminated soil, was unpolluted due to the presence of a protecting wall at the destruction field. Both samples were collected from the top 5-cm layer of soil, homogenized, and stored at 4 °C for 1 month until DNA extraction.

RNA used for microarray hybridizations was native RNA (i.e., total RNA - including rRNA, tRNA, and mRNA - harvested from cellular biomass at the time of lysis) isolated from *P. putida*, and non-native RNA was derived from in vitro transcription of the *P. putida* 16S rRNA gene or 16S rRNA gene sequences amplified from soil DNA. Native RNA of *P. putida* was isolated with the Qiagen RNeasy mini kit (Valencia, CA) following the manufacturer's instructions. Total DNA from *P. putida* was extracted by lysozyme and sodium dodecylsulfate lysis followed by phenol extraction (*13*). DNA was extracted from soil samples with the UltraClean Soil DNA kit from Mo Bio (Carlsbad, CA) following the manufacturer's instructions. DNA isolated from *P. putida* and soil samples was used as a template for PCR amplification of the 16S rRNA gene under conditions described in the Supporting Information. PCR products were purified with the Qiaquick PCR purification kit (Qiagen, Inc.) following the manufacturer's instructions. T7 RNA polymerase (Invitrogen, Inc., Carlsbad, CA, Cat. No. 18033-019) was used for in vitro transcription to produce 16S rRNA according to the manufacturer's protocol. RNA samples were purified with the Qiagen RNeasy mini kit.

Fragmentation and labeling of RNA. RNA was simultaneously fragmented and end-labeled following the protocol of Bavykin et al. (*14*) with modifications given in the Supporting Information.

Microarray fabrication and hybridization. CodeLink and gel-pad microarray formats were used in this study. The two formats are described in the Supporting Information. The gel-pad microarrays contained 29 different PM oligonucleotide probes (13-25 nucleotides in length) that were complementary to universal-, domain-, group-, and species-specific rRNA sequences (Table S-1 in the Supporting Information). Fifteen of these 29 probes were spotted on the CodeLink microarrays (Table S-1). In addition to PM probes, single MM probes were printed on the gel-pad microarray. The design of PM and single MM probes was based on several criteria given in the Supporting Information. Hybridizations were conducted in triplicate for each sample, using 5 μg of RNA that was fragmented, labeled, hybridized, and washed as previously described (*6*), except that the hybridization time was reduced to 1 h.

Analysis of images and thermal dissociation curves. Fluorescence signal intensities were recorded and dissociation curves were generated as described in the Supporting Information. Data of hybridizations to the gel-pad and CodeLink microarrays are available at the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under respectively the accession numbers GSE3499 and GSE3525. Data from each hybridization were analyzed with the functional ANOVA calculator (available at http://stahl.ce.washington.edu) developed by Bugli et al. (manuscript submitted) with Matlab 7.0. Functional ANOVA is an adaptation of linear modeling to analyze curve data (*15*). Functional ANOVA uses penalized spline regressions to smooth data points of individual dissociation curves and obtain continuous curves. The functional ANOVA calculator was used to (i) obtain continuous curves, (ii) identify and reject outlier curves, (iii) normalize individual curves by setting the lowest intensity in a profile at a value of 0 and the highest intensity at 1, (iv) retrieve the T_d value of individual curves, (v) average curves for each probe, (vi) calculate an interval of confidence for each average curve and obtain the *p*-values associated to the comparison of these curves, (vii) calculate, when the confidence bands of two average curves do not overlap, the maximum difference (MAX_{DCSD}, where DCSD stands for "dissociation curves significantly different") in normalized signal intensities between the lower limit of the interval of confidence of the upper dissociation curve and the upper limit of the lower dissociation curve, (viii) calculate the temperature corresponding to the MAX_{DCSD}, and (ix) identify when signal intensities and T_d values between probes were significantly different (unpaired tests described in the Supporting Information).

Calculation of confidence bands was performed with an α value of 0.05. Outlier dissociation curves were not included in the functional ANOVA model, calculation and comparison of T_d values. Outlier curves were detected using leave-one-out error. The "leave-one-out error" is the error of the computed functional ANOVA model determined by successive single exclusion of each of the observed curves. The error was calculated by summing for each point the square difference between observed points and estimated points determined with the model. A boxplot of the leave-one-out errors for each of the curves allowed the detection of extreme values of the errors and consequently outlier curves. The rejection of outliers often resulted in the calculation of different confidence intervals and T_d values than without rejection.

4.4. Results and discussion

Discrimination of PM and MM probes with signal intensity and T_ds. Our ultimate analytical goal is to develop a combined experimental and statistical framework to determine sequence identity of nucleic acid hybridized to each array element (PM vs MM). For example, a common design feature of DNA microarrays has been the incorporation of mismatch variants of probes complementary to specific target sequences, setting a threshold of hybridization between match and mismatch variants as a basis to evaluate hybridization to the intended target sequence (*16*). However, given the complexity and undefined composition of natural microbial communities, we anticipate that such an assignment will not be possible using data collected at a single stringency of hybridization. For example, a highly abundant but closely related nontarget sequence could skew PM/MM hybridization values. Alternatively, unknown microbial species present in the sample may have sequences complementary to a subset of the MM probes. Therefore, identifying hybridization events that deviate from expectations is of utmost importance for correct assignment. In this study, we used a new statistical package (functional ANOVA calculator developed by Bugli et al., manuscript submitted) to analyze data collected using nonequilibrium dissociation of rRNA hybridized to PM and MM probes (Table S-1) on oligonucleotide microarrays.

Initial analysis of collected dissociation curves revealed some curves that deviated from the general sigmoid shape of the majority of the curves. The shapes of these "outlier" curves differed from those reported by Pozhitkov et al. (*17*). Three classes of outliers were detected and are described elsewhere (Bugli et al., manuscript submitted). One hundred fifty-one individual curves from a total of 1548 were statistically identified as outliers and excluded from further analysis (Bugli et al., manuscript submitted). Ongoing studies have identified several contributing factors (movement of the microarray during data collection, sporadic occurrence of fluorescent particles, and the formation of bubbles within the hybridization chamber at high temperature) that are being alleviated by improvements in the experimental protocol and image analysis software (e.g. ref *18*).

Significant differences in initial signal intensities between PM and their respective c-MM and/or g-MM variants[1] were achieved for 24 PM probes using nucleic acid derived from uncontaminated soil and for 23 PM probes with contaminated soil (Table S-2). Thus, initial signal intensity data provided some discriminatory information about the two soil communities. However, dissociation analysis provided the basis for a higher resolution comparison of the two communities. For instance, comparison of dissociation curves from in vitro transcribed 16S rRNA of *P. putida* with in vitro transcribed 16S rRNA of the uncontaminated and TNT-contaminated soil sample showed no differences in the shape of the curves for the *Bacteria* probe Eub338 (Figure 1A). This and previous studies (*6, 8*) demonstrate that a robust and reproducible reference curve can be established for each probe.

[1] c-MM and g-MM probes consist in a substitution of G to C or C to G, respectively (cfr. Supporting Information)

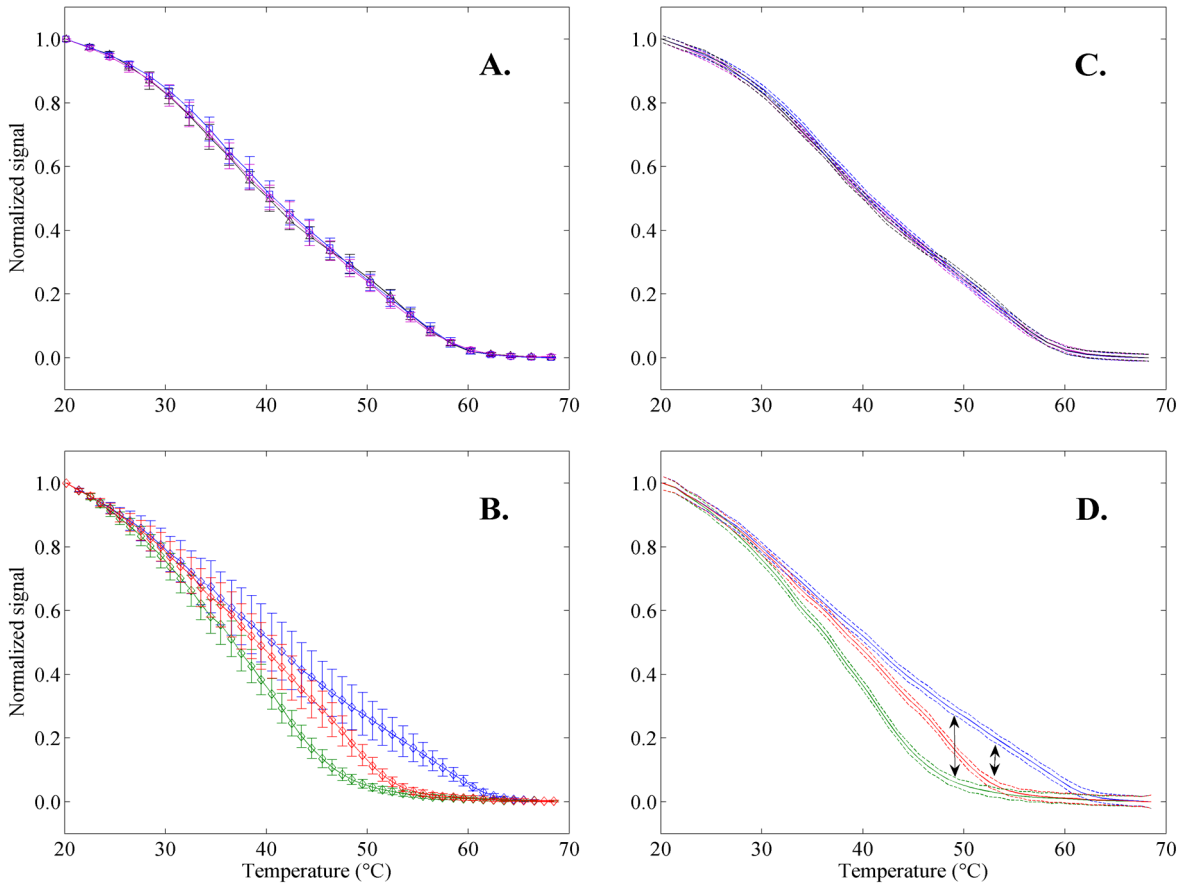


Figure 1. Dissociation curves analysis. Average normalized dissociation curves of probe Eub338 hybridized with in vitro transcribed 16S rRNA of *P. putida* (black triangles, *n*=18), uncontaminated (blue squares, *n*=18), and TNT-contaminated soil sample (pink circles, *n*=18) to CodeLink microarrays (A). Average normalized dissociation curves of probe Eub338 (blue lozenges, *n*=12), Eub338-g11c (green lozenges, *n*=12), and Eub338-t7g (red lozenges, *n*=12) hybridized with in vitro transcribed 16S rRNA of uncontaminated soil sample to gel-pad microarrays (B). Error bars represent the standard deviation. Average normalized dissociation curves with functional ANOVA analysis of probe Eub338 hybridized with in vitro transcribed 16S rRNA of *P. putida* (black line), uncontaminated (blue line), and TNT-contaminated soil sample (pink line) to CodeLink microarrays (C). Average normalized dissociation curves with functional ANOVA analysis of probe Eub338 (blue line), Eub338-g11c (green line), and Eub338-t7g (red line) hybridized with in vitro transcribed 16S rRNA of uncontaminated soil sample to gel-pad microarrays (D). The number of individual dissociation curves of C and D are identical to A and B, respectively. Confidence bands are indicated by non-continuous lines. The big and small arrows represent, respectively, the MAX_{DCSD} of Eub338/Eub338-g11c and Eub338/Eub338-t7g.

Thermal dissociation generates a large data set for each probe, and the analysis of several hundred microarray elements requires computational processing to summarize the data. One

method of data reduction is the determination of T_d for each probe. Although the strategy used in designing single MM probes was to have a 4-6 °C difference in T_ds between PM and MM probes, discrimination was not achieved for all probes using RNA from the complex soil communities, and the extent of discrimination differed between samples. Significant differences between PM probes and their respective c-MM and/or g-MM variants were achieved for 21 PM probes using nucleic acid derived from uncontaminated soil and for only 16 PM probes with contaminated soil (Table S-3). In addition, T_ds of PM probes Eub338-II and Eub338-III, which target *Planctomycetes* and *Verrucomicrobia*, respectively, and differ by a single T or A nucleotide, were not significantly different when hybridizations were realized with in vitro transcribed 16S rRNA of the uncontaminated or contaminated soil (Table S-3). These data suggest that either detection of a single nucleotide change is unattainable for most of the probes or that T_d alone is not a significant measure to discriminate closely related targets from taxonomically distinct microorganisms.

Visual comparisons of normalized dissociation curves of PM and MM probes suggested that discrimination was possible for many probes even when their T_ds were not different. For example, when in vitro transcribed 16S rRNA of the uncontaminated soil sample was hybridized to the microarray, a significant difference in T_d was detected between probes Eub338 and Eub338-g11c but not between probes Eub338 and Eub338-t7g (Table S-3). However, discriminating curves were observed for probes Eub338, Eub338-g11c, and Eub338-t7g (Figure 1B). The reason for the lack of discrimination of T_d between probe Eub338 and Eub338-t7g was that the primary region of difference for these curves occurred at the final stages of probe-target dissociation (e.g., between 50 °C and 60 °C) at temperatures greater than the T_ds for probes Eub338, Eub338-g11c, and Eub338-t7g (40.5 °C, 36.7 °C and 39.1 °C, respectively).

In a previous study, Urakawa et al. (*9*) reported small differences between the T_ds of PM and single MM probes. The authors introduced the DI value (defined as the product of difference and ratio of the signal intensities for PM and MM duplexes at a given wash temperature) as a mechanism to determine the optimum wash temperature for each probe set (*9*). However, the DI approach gives no indication that differences in dissociation curves are statistically significant.

Discrimination of PM and MM probes with functional ANOVA analysis. The functional ANOVA was more efficient than the T_d approach in determining when dissociation curves significantly differed. This is due to the fact that functional ANOVA compares signal intensities of dissociation curves at each temperature, whereas T_ds only take into account signal intensities at 50% dissociation. Examples of the curves produced by the functional ANOVA calculator are shown in Figure 1C,1D. The comparisons of the dissociation curves produced from the hybridizations of different samples to probe Eub338 showed no significant difference with CodeLink (Figure 1C) and gel-pad microarrays (Table 1). The absence of significant differences was also observed with these samples hybridized to probes Univ907, Univ1390, and Eub927 with gel-pad (Table 1) and CodeLink microarrays (results not shown). When the PM and MM Eub338 probes were compared on gel-pad microarrays, significant differences in the dissociation curves were detected between probes Eub338, Eub338-g11c, and Eub338-t7g hybridized with the uncontaminated soil sample (Figure 1D) as shown by non-overlapping confidence bands. In addition to *p*-values (Table 1), MAX_{DCSD} values were calculated to provide an indication of the extent of discrimination between the 95% confidence intervals (e.g., a MAX_{DCSD} value of 0.100 indicates a 10% maximum difference in normalized dissociation curves as they start at 1 and finish at 0). MAX_{DCSD} values were 0.207 and 0.124 for the Eub338/Eub338-g11c and Eub338/Eub338-t7g probe sets, respectively (Table 1). Although not demonstrated, the extent of discrimination may be related to the number of mismatches and/or the presence of multiple targets contributing to the shape of the curve.

Table 1. *p*-Values, MAX_{DCSD}, and temperature at the MAX_{DCSD} of PM/MM probe set[a]

Probe name	Uncontaminated sample						Contaminated sample						Uncontaminated vs contaminated sample		
	PM/c-MM			PM/g-MM			PM/c-MM			PM/g-MM			PM unc./PM cont.		
	p-value	MAX_{DCSD}	*T* (°C)	*p*-value	MAX_{DCSD}	*T* (°C)	*p*-value	MAX_{DCSD}	*T* (°C)	*p*-value	MAX_{DCSD}	*T* (°C)	*p*-value	MAX_{DCSD}	*T* (°C)
Univ907	<0.001	0.099	43.4	0.001	0.036	43.4	<0.001	0.040	45.5	0.302			0.502		
Univ1390	<0.001	0.116	47.5	0.002	0.031	49.5	<0.001	0.121	47.5	0.007	0.020	51.5	0.313		
Euk502	<0.001	0.096	37.5	0.010	0.019	35.5	<0.001	0.108	35.5	0.004	0.029	37.5	0.208		
Arch915	<0.001	0.206	39.4	<0.001	0.091	43.4	<0.001	0.159	41.5	<0.001	0.071	47.5	0.066		
Eury498	<0.001	0.139	34.5	0.004	0.041	34.5	<0.001	0.222	36.6	<0.001	0.138	37.5	0.011	0.026	39.4
Cren499	0.337			<0.001	0.117	39.4	0.267			<0.001	0.111	41.5	0.310		
Eub927	<0.001	0.098	45.5	<0.001	0.051	51.5	<0.001	0.099	47.5	<0.001	0.035	53.6	0.543		
Eub338	<0.001	0.207	47.5	<0.001	0.124	51.5	<0.001	0.205	47.5	<0.001	0.114	53.5	0.164		
Eub338-II	0.259			<0.001	0.073	39.4	0.169			<0.001	0.121	37.5	0.007	0.036	30.5
Eub338-III	<0.001	0.089	39.4	<0.001	0.089	43.4	<0.001	0.092	39.4	<0.001	0.073	43.5	0.017	0.009	37.5
Alf1b	0.008	0.015	55.5	0.097			0.414			<0.001	0.067	43.5	<0.001	0.034	41.5
Alf968	<0.001	0.051	47.5	<0.001	0.075	37.5	<0.001	0.113	39.4	<0.001	0.087	45.5	<0.001	0.090	43.4
Bet42a	0.483			0.001	0.045	34.5	0.338			0.010	0.021	30.5	0.246		
Nmo218	<0.001	0.139	35.5	0.003	0.024	36.6	<0.001	0.222	37.5	<0.001	0.076	35.5	<0.001	0.091	37.5
Nse1472	<0.001	0.102	34.5	0.016	0.008	35.5	<0.001	0.138	34.5	0.424			<0.001	0.070	34.5
Nsv443	0.020	0.019	34.5	0.236			<0.001	0.077	35.5	0.073			0.249		
Gam42a	0.009	0.015	43.4	<0.001	0.123	39.4	0.501			<0.001	0.163	39.4	0.021	0.008	30.5
Nsc825	0.320			0.017	0.009	30.5	<0.001	0.045	33.5	0.191			0.220		
Gpos1192	<0.001	0.072	40.6	0.003	0.044	39.4	0.109			0.074			0.143		
Gpos1200	0.289			0.346			0.134			0.092			0.283		
Hgc69a	0.496			<0.001	0.104	37.5	0.137			0.002	0.058	32.6	0.004	0.045	30.5
Lowgc353	0.002	0.048	34.5	0.001	0.054	43.4	<0.001	0.127	35.5	<0.001	0.083	43.4	0.420		
Bacil1136	0.176			0.002	0.051	47.5	0.006	0.035	32.6	0.070			0.176		
Cf319a	<0.001	0.076	35.5	<0.001	0.059	37.5	<0.001	0.125	35.5	<0.001	0.084	37.5	0.001	0.025	41.5
Ntspa0712	<0.001	0.176	41.5	<0.001	0.105	41.5	<0.001	0.230	39.4	<0.001	0.199	39.4	<0.001	0.030	38.6
Neg. ctrl.		na			na			na			na		0.305		

[a]Values shown are p-values, MAX_{DCSD} (α=0.05), and corresponding temperature of normalized dissociation curves of triplicate hybridizations to gel-pad microarrays. Probes Nso1225, Nso190, Nsm653, and Nsm156 were not analyzed with the functional ANOVA calculator because of the low number of observations (n<4). The number of dissociation curves analyzed are identical to Table S-3 (Supporting Information). na: not applicable.

The functional ANOVA calculator revealed significant differences for practically all the PM-MM probes sets when hybridized with uncontaminated and contaminated soil samples, with MAX_{DCSD} values ranging from 0.008 to 0.230 (Table 1). Twenty-four PM probes were discriminated from their respective c-MM and/or g-MM probes using nucleic acid from the uncontaminated soil sample (Table 1). Of the 24 PM probes, 16 PM probes were discriminated from both their c-MM and g-MM variants. Twenty-three of the 25 PM probes were discriminated in the analysis of the contaminated soil sample (Table 1). Of the 23 probes, 12 PM probes were discriminated from both their c-MM and g-MM probes. In addition, probes Eub338-II and Eub338-III were discriminated from each other in hybridizations done with the uncontaminated and contaminated soil samples (MAX_{DCSD} values of 0.082 and 0.074, respectively). All PM-MM probe sets that had significantly different T_ds were also significantly different as determined using the functional ANOVA calculator (Table S-3 and Table 1). For probes that were significantly different with the functional ANOVA calculator but not by T_d, discrimination generally occurred at higher temperatures than the respective T_ds (e.g., with the uncontaminated soil sample, the temperature of the MAX_{DCSD} for probes Alf968 and Alf968-g7c was 47.5 °C, whereas their respective T_ds were 37.3 and 36.6 °C (Table S-3 and Table 1)). For the probes that were significantly different as determined by both functional ANOVA and T_d, discrimination often occurred at temperatures near the T_ds (e.g., with the uncontaminated soil sample, the temperature of the MAX_{DCSD} between probes Alf968 and Alf968-c13g was 37.5 °C whereas their respective T_ds were 37.3 and 35.2 °C).

The functional ANOVA calculator was also used to examine the presumptive cross hybridization of the 23S rRNA probe Gam42a with in vitro transcribed 16S rRNA of *P. putida*. Dissociation curves from the 23S rRNA probe Gam42a hybridized to native *P. putida* RNA and the in vitro transcribed 16S rRNA molecule of *P. putida* were significantly different (MAX_{DCSD} value of 0.469). The 23S rRNA molecule from native *P. putida* RNA is a suitable reference standard because it is a PM target to the Gam42a probe, whereas the in vitro transcribed 16S rRNA molecule of *P. putida* is a MM target to the Gam42a probe. Of note, the dissociation curves of the Gam42a probe hybridized with in vitro transcribed 16S rRNA of uncontaminated and TNT-contaminated soil samples were also significantly different than the dissociation curve obtained with native *P. putida* RNA (MAX_{DCSD} value of 0.428 and 0.441, respectively).

Comparison of 16S rRNA profiles of uncontaminated and TNT-contaminated soils. Significant differences in the nucleic acids amplified from the uncontaminated and TNT-contaminated soil samples were detected when comparing initial signal intensity and dissociation curves. Initial signal intensities showed that several domain-, group-, and species-specific PM probes were significantly different between in vitro transcribed 16S rRNA of uncontaminated and TNT-contaminated soils (Figure 2). The signal intensity of ten PM probes was significantly lower with the TNT-contaminated soil than the uncontaminated soil. The signal from the species-specific probe Bacil1136 was significantly higher in the TNT-contaminated soil. Significant differences in signal intensities were also detected for probes Bet42a and Gam42a, which target the 23S rRNA molecule and for probes Euk502, Eury498, and Cren499 which target small subunit rRNA sequences from *Eukarya* and *Archaea*. Clearly these results were cross-hybridization events since these hybridizations were realized with in vitro transcribed 16S rRNA amplified with primers designed to the *Bacteria* domain. Indeed, the cross hybridizations with the Gam42a probe were confirmed with native RNA of *P. putida* (see above); however, cross hybridization to the other probes was not verified with microorganisms having a PM target sequence.

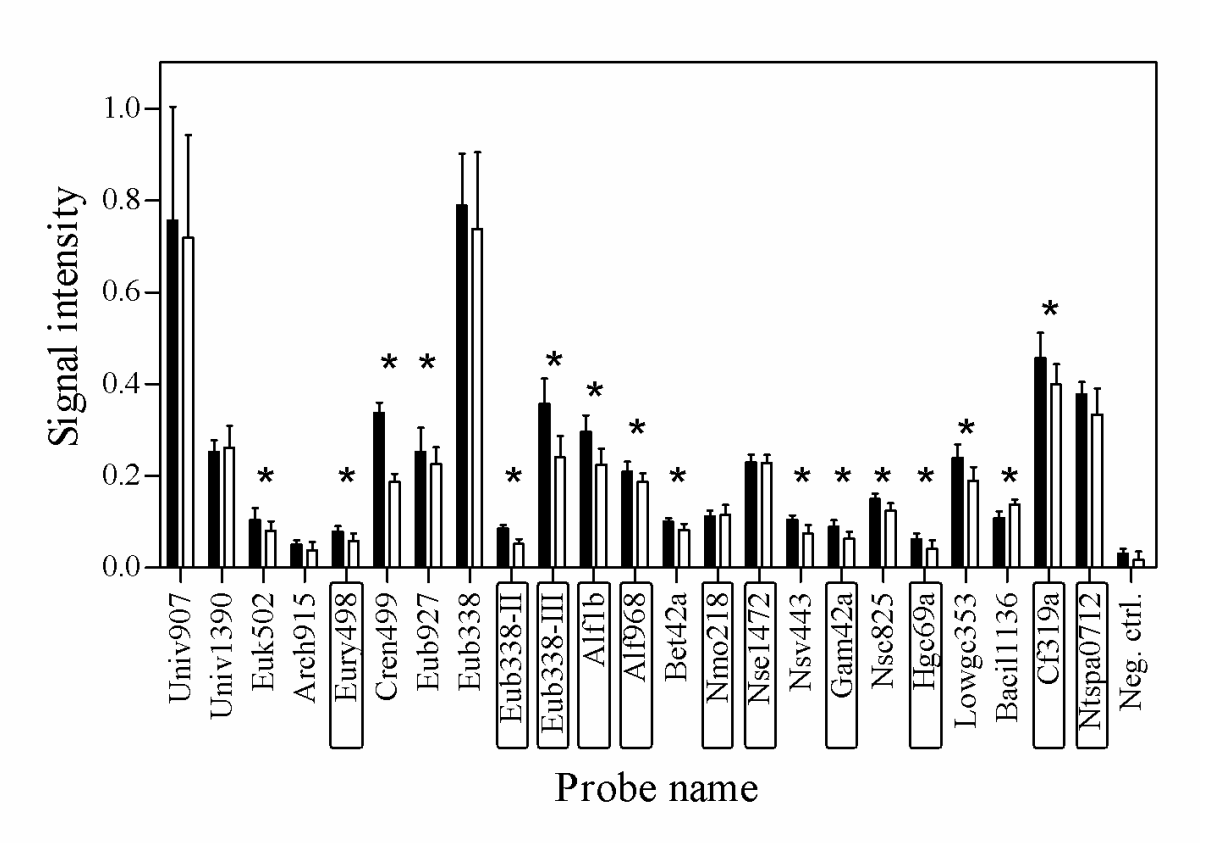


Figure 2. Initial signal intensities from hybridizations with in vitro transcribed 16S rRNA of uncontaminated and TNT-contaminated soil samples to gel-pad microarrays (data of Table S-2). Significant differences in signal intensity between uncontaminated (black bars) and contaminated (white bars) soil samples are marked with asterisks. Error bars represent the standard deviation. Probes with significantly different dissociation curves (data of Table 1) when hybridized with the two soil samples are indicated by a rectangle. Probe names are listed in Table S-1. The PM probes Gpos1192 and Gpos1200 could not be discriminated from their respective c-MM and g-MM with the uncontaminated and/or contaminated soil samples with the functional ANOVA calculator (Table 1). These two probes were not considered for comparison of 16S rRNA profiles of the uncontaminated and TNT-contaminated soil samples.

Dissociation curves of PM probes hybridized to in vitro transcribed 16S rRNA of the uncontaminated and TNT-contaminated soil samples further delineated differences between the target nucleic acids. Eleven PM probes had significantly different dissociation curves between the two soil samples with functional ANOVA (Figure 2; Table 1). Among these 11 probes, no differences in initial signal intensities between the two soil samples were detected for 3 of them (Table S-2), and only 7 of these 11 had significantly different T_ds (Table S-3). This suggested that comparison of signal intensities and individual points in the dissociation curve were ineffective in detecting differences in the sequence of target nucleic acid pools

originating from complex mixtures of unknown microorganisms in unknown abundances. On the other hand, functional ANOVA detected more differences between the dissociation curves and provided a statistical robust method of analysis. Functional ANOVA of dissociation curves of PM and respective MM probe sets revealed subtle differences between the contaminated and uncontaminated sample target sequences (Table 1). Eleven PM/MM probe sets were detected as significantly different with one of the soil samples but not the other one. In addition, two PM/MM probe sets (i.e., Alf968/Alf968-g7c and Alf968/Alf968-c13g) had temperatures more than 5 °C different between the two soil samples at the point of the MAX_{DCSD} (Table 1).

Several parameters may explain the observed differences between the two soil samples: presence or absence of TNT, spatial heterogeneity (e.g., pH, granulometry and carbon content), and PCR bias. Several authors (*19, 20*) have demonstrated that PCR-amplification bias is low when the sample has diverse templates. In another study (*11*), 16S rRNA gene PCR amplicons from the same soil samples were analyzed with denaturing gradient gel electrophoresis (DGGE). The number of bands identified in these two soil samples (more than 15 bands) indicated that the samples contained highly diverse templates. Therefore, we suggest that the influence of PCR bias was probably small. The position of the bands differed between the two soil samples, and three brighter bands were present in the TNT-contaminated but not the uncontaminated soil sample (*11*). These results corroborate the microarray results that differences in amplicon sequences existed between the two soil samples. DGGE was also used to characterize 9 additional soils sampled at the same TNT-destruction field (George et al., manuscript in preparation). Three to seven brighter bands belonging to *Proteobacteria* were present in all TNT-contaminated soils but not in the uncontaminated soils. Several characteristics of these soils were investigated (pH, granulometry, and carbon content), but the differences in DGGE fingerprints between uncontaminated and TNT-contaminated soil samples were explained by the absence or presence of TNT. The influence of TNT in soil microbial population was also supported by additional results of George et al. (manuscript in preparation) who artificially contaminated a pristine soil with TNT and found a rapid selection of *Pseudomonadaceae*.

Fuller and Manning (*21*) showed that TNT was more toxic to Gram-positive than Gram-negative bacteria. A significant decrease in signal intensity for the probes Hgc69a and Lowgc353 targeting Gram-positive bacteria was observed for the contaminated soil sample

(Figure 2; Table S-2), although the magnitude of this difference was relatively small. The significant increase in signal intensity from the probe targeting *Bacillus* (Bacil1136) in the contaminated soil and the difference in the dissociation curves for the Hgc69a probe between the uncontaminated and contaminated soil sample further suggests a possible shift within the Gram positive community. However, detection of specific microbial populations sensitive or resistant to the toxicity of TNT is limited by the relatively small number of specific probes (genus or species) on this prototype microarray. Ultimately, improvements in microarray technology will be needed to not only identify differences among samples (e.g., how differences in amplicon diversity and abundance influence signal intensities and dissociation curves), but also to provide sufficient sequence information to identify specific target sequences, and to determine appropriate stringency temperatures to compare signal intensities of different environmental samples.

Previous studies have documented the general utility of using dissociation analysis to provide additional information about the sequence composition of RNA hybridized to microarrays. However, those studies had not presented a fully elaborated experimental and statistical approach for the generation and analysis of dissociation data sets. This study has more fully developed the experimental and analytical approach. Functional ANOVA was shown to resolve differences between most perfect-match and single mismatch probe sets, providing for much greater resolution of microbial communities inhabiting TNT contaminated and uncontaminated soils than could not be achieved by using either differences in signal intensity or T_d. Experimental replication demonstrated that dissociation curves were highly reproducible (e.g., probe Eub338, Figure 1C). The ability to clearly resolve single nucleotide mismatch variants and to identify dissociation curves that conform with a specific target reference are essential to achieving the ultimate analytical goal of a robust assignment of identity to nucleic acid hybridized to each array element. Although as yet there is no established theoretical or analytical framework for using DNA microarrays to fully resolve complex and undefined mixtures of nucleic acid sequences, we anticipate that the development of a robust statistical framework that captures the additional information of duplex composition available using dissociation analysis will provide a critical foundation for advancing this important area of environmental biotechnology.

Acknowledgements

This work was supported by a short-term visiting fellowship from the Belgian French Community to L.E. Additional support was received from NIH/NIDCR (U01 DE14955) and US DARPA (DABT63-99-1-0009) to D.A.S., and NASA (MSMT-2004-0045-0066) to D.A.S. and J.C.S. We thank J.L. Ramos for providing *P. putida* sp. strain JRL11, E.M. Starke, P.A. Noble and A. Pozhitkov for critical review of the manuscript. We also thank J. Jackman and M. Theodore of Johns Hopkins University Applied Physics Laboratory for assistance throughout this study.

Supporting Information

PCR amplification of the 16S rRNA gene; fragmentation and labeling of RNA samples; description of microarray formats and design of PM and MM probes; recording of signal intensities and generation of dissociation curves; description of unpaired tests to compare signal intensities and T_d values; characteristics of PM and MM probes (Table S-1); initial signal intensities of PM and MM probes (Table S-2); T_d values of PM and MM probes (Table S-3); and images of representative hybridizations (Figure S-1). This material is available free of charge via the Internet at http://pubs.acs.org.

4.5. Literature cited

(1) Keramas, G.; Bang, D. D.; Lund, M.; Madsen, M.; Rasmussen, S. E.; Bunkenborg, H.; Telleman, P.; Christensen, C. B. V. Development of a sensitive DNA microarray suitable for rapid detection of *Campylobacter* spp. *Mol. Cell. Probes* **2003,** *17*, 187-196.

(2) Zhou, J. Microarrays for bacterial detection and microbial community analysis. *Curr. Opin. Microbiol.* **2003,** *6*, 288-294.

(3) Loy, A.; Lehner, A.; Lee, N.; Adamczyk, J.; Meier, H.; Ernst, J.; Schleifer, K.; Wagner, M. Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl. Environ. Microbiol.* **2002,** *68*, 5064-5081.

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

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

(6) El Fantroussi, S.; Urakawa, H.; Bernhard, A. E.; Kelly, J. J.; Noble, P. A.; Smidt, H.; Yershov, G. M.; Stahl, D. A. Direct profiling of environmental microbial populations by thermal dissociation analysis of native rRNAs hybridized to oligonucleotide microarrays. *Appl. Environ. Microbiol.* **2003,** *69*, 2377-2382.

(7) Reyes-Lopez, M. A.; Mendez-Tenorio, A.; Maldonado-Rodriguez, R.; Doktycz, M. J.; Fleming, J. T.; Beattie, K. L. Fingerprinting of prokaryotic 16S rRNA genes using oligodeoxyribonucleotide microarrays and virtual hybridization. *Nucleic Acids Res.* **2003,** *31*, 779-789.

(8) Liu, W. T.; Mirzabekov, A. D.; Stahl, D. A. Optimization of an oligonucleotide microchip for microbial identification studies: a non-equilibrium dissociation approach. *Environ. Microbiol.* **2001,** *3*, 619-629.

(9) Urakawa, H.; El Fantroussi, S.; Smidt, H.; Smoot, J. C.; Tribou, E. H.; Kelly, J. J.; Noble, P. A.; Stahl, D. A. Optimization of single-base-pair mismatch discrimination in oligonucleotide microarrays. *Appl. Environ. Microbiol.* **2003,** *69*, 2848-2856.

(10) Fuller, M. E.; Manning, J. F. Evidence for differential effects of 2,4,6-trinitrotoluene and other munitions compounds on specific subpopulations of soil microbial communities. *Environ. Toxicol. Chem.* **1998,** *17*, 2185-2195.

(11) Eyers, L.; Stenuit, B.; El Fantroussi, S.; Agathos, S. N. Microbial ecology of TNT-contaminated soils and anaerobic TNT biodegradation processes. In *Proceedings of the 103rd general meeting of the American Society for Microbiology*; Washington, DC, 2003.

(12) Esteve-Nunez, A.; Ramos, J. L. Metabolism of 2,4,6-trinitrotoluene by *Pseudomonas* sp. JLR11. *Environ. Sci. Technol.* **1998,** *32*, 3802-3808.

(13) Maniatis, T.; Fritsch, E. F.; Sambrook, J., *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: New York, 1982.

(14) Bavykin, S. G.; Akowski, J. P.; Zakhariev, V. M.; Barsky, V. E.; Perov, A. N.; Mirzabekov, A. D. Portable system for microbial sample preparation and oligonucleotide microarray analysis. *Appl. Environ. Microbiol.* **2001,** *67*, 922-928.

(15) Bugli, C.; Lambert, P. Functional ANOVA with random functional effects: an application to event-related potentials modelling for electroencephalograms analysis. *Stat. Med.* **2006**, in press.

(16) Hacia, J. G.; Brody, L. C.; Chee, M. S.; Fodor, M. S.; Collins, F. S. Detection of heterozygous mutations in *BRCA1* using high density oligonucleotides arrays and two-colour fluorescent analysis. *Nat. Genet.* **1996,** *14*, 441-447.

(17) Pozhitkov, A.; Chernov, B.; Yershov, G.; Noble, P. A. Evaluation of gel-pad oligonucleotide microarray technology by using artificial neural networks. *Appl. Environ. Microbiol.* **2005,** *71*, 8663-8676.

(18) White, A. M.; Daly, D. S.; Willse, A. R.; Protic, M.; Chandler, D. P. Automated microarray image analysis toolbox for matlab. *Bioinformatics* **2005,** *21*, 3578-3579.

(19) Suzuki, M. T.; Giovannoni, S. J. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl. Environ. Microbiol.* **1996,** *62*, 625-630.

(20) Polz, M. F.; Cavanaugh, C. M. Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* **1998,** *64*, 3724-3730.

(21) Fuller, M. E.; Manning, J. F. Aerobic Gram-positive and Gram-negative bacteria exhibit differential sensitivity to and transformation of 2,4,6-trinitrotoluene (TNT). *Curr. Microbiol.* **1997,** *35*, 77-83.

4.6. Supporting information

Discrimination of shifts in a soil microbial community associated with TNT-contamination using a functional ANOVA of 16S rRNA hybridized to oligonucleotide microarrays

LAURENT EYERS,[1] JAMES C. SMOOT,[2] LAURA M. SMOOT,[2] CÉLINE BUGLI,[3] HIDETOSHI URAKAWA,[4] ZACK MCMURRY,[2] SLIL SIRIPONG,[5] SAID EL FANTROUSSI,[1] PHILIPPE LAMBERT,[3] SPIROS N. AGATHOS,[1] AND DAVID A. STAHL[2,*]

[1]*Unit of Bioengineering, University of Louvain, 1348 Louvain-la-Neuve, Belgium,* [2]*Civil and Environmental Engineering, University of Washington, Seattle, Washington 98195,* [3]*Institute of Statistics, University of Louvain, 1348 Louvain-la-Neuve, Belgium,* [4]*Ocean Research Institute, University of Tokyo, Tokyo 164-8639, Japan,* [5]*Civil and Environmental Engineering, Northwestern University, Evanston, Illinois 60208.*

Content

Total number of pages: 13

Materials and Methods

Sample collection and processing of nucleic acids. PCR amplification of the 16S rRNA gene was done with primers BACT 11 F (*S1*) containing the T7 promoter sequence and S-D-Bact-1512-a-A-16 (*S2*) in a thermal cycler (Thermo Hybaid US, Franklin, MA) in 100 µl reactions under the following conditions (reagents were from Invitrogen (Carlsbad, CA)): 1 X PCR buffer, 2 µl of deoxynucleoside triphosphate mixture (2.5 mM each), 2.5 µM of each primer, 5 U of *Taq* DNA polymerase, and 100 ng of template DNA. An initial denaturation step of 3 min at 94°C was followed by 30 cycles of 30 s at 94°C, 30 s at 55°C, and 1 min at 72°C. A final extension at 72°C for 7 min was also done.

Fragmentation and labeling of RNA. RNA was fragmented in solution rather than on-column. Fragmentation and labeling of RNA samples of uncontaminated and TNT-contaminated soils was done in parallel using the same freshly prepared reactants. In brief, RNA (30 µg) was suspended in 20 µl of diethyl pyrocarbonate (DEPC)-treated water and incubated at 95°C for 4 min. The RNA solution was mixed with 122 µl of preheated (95ºC for 30 s), filtered (0.22 µm) fragmentation and labeling solution that contained 700 µl of 25 mM sodium phosphate buffer (pH 7.0), 30 µl of 150 mM 1,10-phenanthroline, 3 µl of 150 mM $CuSO_4$, and 6 µl of 100 mg/ml lissamine rhodamine B ethylenediamine (Molecular Probes, Inc., Eugene, OR). Three microliters of 200 mM H_2O_2 and 5 µl of 600 mM $NaCNBH_3$ were added, and the mixture was incubated at 95°C for 10 min. The reaction was stopped with 540 µl of cold 100% ethanol and 20 µl of stop solution (2.6 M sodium acetate and 68 mM EDTA, pH 8.0). Following precipitation at -80°C for 10 min, RNA was centrifuged and washed three times with 100% ethanol. The RNA pellet was suspended in 20 µl of DEPC-treated water and stored before use at -80°C.

Microarray fabrication and hybridization. The CodeLink microarray contained oligonucleotide probes deposited on CodeLink slides of uniformly coated hydrophilic polymer containing amine-reactive groups (Amersham Biosciences, Buckinghamshire, UK). Oligonucleotide probes were manufactured by Argonne National Laboratory and Qiagen, Inc., and approximately 6 fmol of each probe was printed in duplicate with a SpotBot arrayer (Telechem International, Inc., Sunnyvale, CA). Probes were covalently attached to the slide via a 3' C7 primary amine linker. Probe attachment, slide hydration, and blocking were done as described by the slide manufacturer's instructions. Gel-pad microarrays, which were fabricated at Argonne National Laboratory, consisted of 100 x 100 x 20 µm polyacrylamide

gel pads photopolymerized to a glass slide with a center-to-center distance of 300 μm (*S3*). After activation of gel pads (*S4*), oligonucleotide probes were spotted (2 pmol deposited) and immobilized on each gel pad by Schiff base coupling to aldehyde groups within the gel pads (*S5*). The coupling was stabilized by reduction with $NaCNBH_3$ (*S4*). The design of oligonucleotide probes targeting rRNA molecules was done as described previously (*S6*). Design of single MM probes was based on several criteria: i) GC content was not changed (e.g., substitution of G to C (c-MM probes) or C to G (g-MM probes)), ii) mismatches at terminal positions were avoided (*S7*), iii) non-complementarity of MM probes to any microorganism was checked by comparing MM probe sequences to the GenBank database (http://www.ncbi.nih.gov/Genbank), and iv) approximately 4°C differences in theoretical T_ds between PM and c-MM and about 6°C differences between PM and g-MM were chosen. However, these are general rules, and often they cannot all be met when designing a probe. For example, many of the MM probes to the general PM probes cannot simultaneously retain the GC content of the PM probe and be a mismatch to all sequences in the databases. Theoretical differences in T_ds were calculated with MeltCalc software (http://www.meltcalc.de). Several of the PM probes used in this study were validated by El Fantroussi et al. (*S6*) and Kelly et al. (*S8*) with PM and MM reference RNA and RNA from environmental samples.

Analysis of images and thermal dissociation curves. Fluorescence signal intensities were recorded for each probe on the microarray with a custom epifluorescence microscope and software (*S3*). Dissociation curves were generated by increasing the temperature of the microarray (1°C/min) from 20°C to 68°C and recording signal intensities every 1°C or 2°C (*S7*). The hybridization signals were corrected by subtracting the local background signal and normalizing the corrected hybridization signal to the background (*S9*). To compare initial signal intensities of hybridizations of in vitro transcribed 16S rRNA derived from uncontaminated and TNT-contaminated soil samples, no normalization was carried out since the same amount of RNA (5 μg) was used for each hybridization. Liu et al. (*S10*) showed that a temperature correction was necessary to compare dissociation curves collected with different dyes. In our study, we used a single dye to label all samples, thus preventing the need of temperature correction.

Unpaired Student's *t* tests were used to compare initial signal intensities between probes, as most of the data followed a normal distribution (146 of 152 signal intensity measurements

from PM and MM probes hybridized with in vitro transcribed RNA of the two soil samples showed a normal distribution). In the case of T_d values, some of the data (41 of 152 T_d measurements) did not follow a normal distribution, especially with in vitro transcribed RNA of the contaminated soil sample (35 of the 41 T_d measurements did not have a normal distribution). Unpaired Wilcoxon and Student's *t*-tests were done to compare T_d values. The results obtained (i.e. hypothesis of equality of compared means accepted or rejected) using Wilcoxon or Student's *t*-tests were similar, except for 6 of 126 comparisons of T_d values between PM probes hybridized with RNA from different soil samples, and between PM and MM probes hybridized with RNA from the same soil sample. Unpaired tests were performed with an α value of 0.05.

Literature Cited

(S1) Kane, M. D.; Poulsen, L. K.; Stahl, D. A. Monitoring the enrichment and isolation of sulfate-reducing bacteria by using oligonucleotide hybridization probes designed from environmentally derived 16S rRNA sequences. *Appl. Environ. Microbiol.* **1993,** *59*, 682-686.

(S2) Fry, N. K.; Frederickson, J. K.; Fishbain, S.; Wagner, M.; Stahl, D. A. Population structure of microbial communities associated with two deep, anaerobic, alkaline aquifers. *Appl. Environ. Microbiol.* **1997,** *63*, 1498-1504.

(S3) Guschin, D. Y.; Yershov, G. M.; Zaslavsky, A.; Gemmell, A.; Shick, V.; Proudnikov, D.; Arenkov, P.; Mirzabekov, A. D. Manual manufacturing of oligonucleotide, DNA, and protein microchips. *Anal. Biochem.* **1997,** *250*, 203-211.

(S4) Proudnikov, D.; Timofeev, E.; Mirzabekov, A. D. Immobilization of DNA in polyacrylamide gel for the manufacture of DNA and DNA-oligonucleotide microchips. *Anal. Biochem.* **1998,** *259*, 34-41.

(S5) Bavykin, S. G.; Akowski, J. P.; Zakhariev, V. M.; Barsky, V. E.; Perov, A. N.; Mirzabekov, A. D. Portable system for microbial sample preparation and oligonucleotide microarray analysis. *Appl. Environ. Microbiol.* **2001,** *67*, 922-928.

(S6) El Fantroussi, S.; Urakawa, H.; Bernhard, A. E.; Kelly, J. J.; Noble, P. A.; Smidt, H.; Yershov, G. M.; Stahl, D. A. Direct profiling of environmental microbial populations by thermal dissociation analysis of native rRNAs hybridized to oligonucleotide microarrays. *Appl. Environ. Microbiol.* **2003,** *69*, 2377-2382.

(S7) Urakawa, H.; Noble, P. A.; El Fantroussi, S.; Kelly, J. J.; Stahl, D. A. Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses. *Appl. Environ. Microbiol.* **2002,** *68*, 235-244.

(S8) Kelly, J. J.; Siripong, S.; McCormack, J.; Janus, L. R.; Urakawa, H.; El Fantroussi, S.; Noble, P. A.; Sappelsa, L.; Rittmann, B. E.; Stahl, D. A. DNA microarray detection of nitrifying bacterial 16S rRNA in wastewater treatment plant samples. *Water Res.* **2005,** *39*, 3229-3238.

(S9) Fotin, A. V.; Drobyshev, A. L.; Proudnikov, D. Y.; Perov, A. N.; Mirzabekov, A. D. Parallel thermodynamic analysis of duplexes on oligodeoxyribonucleotide microchips. *Nucleic Acids Res.* **1998,** *26*, 1515-1521.

(S10) Liu, W. T.; Wu, J. H.; Li, E. S.; Selamat, E. S. Emission characteristics of fluorescent labels with respect to temperature changes and subsequent effects on DNA microchip studies. *Appl. Environ. Microbiol.* **2005,** *71*, 6453-6457.

Table S-1. Characteristics of PM and respective MM probes printed on the gel-pad microarray

Probe name	Sequence (5' to 3')	% G+C	Target group	Location(s) on the array [τ]	c-mismatch probe [¥]	Location(s) on the array [τ]	g-mismatch probe [¥]	Location(s) on the array [τ]
Univ907*	CCCCGTCAATTCCTTTGAGTTT	46	All Life	C3, O3, J5, J12, J17	Univ907-a9c	L5, L12, L17	Univ907-a18g	K5, K12, K17
Univ1390*	GACGGGCGGTGTGTACAA	61	All Life	B3, N3, G5, G12, G17	Univ1390-g13c	I5, I12, I17	Univ1390-t10g	H5, H12, H17
Euk502	ACCAGACTTGCCCTCC	63	*Eukarya*	A11, M12, A17, M17	Euk502-a4c	C11, O12, C17, O17	Euk502-c12g	B11, N12, B17, N17
Arch915	GTGCTCCCCCGCCAATTCCT	65	*Archaea*	D10, P11, D16, P16	Arch915-g3c	F10, R11, F16, R16	Arch915-c9g	E10, Q11, E16, Q16
Eury498	CTTGCCCRGCCCTT	75	*Euryarchaeota*	G10, G15	Eury498-g4c	I10, I15	Eury498-c6g	H10, H15
Cren499	CCAGRCTTGCCCCCCGCT	75	*Crenarchaeota*	J10, J15	Cren499-g4c	L10, L15	Cren499-c11g	K10, K15
Eub927*	ACCGCTTGTGCGGGCCC	77	*Bacteria*	D3, P3, A5, M5, D11, P12, D17, P17	Eub927-g8c	C5, O5, F11, R12, F17, R17	Eub927-t9g	B5, N5, E11, Q12, E17, Q17
Eub338*	GCTGCCTCCCGTAGGAGT	67	*Bacteria*	D5, P5, G11, G16	Eub338-g11c	F5, R5, I11, I16	Eub338-t7g	E5, Q5, H11, H16
Eub338-II*	GCAGCCACCCGTAGGTGT	67	*Planctomycetes*	G4, J11, J16	Eub338-II-g11c	I4, L11, L16	Eub338-II-c9g	H4, K11, K16
Eub338-III*	GCTGCCACCCGTAGGTGT	67	*Verrucomicrobia*	J4, A10, M11, A16, M16	Eub338-III-g11c	L4, C10, O11, C16, O16	Eub338-III-c9g	K4, B10, N11, B16, N16
Alf1b	CGTTCGYTCTGAGCCAG	53	α-*Proteobacteria*	A4, M4, D8, P9, D14, P14	Alf1b-t10c	C4, O4, F8, R9, F14, R14	Alf1b-c9g	B4, N4, E8, Q9, E14, Q14
Alf968*	GGTAAGGTTCTGCGCGTT	56	α-*Proteobacteria*	D4, P4, G8, G13	Alf968-g7c	F4, R4, I8, I13	Alf968-c13g	E4, Q4, H8, H13
Bet42a*	GCCTTCCCACTTCGTTT	53	β-*Proteobacteria*, 23S	G3, J8, J13	Bet42a-g14c	I3, L8, L13	Bet42a-c8g	H3, K8, K13
Nso1225	CGCCATTGTATTACGTGTGA	45	β-ammonia-oxidizers	J2	Nso1225-g8c	L2	Nso1225-c14g	K2
Nso190*	CGATCCCCTGCTTTTCTCC	58	β-ammonia-oxidizers	G2	Nso190-g10c	I2	Nso190-c11g	H2
Nsm653*	CCCCTCTGCTGCACTCTA	61	*Nitrosomonas* spp.	J1	Nsm653-g11c	L1	Nsm653-c12g	K1
Nmo218	CGGCCGCTCCAAAAGCAT	61	*Nitrosomonas oligotropha*	D1, P1	Nmo218-g6c	F1, R1	Nmo218-c10g	E1, Q1
Nse1472	ACCCCAGTCATGACCCCC	67	*Nitrosomonas europaea*	A1, M1	Nse1472-g7c	C1, O1	Nse1472-c15g	B1, N1
Nsm156*	TATTAGCACATCTTTCGAT	32	*Nitrosomonas* spp.	G1	Nsm156-g17c	I1	Nsm156-c12g	H1
Nsv443*	CCGTGACCGTTTCGTTCCGG	63	*Nitrosospira* spp.	D2, P2	Nsv443-g9c	F2, R2	Nsv443-c7g	E2, Q2
Gam42a*	GCCTTCCCACATCGTTT	53	γ-*Proteobacteria*, 23S	J3, A7, M8, A13, M13	Gam42a-a11c	L3, C7, O8, C13, O13	Gam42a-c6g	K3, B7, N8, B13, N13
Nsc825	CCCTCCCAACGTCTAGTT	56	γ-ammonia-oxidizers	A2, M2	Nsc825-g11c	C2, O2	Nsc825-c10g	B2, N2
Gpos1192	TAAGGGGCATGATGATTTGACG	46	Gram-positive bacteria	A9, M10, A15, M15	Gpos1192-g14c	C9, O10, C15, O15	Gpos1192-t10g	B9, N10, B15, N15
Gpos1200	AAGGGGCATGATG	54	Gram-positive bacteria	G6, G7	Gpos1200-g5c	I6, I7	Gpos1200-t9g	H6, H7
Hgc69a	TATAGTTACCACCGCCGT	47	*Actinobacteria*	D9, P10, D15, P15	Hgc69a-g5c	F9, R10, F15, R15	Hgc69a-c9g	E9, Q10, E15, Q15
Lowgc353*	GGAAGATTCCCTACTGCT	50	Low G+C Gram-positive bacteria	G9, G14	Lowgc353-a6c	I9, I14	Lowgc353-c9g	H9, H14
Bacil1136*	AGTCACCTTAGAGTGCCCAACT	50	*Bacillus*	N6, N7	Bacil1136-g13c	P6, P7	Bacil1136-c17g	O6, O7
Cf319a	TGGTCCGTGTCTCAGTAC	56	*Cytophaga-Flavobacteria*	A8, M9, A14, M14	Cf319a-g9c	C8, O9, C14, O14	Cf319a-c13g	B8, N9, B14, N14
Ntspa0712	CGCCTTCGCCACCGGCCTTCC	76	*Nitrospira*	J9, J14	Ntspa0712-g8c	L9, L14	Ntspa0712-c10g	K9, K14
Neg. ctrl.	GATGATGATGATGA	35	Negative control	E3, Q3, E7, Q8, E13, Q13	-		-	

* These probes were also spotted on CodeLink microarrays. [τ] Indicates position as show non Figure S-1. [¥] Descriptive names of MM probes are given. Nomenclature following the hyphen indicates the nucleotide change and the position of MM from 5'-end.

Table S-2. Initial signal intensities of PM and MM probes*

Probe name	Uncontaminated sample						Contaminated sample					
	PM		c-MM		g-MM		PM		c-MM		g-MM	
	Signal intensity ± SD	*n* τ	Signal intensity ± SD	*n*	Signal intensity ± SD	*n*	Signal intensity ± SD	*n*	Signal intensity ± SD	*n*	Signal intensity ± SD	*n*
Univ907	0.756 ± 0.248	15	0.714 ± 0.124	9	0.746 ± 0.114	9	0.719 ± 0.224	15	**0.561 ± 0.112**	9	0.634 ± 0.151	9
Univ1390	0.253 ± 0.024	15	**0.209 ± 0.035**	9	**0.104 ± 0.0.20**	9	0.262 ± 0.047	15	**0.201 ± 0.038**	9	**0.121 ± 0.020**	9
Euk502	0.104 ± 0.026	12	0.104 ± 0.017	12	0.087 ± 0.026	12	0.081 ± 0.020	12	0.072 ± 0.020	12	0.065 ± 0.019	12
Arch915	0.050 ± 0.010	12	**0.490 ± 0.060**	12	**0.416 ± 0.043**	12	0.038 ± 0.018	12	**0.305 ± 0.047**	12	**0.318 ± 0.052**	12
Eury498	0.079 ± 0.012	6	**0.111 ± 0.013**	6	**0.061 ± 0.013**	6	0.059 ± 0.016	6	0.078 ± 0.018	6	0.058 ± 0.016	6
Cren499	0.338 ± 0.022	6	**0.442 ± 0.033**	6	**0.162 ± 0.013**	6	0.187 ± 0.018	6	**0.252 ± 0.013**	6	**0.143 ± 0.016**	6
Eub927	0.253 ± 0.052	24	**0.076 ± 0.013**	18	**0.141 ± 0.019**	18	0.226 ± 0.037	24	**0.062 ± 0.016**	18	**0.118 ± 0.024**	18
Eub338	0.789 ± 0.113	12	**0.567 ± 0.080**	12	**0.590 ± 0.073**	12	0.738 ± 0.167	12	**0.427 ± 0.079**	12	**0.470 ± 0.076**	12
Eub338-II	0.086 ± 0.007	9	**0.065 ± 0.008**	9	**0.059 ± 0.006**	9	0.052 ± 0.011	9	0.048 ± 0.006	9	0.044 ± 0.009	9
Eub338-III	0.357 ± 0.055	15	0.395 ± 0.055	15	**0.072 ± 0.011**	15	0.241 ± 0.046	15	**0.291 ± 0.038**	15	**0.046 ± 0.013**	15
Alf1b	0.296 ± 0.036	18	**0.113 ± 0.027**	18	**0.094 ± 0.017**	18	0.225 ± 0.034	18	**0.091 ± 0.028**	18	**0.072 ± 0.015**	18
Alf968	0.209 ± 0.022	12	**0.049 ± 0.009**	12	**0.090 ± 0.010**	12	0.187 ± 0.019	12	**0.056 ± 0.013**	12	**0.096 ± 0.016**	12
Bet42a	0.101 ± 0.007	9	**0.113 ± 0.010**	9	**0.065 ± 0.008**	9	0.083 ± 0.012	9	0.084 ± 0.007	9	**0.055 ± 0.013**	9
Nmo218	0.112 ± 0.013	6	**0.147 ± 0.019**	6	0.095 ± 0.013	6	0.115 ± 0.022	6	0.119 ± 0.014	6	**0.075 ± 0.016**	6
Nse1472	0.230 ± 0.017	6	**0.182 ± 0.015**	6	**0.160 ± 0.011**	6	0.228 ± 0.018	6	**0.131 ± 0.009**	6	**0.160 ± 0.009**	6
Nsv443	0.105 ± 0.009	6	**0.088 ± 0.014**	6	**0.054 ± 0.009**	6	0.075 ± 0.019	6	0.070 ± 0.016	6	**0.036 ± 0.013**	6
Gam42a	0.089 ± 0.015	15	**0.189 ± 0.038**	15	**0.062 ± 0.014**	15	0.064 ± 0.014	15	**0.139 ± 0.028**	15	**0.042 ± 0.015**	15
Nsc825	0.150 ± 0.012	6	**0.133 ± 0.013**	6	0.147 ± 0.009	6	0.124 ± 0.018	6	**0.089 ± 0.013**	6	0.127 ± 0.009	6
Gpos1192	0.074 ± 0.012	12	**0.051 ± 0.009**	12	**0.055 ± 0.013**	12	0.055 ± 0.011	12	**0.040 ± 0.011**	12	**0.044 ± 0.012**	12
Gpos1200	0.046 ± 0.006	6	**0.034 ± 0.007**	6	0.042 ± 0.005	6	0.033 ± 0.010	6	0.027 ± 0.008	6	0.031 ± 0.013	6
Hgc69a	0.062 ± 0.013	12	0.057 ± 0.012	12	**0.043 ± 0.011**	12	0.041 ± 0.019	12	0.043 ± 0.018	12	0.033 ± 0.017	12
Lowgc353	0.240 ± 0.028	6	**0.133 ± 0.011**	6	**0.051 ± 0.008**	6	0.190 ± 0.030	6	**0.111 ± 0.018**	6	**0.040 ± 0.014**	6
Bacil1136	0.109 ± 0.014	6	**0.054 ± 0.014**	6	**0.050 ± 0.004**	6	0.138 ± 0.011	6	**0.046 ± 0.012**	6	**0.034 ± 0.013**	6
Cf319a	0.457 ± 0.055	12	**0.080 ± 0.013**	12	**0.171 ± 0.021**	12	0.399 ± 0.044	12	**0.053 ± 0.016**	12	**0.131 ± 0.019**	12
Ntspa0712	0.378 ± 0.027	6	**0.503 ± 0.053**	6	**0.159 ± 0.017**	6	0.334 ± 0.056	6	0.380 ± 0.037	6	**0.091 ± 0.017**	6
Neg. ctrl.	0.031 ± 0.011	18	na		na		0.017 ± 0.018	18	na		na	

* Values shown are average initial signal intensities ± standard deviation of triplicate hybridizations to gel-pad microarrays. Signal intensities of MM probes significantly different (Student's *t*-tests) than the ones of corresponding PM probes are indicated in bold. Shaded zones correspond to signal intensities of PM probes significantly different (Student's *t*-tests) when hybridized with uncontaminated or TNT-contaminated soil. Probes Nso1225, Nso190, Nsm653 and Nsm156 were not analyzed with the functional ANOVA calculator because of the low number of observations (n<4). τ Number of hybridized microarray elements. na: not applicable.

Table S-3. T_d values of PM and MM probes*

Probe name	Uncontaminated sample						Contaminated sample					
	PM		c-MM		g-MM		PM		c-MM		g-MM	
	T_d ± SD	n^{τ}	T_d ± SD	n	T_d ± SD	n	T_d ± SD	n	T_d ± SD	n	T_d ± SD	n
Univ907	38.5 ± 1.3	12	**36.2 ± 1.2**	9	**37.2 ± 1.4**	9	38.6 ± 2.2	15	37.8 ± 0.9	7	38.9 ± 1.2	7
Univ1390	39.2 ± 1.8	13	**37.8 ± 1.0**	9	38.6 ± 2.1	9	40.2 ± 2.5	14	**38.6 ± 0.5** ¥	7	40.5 ± 1.5	7
Euk502	36.6 ± 1.7	12	**32.9 ± 0.7**	11	**38.0 ± 1.4**	11	36.0 ± 3.5	11	**32.2 ± 2.5**	11	37.6 ± 4.3 ¥	11
Arch915	38.3 ± 2.7	12	**33.0 ± 0.6**	12	**36.1 ± 0.7**	11	37.0 ± 3.0	7	**32.5 ± 2.6**	12	35.7 ± 3.2	11
Eury498	35.7 ± 1.4	6	**31.9 ± 0.6**	6	**33.6 ± 1.1**	5	36.9 ± 3.9	5	**30.6 ± 1.9** ¥	6	32.2 ± 2.5	5
Cren499	32.9 ± 0.4	6	33.2 ± 0.4	6	**35.0 ± 0.9**	6	32.7 ± 1.4	5	33.2 ± 1.7	5	34.5 ± 3.1	5
Eub927	40.0 ± 1.8	20	**38.2 ± 1.6**	17	**38.7 ± 1.7**	17	40.1 ± 2.1	24	**39.0 ± 0.9**	14	40.2 ± 1.4	14
Eub338	40.5 ± 3.1	12	**36.7 ± 1.3**	12	39.1 ± 2.1	12	41.4 ± 1.7	10	**37.9 ± 1.3**	10	40.8 ± 1.8	10
Eub338-II	35.7 ± 1.1	9	36.6 ± 2.2	9	**39.0 ± 2.2**	9	34.1 ± 3.2	6	35.4 ± 3.4	7	**39.6 ± 1.1**	5
Eub338-III	35.3 ± 0.8	15	**37.0 ± 0.9**	14	36.0 ± 1.4	14	35.9 ± 0.7 ¥	12	**37.4 ± 2.0**	13	35.0 ± 2.4	11
Alf1b	36.5 ± 1.3	18	37.1 ± 1.8	18	36.6 ± 1.8	18	37.5 ± 1.6	16	38.0 ± 0.9	12	36.8 ± 2.3	15
Alf968	37.3 ± 1.2	12	36.6 ± 1.7	11	**35.2 ± 1.0**	12	39.6 ± 0.7	10	**36.3 ± 1.4**	9	**37.2 ± 2.2**	11
Bet42a	32.1 ± 0.4	9	31.7 ± 0.4	8	**33.6 ± 1.2**	8	31.7 ± 2.1	8	31.4 ± 1.5	8	32.8 ± 1.1	6
Nmo218	35.0 ± 1.5	6	**31.6 ± 0.7**	6	33.8 ± 1.2	6	37.8 ± 0.6	6	**32.7 ± 0.4**	6	**35.4 ± 0.9**	6
Nse1472	33.8 ± 1.0	6	**31.5 ± 0.4**	5	34.4 ± 1.1	5	35.6 ± 0.4	6	**32.4 ± 0.4**	6	35.8 ± 0.5	6
Nsv443	34.8 ± 1.4	6	**33.2 ± 1.1** ¥	6	35.0 ± 1.7	6	36.0 ± 0.9	6	**33.5 ± 0.8**	5	34.6 ± 1.5	4
Gam42a	31.8 ± 0.5	14	31.7 ± 0.3	14	**35.6 ± 3.3**	15	31.3 ± 0.8	13	31.2 ± 0.6	13	**34.3 ± 2.5**	12
Nsc825	33.8 ± 0.8	6	32.9 ± 0.1	3	34.4 ± 0.7	4	34.6 ± 0.5	6	**33.2 ± 0.6**	6	34.5 ± 0.6	6
Gpos1192	35.4 ± 1.2	12	**38.0 ± 2.8**	12	34.5 ± 7.9	12	35.2 ± 2.9	11	34.4 ± 3.6	9	36.0 ± 3.5	10
Gpos1200	38.2 ± 5.6	6	41.2 ± 6.2	6	39.3 ± 3.1	6	38.9 ± 3.9	6	35.7 ± 5.6	6	34 ± 12	6
Hgc69a	33.3 ± 1.7	12	33.4 ± 1.3	12	**37.5 ± 3.6**	11	31.3 ± 2.0	10	31.0 ± 2.4	9	34.8 ± 5.2	8
Lowgc353	35.4 ± 0.6	6	**33.8 ± 0.7**	6	35.1 ± 0.7	4	35.0 ± 2.7	5	32.0 ± 2.7	6	34.1 ± 2.7	4
Bacil1136	35.8 ± 0.9	6	36.1 ± 1.5	6	37.3 ± 1.5	6	35.5 ± 1.3	6	34.2 ± 1.5	5	36.0 ± 2.2	5
Cf319a	36.8 ± 1.2	12	**34.9 ± 1.5**	12	**35.1 ± 0.8**	12	37.6 ± 2.3 ¥	11	**34.2 ± 1.1**	6	**35.4 ± 1.5**	11
Ntspa0712	37.7 ± 0.9	6	**33.6 ± 0.5**	6	**35.4 ± 0.5**	6	39.1 ± 0.5	4	**33.2 ± 1.7**	5	**34.2 ± 1.3**	5
Neg. ctrl.	45 ± 21	18	na		na		43 ± 13	18	na		na	

* Values shown are average T_d ± standard deviation of triplicate hybridizations to gel-pad microarrays. T_d values of MM probes significantly different (Student's *t*-tests) than the ones of corresponding PM probes are indicated in bold. Shaded zones correspond to T_d values of PM probes significantly different (Student's *t*-tests) when hybridized with uncontaminated or TNT-contaminated soil. Probes Nso1225, Nso190, Nsm653 and Nsm156 were not analyzed with the functional ANOVA calculator because of the low number of observations (n<4). $^{\tau}$ Number of dissociation curves after rejection of outliers. ¥ Results (i.e. hypothesis of equality of the compared means accepted or rejected) of Wilcoxon tests differed from Student's *t*-tests. na: not applicable.

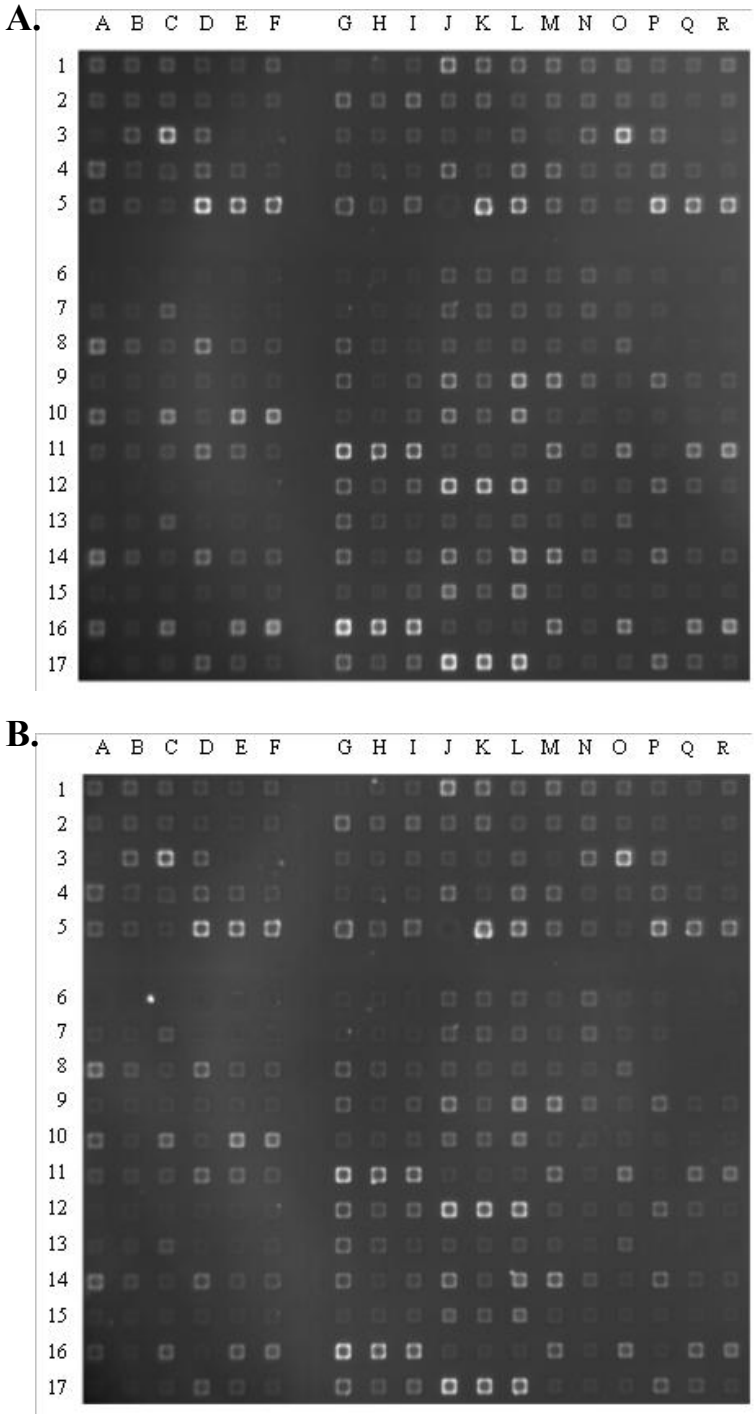


Figure S-1. Images of representative hybridizations of in vitro transcribed 16S rRNA of uncontaminated soil (A) and TNT-contaminated soil (B) to gel-pad microarrays.