Journal of Microbiological Methods 75 (2008) 92–102

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Simultaneous quantification of multiple nucleic acid targets in complex rRNA mixtures using high density microarrays and nonspecific hybridization as a source of information

Alex E. Pozhitkov [a], Georg Nies [b], Barbara Kleinhenz [b,c], Diethard Tautz [b,c], Peter A. Noble [d,*]

[a] College of Marine Sciences, P.O. Box 7000, University of Southern Mississippi, Ocean Springs, MS 39566, USA
[b] Institute for Genetics, Cologne, D-50674, Germany
[c] Max-Planck-Institut fuer Evolutionsbiologie, August-Thienemannstrasse 2, 24306 Ploen, Germany
[d] Civil and Environmental Engineering, University of Washington, Seattle, WA 98195, USA

ARTICLE INFO

Article history:
Received 21 March 2008
Received in revised form 1 May 2008
Accepted 7 May 2008
Available online 19 May 2008

Keywords:
Pattern recognition
Nonspecific hybridizations
Oligonucleotide arrays

ABSTRACT

To date, it has been problematic to accurately quantify multiple nucleic acid sequences, representing microbial targets, in multi-target mixtures using oligonucleotide microarrays, primarily due to nonspecific target binding (*i.e.*, cross-hybridization). While some studies ignore the effects of nonspecific binding, other studies have developed approaches to minimize nonspecific binding, such as physical modeling to design highly specific probes, subtracting nonspecific signal using mismatch probes, and/or removing nonspecific duplexes by scanning through a range of wash stringencies. We have developed an alternative approach that, in contrast to previous approaches, uses nonspecific target binding as a source of information. Specifically, the new approach uses hybridization patterns (fingerprints) to quantify specific nucleic acid targets in complex target mixtures. We evaluated the approach by mixing together *in vitro* transcribed 28S rRNA targets at varying concentrations (up to 1.0 nM), and hybridizing the 24 mixtures to microarrays ($n=3160$ probes, in duplicate). Three independent Latin-square-designed experiments revealed accurate quantification of the targets. The regression between actual concentration of targets and those determined by the approach were highly positively correlated with high R^2 values (*e.g.*, $R^2=0.90$, $n=6$ targets; $R^2=0.84$, $n=8$ targets; $R^2=0.82$, $n=10$ targets).

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Accurate quantification of specific nucleic acid targets is important to many areas of biomedical science because they indirectly reflect the condition of a biological system. For example, in the case of quantitative gene expression profiling, the abundance of a specific gene provides mechanistic insights into cellular function (Lipshutz et al., 1999). In the case of ribosomal RNA molecules extracted from a biomedical or environmental sample, the quantities of specific sequences provide a snapshot of microbial population dynamics (Noble et al., 1997).

High-throughput technologies, such as oligonucleotide microarrays, offer significant potential for the simultaneous quantification of multiple nucleic acid targets within the same sample. High-density microarrays typically contain hundreds of thousands of different immobilized probes, with each probe acting as an individual sensor with its own specificity and sensitivity to different nucleic acid targets in solution. The intensity of the signal from each probe provides a measure of the amount of bound nucleic acid target in a sample. It is difficult to interpret the signal from a probe in terms of target quantity, because there is a multitude of possible interactions that can occur between the nucleic acid targets in solution and oligonucleotide probes on an array surface. Additionally, each probe (*i.e.*, sensor) has its own unique binding properties (Pozhitkov et al., 2006). Specifically, the extent of nonspecific hybridization is usually not known, and therefore it is not possible to accurately determine how much of the signal can be attributed to a particular target, or how much of it is due to nonspecific binding (Zhang et al., 2005; Pozhitkov et al., 2007b).

Some studies (Palmer et al., 2006, 2007) have chosen to ignore the effects of nonspecific binding, noting that array design and analysis methods are imperfect and still evolving, while others have attempted to minimize or compensate for nonspecific binding (see Discussion for details). Given the complexity of the samples tested as well as lack of statistical models for identification and/or quantification, it is difficult to ascertain the extent of false-positive and false-negative results in these studies. They can neither be proven nor disproved.

Since nonspecific target binding appears to be inherent to all microarray experiments, we have recently begun devising an analytical approach that uses intensities of all oligonucleotide probes on a

* Corresponding author.
E-mail addresses: Alexander.Pozhitkov@usm.edu (A.E. Pozhitkov), gnies@uni-koeln.de (G. Nies), b.kleinhenz@uni-koeln.de (B. Kleinhenz), tautz@uni-koeln.de (D. Tautz), panoble@washington.edu (P.A. Noble).

0167-7012/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.mimet.2008.05.013

microarray as a source of information (Pozhitkov et al., 2007a). The approach permits direct analysis of hybridization patterns of targets in mixed target samples. The basic principle of the approach is that a fluorescence pattern of a mixed sample is a superposition of individual fluorescence patterns (*i.e.*, fingerprints) of each target. The superposition can be deconvoluted in terms of individual target mass or concentration, and the quality of the quantification can be assessed by linear regression analysis (*e.g.*, R^2).

Our previous study (Pozhitkov et al., 2007a) showed only basic proof-of-principle and examined a very limited number of rRNA from three microorganisms (n=3) at the same concentration and in one mixture (in triplicate). The present study rigorously evaluates the approach using pure *in vitro* transcribed rRNA from a clone library to accurately quantify mixed nucleic acid targets hybridized to oligonucleotide microarrays at a range of concentrations. Our specific objectives were: (i) to perform an additional analysis of the data published in our previous work (Pozhitkov et al., 2007a) in order to further verify the stability of the approach, (ii) to determine the relationship between the number of probes on a microarray and the accuracy and precision for quantifying specific targets using the new approach, (iii) to assess the ability to quantify specific targets at varying concentrations using a Latin-square experimental design, (iv) to determine if prescreening of probes, based on selectivity and sensitivity measurements, would improve target quantification, and (v) to establish the limits of our approach in quantifying specific targets in multi-target mixtures of varying concentrations containing both known and unknown targets.

We demonstrate accurate quantification of nucleic acid targets at varying concentrations in multi-target mixtures. The advantage of our quantification approach over existing approaches is that it provides a measure of the quality of the quantification, which can prevent false-positive or false-negative calls.

2. Materials and methods

2.1. Target preparation

The targets originated from 12 protozoa species coming from cultures of the Heterotroph Flagelates Cologne Collection (HFCC, Germany). The 12 protozoa are (HFCC number, species): HFCC9, *Neobodo designis*; HFCC12, *Bodo saltans*; HFCC13, *Bodo species*; HFCC20, *Dimastigella mimosa*; HFCC31, *Pteridomonas danica*; HFCC37, *Neobodo curvifilus*; HFCC44, *Monosiga species*; HFCC45, *undefined clone;* HFCC55, *Bodomorpha minima;* HFCC93, *Thaumatomonas coloniensis;* HFCC94, *Gyromitus sp.*; and HFCC97, *Bodo designis*. Ribosomal RNA templates of the protozoa species were derived from a project in which the D3–D5 region of the 28S rRNA was sequenced. The sequences were amplified using universal primers (28S 485wT7 5′-TAATACGACTCACTATAGGGG-ACCCGTCTTGAAACACGGA-3′; 28S 689 5′-ACACACTCCTTAGCGGA-3′) of which the forward primer carries a tail with a T7-RNA Polymerase initiation site at its 5′ end. PCR products obtained with these primers were directly used for *in vitro* transcription. The transcription was performed with the MEGAscript Kit (Ambion) according to the instructions of the supplier. The master-mix for all Latin-square experiments (see below) was supplemented with 1.8 mM biotin-conjugated UTP to label all transcripts. The RNA transcript was purified using Qiagen Rneasy kit and the yield was determined using a spectrophotometer (Thermo Scientific NanoDrop™, USA) and Bioanalyzer (Agilent Technologies, Inc., USA). Based on these RNA stocks, a special pipetting scheme was used to generate 8 Latin-square target mixtures (see Table 1), each with different concentrations of the HFCC targets.

2.2. Hybridization

The hybridizations were performed following the protocols routinely used for the GENIOM ONE® instrument (Febit GmbH, Heidelberg, Germany). The RNA fingerprint and mixture samples (final volume 16 μL) were heated at 95 °C for 3 min for denaturation before being placed in the arrays, which had been prehybridized for 30 min with 1% (w/v) BSA in MES-hyb at 20 °C. Hybridizations were carried out at 45 °C for 4 h without agitation (active mixing). After removing the hybridization solutions, the microarray was washed at 45 °C (0.5× SSPE), fluorescently stained with SAPE-solution (streptavidin-phycoerythrin, 5 μg/mL in 6× SSPE; 15 min), and then washed at 25 °C (6× SSPE) to remove unbound SAPE. The microarray was than scanned using the CCD-based detection system of the Geniom device (filter set Cy3). The images of the hybridized microarrays were recorded by auto-exposure, and the raw data of the signals were converted to MS Excel files.

Table 1
Expected concentrations of target by mixture and Latin-square experiment

Latin-square	HFCC	Amount of target (pM) by mixture							
Experiment	Number	1	2	3	4	5	6	7	8
1	9	64	0	1	2	4	8	16	32
	20	32	65	2	5	10	20	40	80
	31	16	32	64	0	1	2	4	8
	37	8	16	32	64	0	1	2	4
	44	6	12	24	48	95	64	1	3
	55	1	2	4	8	16	32	64	0
2	9	255	127	64	32	16	8	4	0
	12	127	64	32	16	8	4	0	255
	13	64	32	16	8	4	0	255	127
	20	32	16	8	4	0	255	127	64
	31	16	8	4	0	255	127	64	32
	37	8	4	0	255	127	64	32	16
	45	4	0	255	127	64	32	16	8
	55	0	255	127	64	32	16	8	4
3	45	1000	500	250	125	63	31	16	8
	94	500	250	125	63	31	16	8	4
	90	250	125	63	31	16	8	4	0
	13	125	63	31	16	8	4	0	1000
	12	63	31	16	8	4	0	1000	500
	31	31	16	8	4	0	1000	500	250
	93	16	8	4	0	1000	500	250	125
	55	8	4	0	1000	500	250	125	63
	9	4	0	1000	500	250	125	63	31
	97	0	1000	500	250	125	63	31	16

2.3. Oligonucleotide microarrays

A set of oligonucleotide probes for the microarrays was generated using a C^{++} program written for a previous study (Pozhitkov et al., 2005b). The set consisted of perfect match (25-mer) probes that were complementary to the 12 rRNA targets (see Target preparation section) as well as several other HFCC species. All probes were replicated two times to provide a measure of intra-array reproducibility. In total, 6320 oligonucleotide probes were synthesized on the microarray by the GENIOM One® instrument (Febit GmbH, Heidelberg, Germany).

A light-activated *in situ* oligonucleotide synthesis was performed within the GENIOM instrument on the activated three-dimensional reaction carrier, which contained a glass-silicon-glass sandwich, using a digital micromirror device (Texas Instruments).

2.4. Latin-square design

The Latin-square design was used to explore a variety of concentrations for each target in a mixture. A Latin square design is a $n \times n$ table filled with n different concentrations in such a way that each concentration occurs exactly once in each row and exactly once in each column. Three independent Latin-square experiments were conducted, using a different range of concentrations. The first Latin-square experiment used a range of target concentrations from 0.0 to 0.08 nM, while the second and third experiments used a range of concentrations from

0.0 to 0.25 nM and from 0.0 to 1.0 nM, respectively. Table 1 shows the exact amount of targets dispensed into each mixture and experiment.

2.5. Availability of data

We developed a Microsoft Access database for fingerprint and mixture signal intensities that is publicly available at http://faculty.washington.edu/pozhit/default.htm. The database also contains queries that can be used to analyze the data.

2.6. Numeric solution

The software for calculating the numeric solution was originally designed in Mathcad and later converted to C++. The C++ code allowed us to perform multiple simulation tests otherwise not possible in Mathcad. A description of the analytical approach can be found in the original publication (Pozhitkov et al., 2007a). Briefly, in this approach, the entire set of probes and their corresponding signal intensities are viewed as a whole and considered to be a "fingerprint". The fingerprint of each target that is suspected to be present in the mixture is recorded into a "library of fingerprints". A pattern of an unknown mixture is then numerically solved in terms of relative contributions of each pattern from the "library of fingerprints." These contributions can be easily converted into mass or concentration units.

2.7. Simulation studies

To evaluate the effects of unknown targets on quantification of specific targets, we conducted simulation studies. We generated "unknown targets" by randomly shuffling the probe signal intensities of the original target fingerprints. The generated fingerprints did not resemble any of the target fingerprints in the library. The reason for using the original target fingerprints was to ensure that probe signal intensities of the generated fingerprints had similar intensity distributions to those of the originals. To distinguish these fingerprints from those produced from the real targets, we re-labeled them with 'r' in front of the target it was generated from. For example, the randomly-shuffled fingerprint of HFCC9 was labeled as 'r9'. Once the 'unknown' targets were generated, the fingerprint of each target was then independently added to the hybridization patterns of the Latin-square mixtures at concentrations ranging from 0.0 nM to 0.06 nM. The concentration of added simulated fingerprint was defined as a factor that was multiplied to the vector of signal intensities of the simulated pattern, and then added to the pattern of the mixture. The factor was calculated as a product of the desired concentration (*e.g.*, 0.06 nM) and the system specific coefficient was calculated from the quantification of real mixtures and real fingerprints.

Specifically, this coefficient was calculated as a sum of all numerical solutions for a mixture divided by the total known RNA concentration. We then deconvoluted each mixture to determine the concentration of all six 'known' fingerprints. We assessed the effect of unknown targets on the analytical approach by evaluating (i) the change in the R^2s of the numeric solutions, (ii) the R^2 of the composite linear regressions of all mixtures, and (iii) the predicted quantities of the known targets in the mixtures.

2.8. Selectivity and sensitivity

Sensitivity was defined as the maximum signal intensity of a probe for all targets divided by the pixel saturation maximum (65,535 a.u.), whereas selectivity was defined as the maximum signal intensity of a probe for all targets divided by the sum of all intensities for that probe.

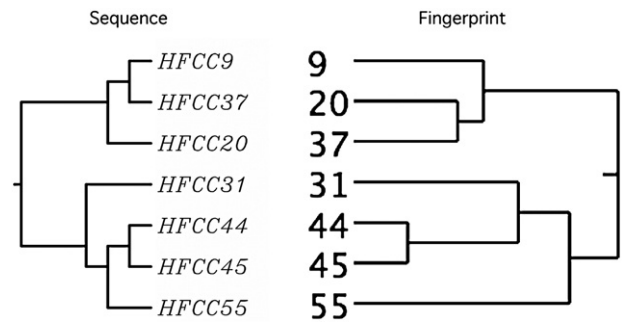


Fig. 1. Tree showing the relationships among seven 18S rRNA sequences (left panel) and the dendrogram generated by comparing fingerprints (n = 3160 probes, right panel). The DNA sequences were aligned using ClustalW. The dendrogram was constructed using SAS/JMP using hierarchical clustering (UPGMA). Fingerprints of HFCC44 and HFCC45 were found to be highly similar (right panel), which was consistent with sequence comparisons (left panel). The presumed reason that 9, 20, and 37 do not have mirror symmetry is because the fingerprints were based on a limited number of probes on the array while the sequence was based on nucleotide comparisons.

Hence, a probe with high sensitivity would yield high signal intensity for one or more targets examined, whereas a probe with low sensitivity would not yield a high intensity for any of the targets. In contrast, a probe with high selectivity would yield *relatively* high signal intensity for only one of the targets examined, whereas a probe with low selectivity would yield similar signal intensities for all targets.

The sensitivity (Sn) and selectivity (Sl) measures can be represented by the following equations:

$$\mathrm{Sl} = \frac{\max(S_i)}{\sum_i S_i},$$

$$\mathrm{Sn} = \frac{\max(S_i)}{65,535},$$

where S — signal intensity of the probe in the i-th fingerprint.

Selectivity, therefore, can be expressed via sensitivity:

$$\mathrm{Sl} = \frac{\mathrm{Sn}}{\mathrm{Sn} + \sum_j S_j/65,535}$$

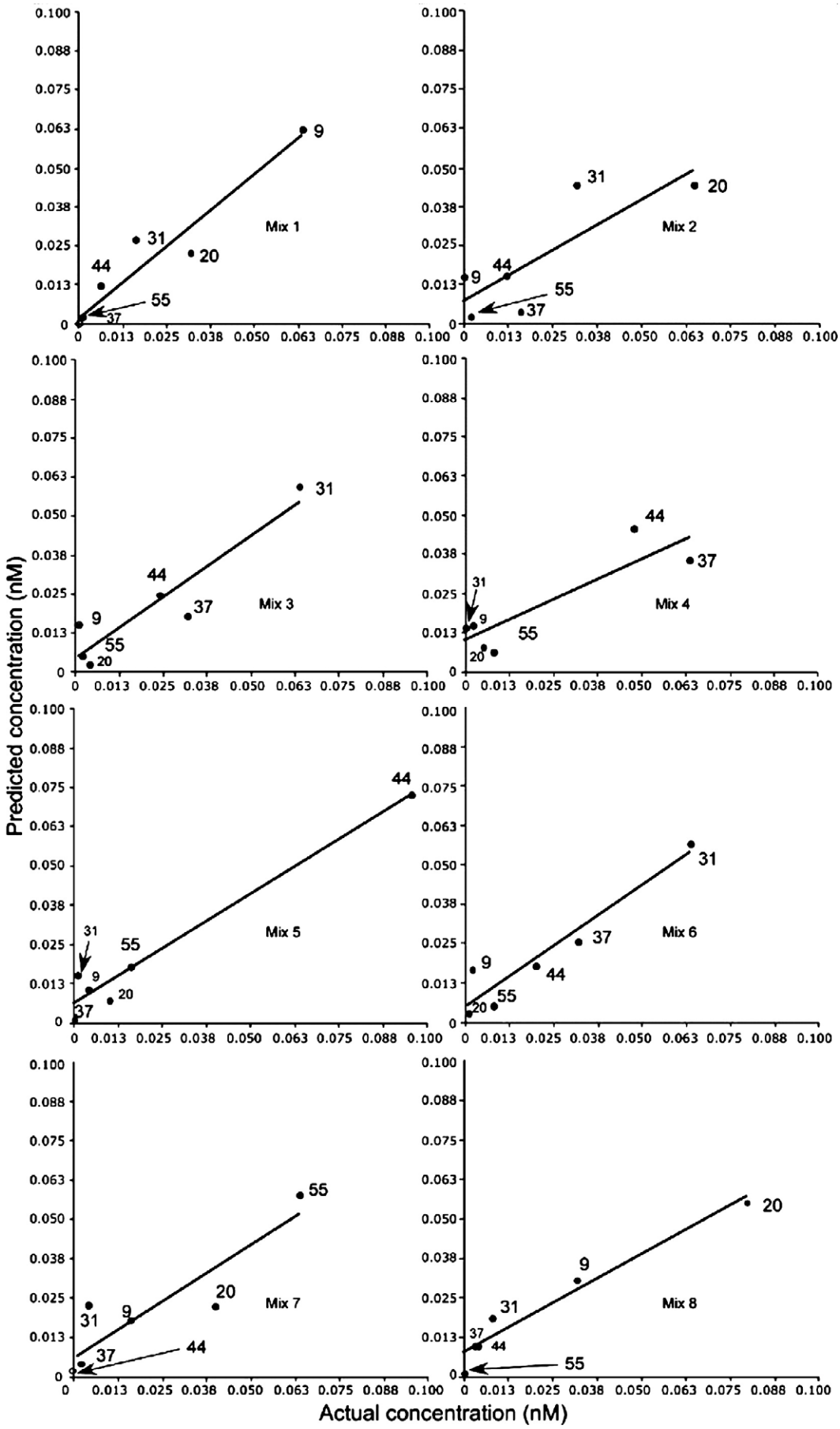
with j enumerating all fingerprints but excluding the one where maximum of signal intensity occurred. The sum of Sj in the denominator is probe specific. The minimum of this sum (divided by 65,535) among all probes in the dataset was found to be 0.03, which determines the position and shape of the envelope line (see green line in Fig. 6).

3. Results

3.1. Determining the stability of the target quantification approach

Data from our previous study (Pozhitkov et al., 2007a) were used to determine the stability of target quantification by the numeric solution. The arrays used in the previous study were different from those used in this study because they had a larger number of probes (n = 86,652 for NimbleGen arrays *versus* 3160 probes for Febit arrays). We iteratively selected different subsets of the NimbleGen data, and we also varied the number of probes used for the analysis. This was accomplished by generating ten size classes using the bisection method (Burden and Faires, 2000) (starting with 86,652 and ending with 84 probes). In each size class, we randomly extracted ten subsets from the original probe set. Each subset was then deconvoluted to determine the quantity of targets in a mixture.

Fig. 2. Calculated (from the numeric solution) and actual concentrations (nM) of targets for eight independent mixtures (Experiment #1) (n = 3160 oligonucleotide probes). Ribosomal RNA targets HFCC9, HFCC20, HFCC31, HFCC37, HFCC44, and HFCC55 are shown as numbers associated with each datum point.


Predicted concentration (nM)
Actual concentration (nM)
Mix 1
Mix 2
Mix 3
Mix 4
Mix 5
Mix 6
Mix 7
Mix 8
0.100
0.088
0.075
0.063
0.050
0.038
0.025
0.013
0
0.013 0.025 0.038 0.050 0.063 0.075 0.088 0.100

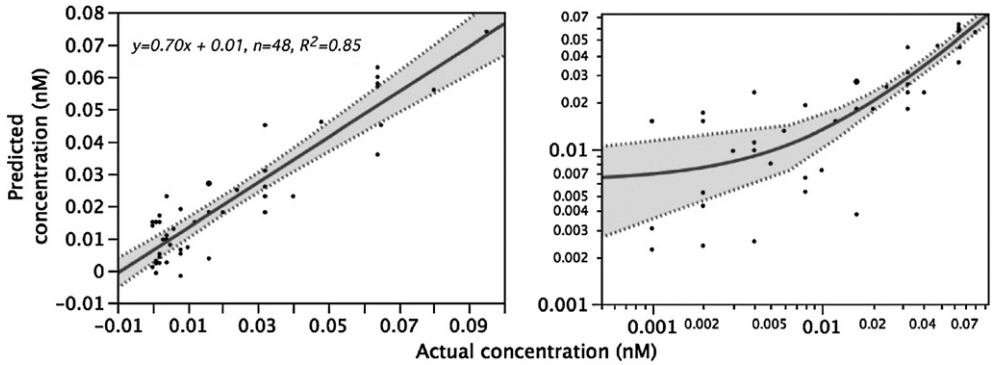


Fig. 3. Composite relationship between predicted and actual concentrations based on eight mixtures (n=3160 oligonucleotide probes) shown in Fig. 2 (Experiment #1). The grey lines and shade indicate the 99% confidence limits of the regression line. Left panel, linear plot; Right panel, log–log plot.

The predicted quantities of each target was more or less consistent when 1000 to 86,652 probes were used, indicating that the analytical approach was very stable and that far fewer probes were required than previously anticipated (see Fig. S1). Below, we will show that the accuracy of the quantifications can be further improved, and that the minimum number of probes needed for accurate quantification can be further reduced by using a probe selectivity measure.

3.2. Latin Square mixtures of targets

In the first Latin-square experiment, we attempted to deconvolute 8 different mixtures of 7 targets (Table 1). However, the numeric solution yielded negative or inconsistent quantities from expected values. The source of this problem was found to be one pair of targets that had similar fingerprints (specifically, Pearson's correlation of the signal intensities for HFCC44 and HFCC45 was 0.99, see also a dendrogram in Fig. 1, right panel). Sequence alignment of the corresponding DNAs indicated that they differed by less than 2% (see Fig. 1, left panel). Better resolution of the fingerprints of these sequences might be possible using more probes. In this study, we were limited to 1580 probes that covered 12 rRNA targets. The problem with highly correlated fingerprints is that the targets cannot be quantified by solving an over-defined system of linear equations because they cause numeric instability. Obviously, since targets having identical fingerprints act as one target, we added the concentrations of these targets in the mixtures together, treating them as one target. This procedure resulted in reducing the number of targets in the first Latin-square experiments from 7 to 6.

The relationship between the actual versus predicted concentrations of the 6 targets in 8 different Latin-square mixtures, using 3160 probes, is shown in Fig. 2. For statistical analysis of a Latin-square experiment, the plots shown in Fig. 2 were overlaid to form a composite plot, and the regression of the predicted and actual target concentrations for all mixtures was calculated (Fig. 3, left panel). The R^2 of the linear relationship of the mixtures indicated that about 85% of the variability in the Latin-square experiments was explained by the data, with the remainder attributed to undefined experimental error. The log–log plot of the same data (Fig. 3, right panel) showed that the distribution of the regression increased at lower concentrations presumably because targets were affected by noise in the signal at the detection limits of the system. Note that the removal of the highest target concentrations and reanalysis of the data would not have significantly altered the slope of the regression. These results demonstrate that accurate quantification of targets in complex target mixtures was possible using our approach.

3.3. Effects of adding additional fingerprints on quantification

We investigated the effects of adding 6 randomly-shuffled fingerprints (n=3160 probes) to the library on the quantification of actual and "simulated" targets in mixtures. The results showed that the number of fingerprints in the library had no effect on either quantification of specific targets, or the R^2s of the numeric solution and the composite linear regression (data not shown).

3.4. Effects of unknown targets on the accuracy of quantification

3.4.1. Experimental

Most real-world samples contain unknown targets whose fingerprints are not in the library. We found that removing a known fingerprint from the library of fingerprints (*e.g.*, HFCC9), although physically present in the mixture, did not have a significant effect on the quantification of other targets in the library (Fig. 4), but it did affect the R^2 of the numeric solution. For example, in the case of removing HFCC9 from the library, the R^2 of the numeric solution dropped from 0.93 to 0.75, but only for one mixture that had the HFCC9 target at a high concentration relative to other targets in the mixture (~0.063 nM; the actual concentration of HFCC9 is shown in Fig. 2, top left panel). Removing other fingerprints from the library yielded similar results (data not shown).

3.4.2. Simulation

In order to further evaluate the effects of unknown targets, we performed a simulation study by introducing randomly-shuffled unknown targets to the mixtures (see Materials and methods). In general, as the concentration of the unknown target in the mixtures increased (the treatment), R^2-values of both the numeric solution and the composite linear regression of the mixtures decreased (black circles, Fig. 5). Note that the response to the treatment differed among unknown targets, with some random fingerprints having more effect than others (*e.g.*, r55 versus r31 and r9 versus r44, Fig. 5 upper and lower panels, respectively). These differences are presumably due to the distribution of signal intensities in the random target's fingerprint. In contrast to the treatment, the control experiments included a fingerprint of the unknown target in the library (open circles). In this situation, quantifications of real and random targets resulted in high R^2. As shown in Fig. 5, the R^2 of both the numeric solution and the composite linear regression slightly increased when the unknown target increased beyond ~0.038 nM, which could be attributed to the random target becoming the dominating factor in the numeric solution. One-tailed Student t-tests of the R^2 values of the numeric solutions for the treatment and control groups revealed significant differences (P<0.001, n=48) when more than 0.006 nM of the unknown target was added to the mixtures. Yet, the R^2 of the composite linear regressions of the mixtures was found to be less sensitive to unknowns than R^2 of the numeric solutions. One-way Student t-tests of the R^2 of the composite linear regression of the mixtures revealed significant differences between treatment and control groups when the concentration of the unknowns was greater than 0.063 nM (P<0.01, n=6).

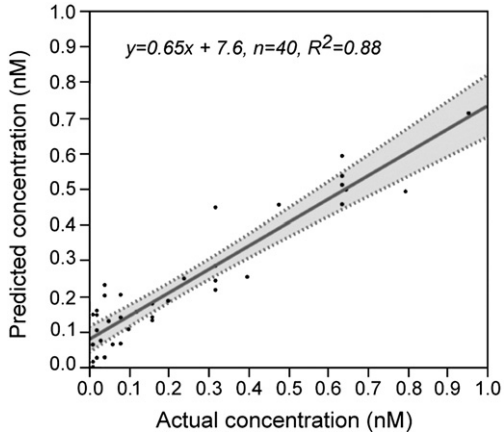


Fig. 4. Composite relationship between predicted and actual concentrations based on eight mixtures (n=3160 oligonucleotide probes) shown in Fig. 2 (Experiment #1) with HFCC9 removed. The grey lines and shade indicate the 99% confidence limits of the regression line. Left panel, linear plot; Right panel, log–log plot.

Since the total concentration of the original mixtures was 0.13 nM, this result implies that the approach might be significantly affected by the patterns of unknown targets when their combined concentration was more than 50% of all targets in the library. It is important to realize that these simulation results, though important for providing a theoretical framework to develop testable hypotheses, might not be suitable interpretations for real experiments because they assume that the signal intensity level of unknown targets is the same as that of known targets. That is, the simulation results do not consider the actual binding affinities of targets to probes. Actual experiments are needed to further verify the simulations.

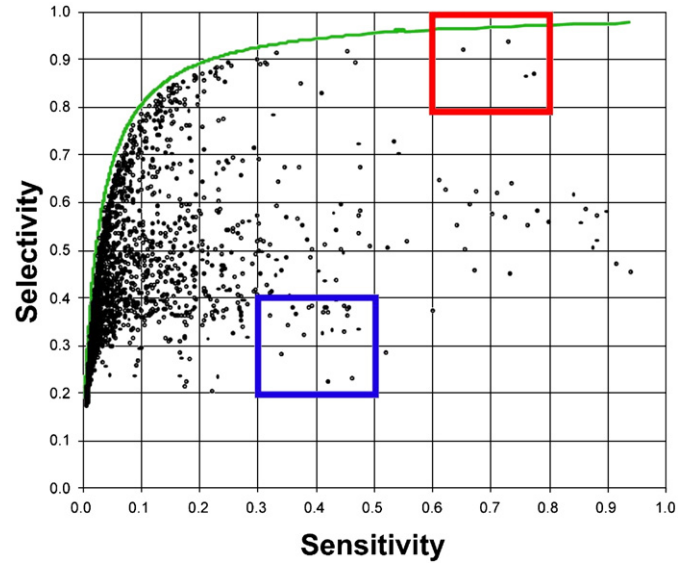


Fig. 6. Relationship between selectivity versus sensitivity for 3160 probes, first Latin Square experiment, 6 fingerprints (see text for details; Experiment #1). Probes with high selectivity and sensitivity (red) had both perfect match duplexes as well as those with multiple mismatches to the targets. Probes that hybridized to all targets (blue; universal probes) also had both perfect match duplexes as well as those with multiple mismatches to the targets. The green line depicts the limits of the relationship between selectivity and sensitivity for our data set (see Materials and methods).

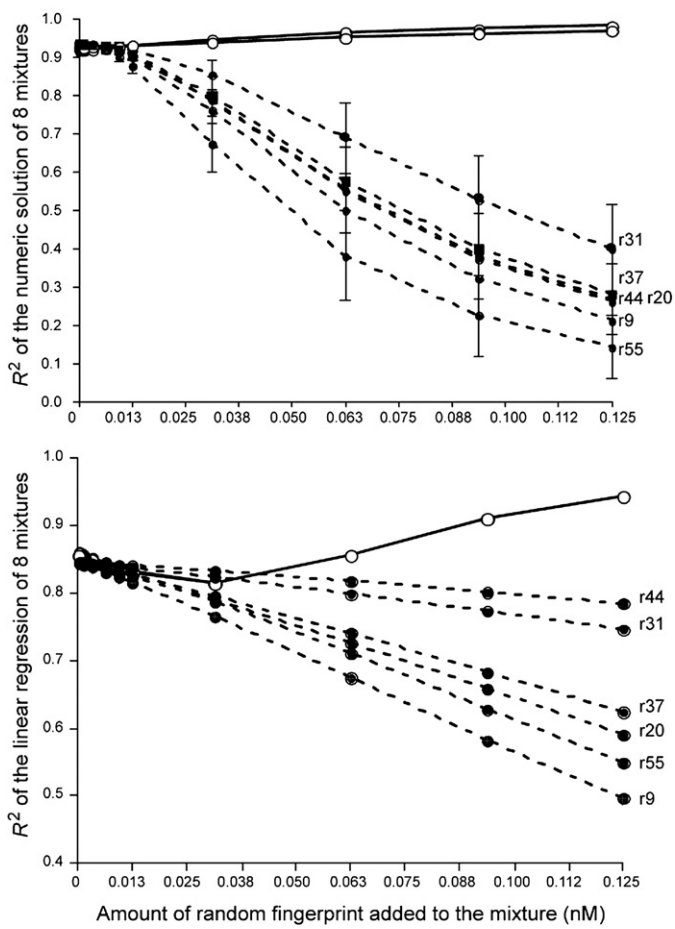


Fig. 5. Effects of increasing concentration of a random target (an unknown) added to each of the 8 Latin-square mixtures (Experiment #1, n=3160 probes). Shown are the average R2 of the numeric solution (±SD; upper panel) and R2 of the linear regression between predicted and observed concentrations of targets in the first Latin-square experiments (lower panel). Open circles, results obtained using a fingerprint library consisting of HFCC9, HFCC20, HFCC31, HFCC37, HFCC44, and HFCC55, and, close circles, randomly-shuffled fingerprints indicated by the 'r' in front of the HFCC number. Increasing the concentration of an "unknown" target (as represented by a randomly-shuffled targets) in a mixture decreased the R2 of both the numeric solution and the composite linear regression based on Latin-square experiments.

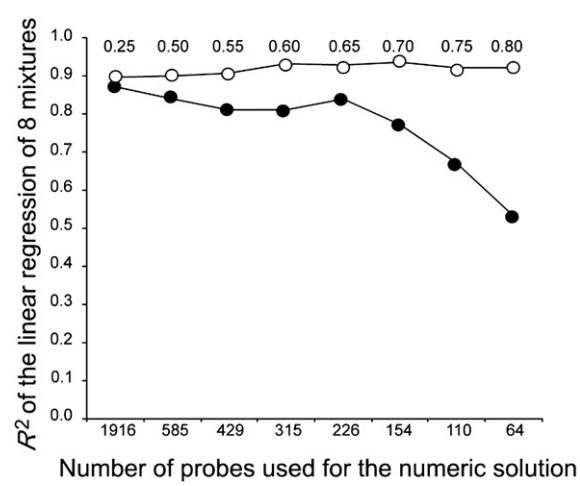


Fig. 7. Effects of selectivity on R2 of the linear regression for 8 Latin-square experiments (Experiment #1). Increasing the value of selectivity used to pick probes (from 0.25 to 0.80), had no apparent effects on the R2 of the linear regressions, even though it reduced the number of probes available for the analysis. However, in the case of experimental controls where no selectivity value was used, a reduction in the number of probes used for the numeric solution did affect the R2 of the linear regression (close circles). The R2 of the regression line of 8 different mixtures decreased as the number of probes decreased. Thus, selecting probes based on their selectivity can reduce the number of probes required to make accurate quantification of specific microbial targets in mixtures.

3.5. Probe selectivity and sensitivity

An oligonucleotide microarray is essentially a collection of probes-sensors that respond to the concentration of a target in a certain manner that can be characterized by selectivity and sensitivity measures (see Materials and methods section for the definitions). Fig. 6 shows the relationship between selectivity and sensitivity for all 3160 probes based on fingerprint data collected from six targets at 1.0 nM concentration. Sensitivity values in Fig. 6 ranged from 0.01 to 0.94 and selectivity values ranged from 0.17 to 0.94. Four of the 3160 probes had both high sensitivity and high selectivity (red box in Fig. 6). Two of these probes were perfectly matched to the target, while the remaining probes had at least seven mismatches with a 10-base complementary overlap. On the other hand, of the probes that formed duplexes with all six targets, *i.e.* universal probes (22 probes in the blue box in Fig. 6), six were perfectly matched to one or more of the six targets, while the remaining 16 probes had at least five mismatches to the targets. These findings indicate that the number of mismatches had little to do with probe selectivity as well as sensitivity.

3.6. Selectivity measure helps reduce the number of probes

In order to further reduce the number of probes needed to accurately quantify targets in mixtures, we investigated two approaches: (i) randomly-selecting probes, and (ii) picking probes based on their selectivity. As Fig. 7 shows, random picking of probes significantly decreased the R^2 of the linear regression of the composite Latin-square mixtures (black circles, Fig. 7). At the same time, selectivity-based picking of probes did not change the R^2 at all (white circles, Fig. 7) — despite a 30-fold reduction in the number of probes. These results indicate that selectivity is essential for reducing the number of probes required to make accurate quantifications.

Re-evaluation of the data used to make Fig. 3 with a probe selectivity cutoff of 0.75, reduced the number of probes from 3160 to 110. However, this reduction of probes had no effect on the R^2s of the regression lines for the numeric solutions because it was high (0.90 ± 0.08, average ± standard deviation) indicating great resolution. Furthermore, the R^2 of the composite linear regression increased by 5% (R^2=0.90, Fig. 8, top panel) the goodness of fit over that obtained in Fig. 3.

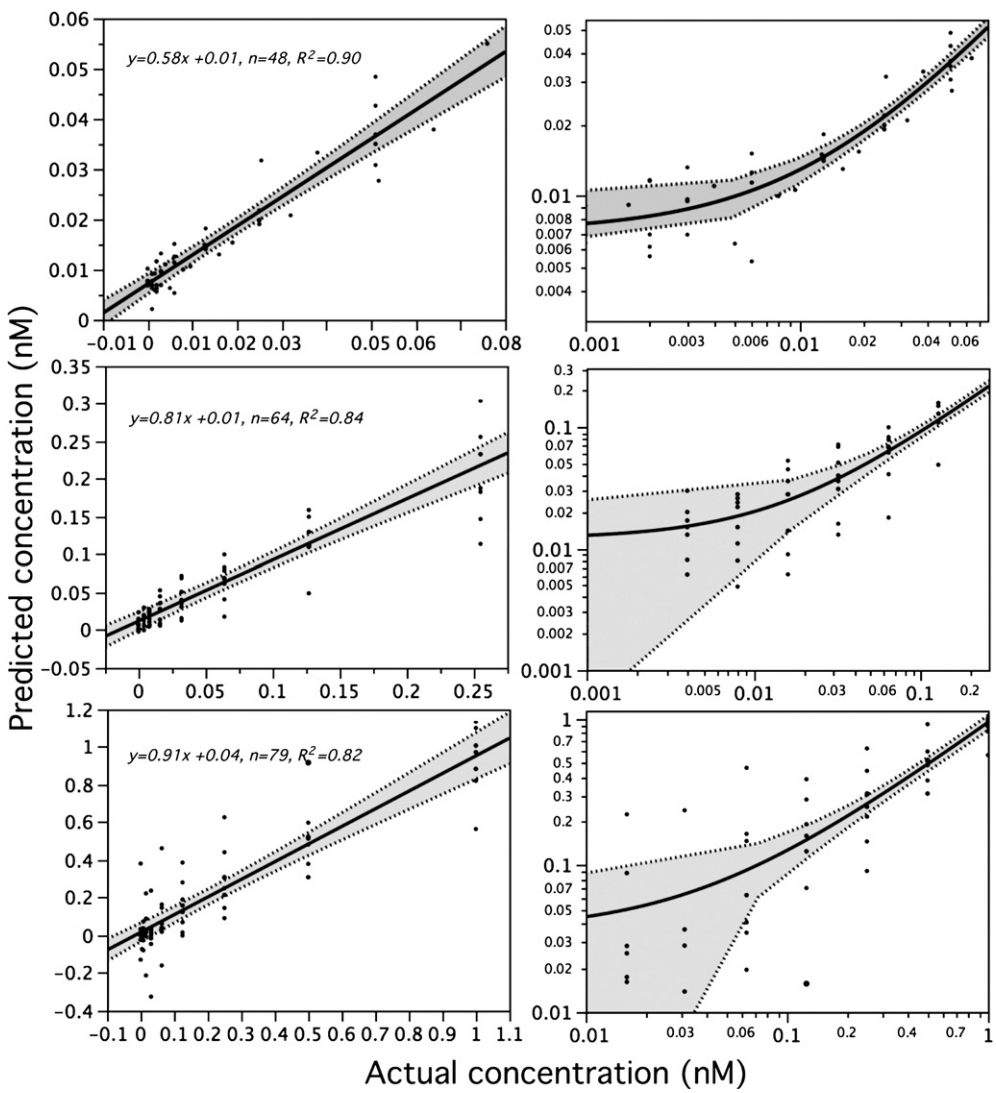


Fig. 8. Composite relationship between actual and predicted concentrations for three independent Latin-square experiments. Left panels, linear plots; right panels, corresponding log–log plots. Top panel shows results from Experiment #1, which was based on 6 targets × 8 mixtures × 110 probes; middle panel shows results from Experiment #2, which was based on 8 targets × 8 mixtures × 440 probes; and, lower panel shows results from Experiment #3, which was based on 10 targets × 8 mixtures × 400 probes. The grey lines represent the 99% confidence limits of the regression line. Left panels, linear plots, right panels, log–log plots.

3.7. Selectivity aids in distinguishing targets

We reasoned that the above-mentioned selectivity measure should not only reduce the number of probes needed for accurate quantification, it should also contrast differences among fingerprints in the library. So, for the second Latin-square experiment, we calculated probe selectivity, sorted the probes, and excluded all but the top 110 most selective probes (selectivity value of ~0.54). These probes were then used to quantify all 8 targets in all 8 Latin square mixtures (Table 1; Experiment #2). Fig. 8, middle panel shows that the composite linear regression for the 8 mixtures was moderately high: R^2=0.84 for 110 probes.

A third Latin square experiment was conducted using 8 different mixtures and 10 targets. We calculated probe selectivity, sorted the probes, and excluded all but the top 400 most selective probes (selectivity value of ~0.34). These probes were then used to quantify 10 different targets in all 8 Latin square mixtures (Table 1; Experiment #3). Fig. 8, lower panel, shows that the composite linear regression for the 8 mixtures was moderately high: R^2=0.82 for 400 probes.

In summary, these three independent Latin-squared experiments indicate that it is possible to quantify specific nucleic acid targets in target mixtures based on our analytical approach and that a linear relationship exists between actual and predicted concentration of targets.

4. Discussion

The application of microarray technology for accurate quantification of specific nucleic acid targets in complex target mixtures has been a major challenge to researchers for over a decade. Accurate and reliable quantification of specific sequences is highly desired by microbiologists, for example, because it can be used to determine how different factors (*e.g.*, biotic and abiotic) affect the abundance of multiple microbial groups and species in natural samples, and how they affect the expression of certain genes. Microarrays offer a tremendous potential over other approaches because a multitude of different targets can be simultaneously quantified in a single experiment. The main obstacle preventing accurate and reliable quantification in natural samples is nonspecific target binding because it is not possible to determine what portion of the observed signal can be attributed to specific targets versus that portion attributed to nonspecific targets.

This study builds on our previous study (Pozhitkov et al., 2007a) because it demonstrates that our approach is able to accurately quantify several targets at varying concentrations (up to 1.0 nM). The three Latin-square experiments were based on 8 different mixtures of 6, 8, and 10 targets (24 different mixtures), whereas the previous study (Pozhitkov et al., 2007a) was based on 1 mixture (in triplicate) of 3 targets at similar concentrations. These results suggest that our analytical approach is independent of the form of RNA used as target, since this study was based on *in vitro* transcribed rRNA, whereas the previous study was based on total RNA directly extracted from microorganisms. The analytical approach also seems to be independent of the array platform because Febit (Geniom) microarrays were used in this study, whereas our previous study used NimbleGen arrays. The reason to emphasize this point is because the mathematics of the analysis relies on the assumption that a hybridization pattern of the mixture is a superposition of fingerprints, which may not be necessarily true for all microarray platforms. The two investigated platforms apparently satisfied this assumption. These findings suggest that the analytical approach demonstrated in this study might have widespread applicability to a variety of microarray platforms and nucleic acid targets.

4.1. Historical attempts to deal with nonspecific binding

We preface this section by emphasizing to the reader the difference between 'selectivity', introduced in this study, and 'specificity'. Selectivity refers the uniqueness of a probe to a target in terms of relative signal to that of other targets, whereas specificity refers to the extent a probe binds to its *complementary* target to the exclusion of all other targets.

A number of different approaches have been implemented to remove or minimize the effects of nonspecific binding (*i.e.*, cross-hybridization). The three most common approaches are: (i) attempting to design probes that have high specificity and sensitivity to the target of interest (Mei et al., 2003; Pozhitkov et al., 2005b; Rouillard et al., 2002; He et al., 2005; Li et al., 2005; Binder and Preibisch, 2005; Matveeva et al., 2003), (ii) subtracting nonspecific signal from specific signal by using probes containing one- or two-internal mismatches (MM) (Lipshutz et al.,1999), and (iii) scanning through a broad range of stringency conditions (*i.e.*, nonequilibrium thermal dissociation) so that nonspecific targets are removed (Pozhitkov et al., 2005a; Urakawa et al., 2002, 2003).

Probe design approaches, such as picking probes based on Gibbs free energy terms (Matveeva et al., 2003), have been shown to result in very poor (or no) correlation between expected and actual signal intensities (Pozhitkov et al., 2006). In fact, analysis of the current literature reveals that design of specific probes is a very challenging task. For example, a review by Halperin et al. (2006) on existing physical modeling studies indicates that the nature of surface-tethered hybridizations is an extremely complex matter, where several factors have to be taken into account. Specifically, these factors include: competitive hybridization of multiple targets to a probe, interaction of bound targets on a microarray spot, interaction of probes within a microarray spot, dissociation of the targets during washing, and secondary structures of both probes and targets. Some success has been made in understanding these factors. For example, Stedtfeld et al. (2007) examined the effects of dangling ends on probe-gene target formation to find that a 7-base-long dangling end sequence with perfect homology to nontarget sequence could generate nonspecific signal that had similar intensity to that of a perfect-match probe target duplex. This finding demonstrates the complexity of probe-target interactions taking place on an array surface. Another strategy is to use only those probes having Gibbs free energy values that are less than a certain threshold (Miller et al., 2008). It is possible that future research examining the factors affecting surface tethered hybridization will lead to better strategies to design highly specific probes that are less affected by nonspecific target binding.

Subtracting the intensity of mismatch probes from the intensity of perfect match probes prior to downstream analysis has also been shown to be not suitable for dealing with nonspecific target binding because mismatch intensities do not adequately account for nonspecific signal (Wu et al., 2005). One reason mismatch duplexes were found to be inadequate is that their intensities depend on the nucleotide type as well as the composition of neighboring bases flanking the mismatch (Pozhitkov et al., 2006; Urakawa et al., 2002). Another reason is that the Wu et al. (2005) study was conducted using 25-nt array probes, which are not optimal for perfect match–mismatch discrimination. Suzuki et al. (2007) recently showed that probe lengths of 19- to 21-nt are optimal for perfect match–mismatch discrimination and that this probe length provides increased probe specificity over other lengths. Clearly, research focused on understanding the physicochemistry of short MM duplexes in complex mixtures might provide a suitable solution for dealing with nonspecific target binding.

Stringent washing approaches, which were once believed to remove nonspecific targets, have also been shown to be inadequate (Pozhitkov et al., 2007c, 2008). For example, a recent physicochemical study (Pozhitkov et al., 2007c) showed that in 20% of the cases examined, washing removed specific targets before nonspecific ones, which goes against the main assumption of the approach. Hence, thus far, not one of the above approaches has been shown to be effective at avoiding or accounting for nonspecific binding, particularly in mixed target samples.

4.2. Our new analytical approach

In contrast to other studies (*e.g.*, DeSantis et al., 2005; Wilson et al., 2002a,b; Palmer et al., 2006, 2007), our analytical approach does not attempt to minimize, reduce, or ignore nonspecific target binding, rather all probes (perfect match and mismatch) are used as a source of information for quantifying rRNA targets. Our decision to deviate from the path of previous studies was based on the realization that nonspecific binding appears to be inherent to all microarray studies, and there is no effective method to avoid or account for this phenomenon. To our surprise, the results demonstrate that the inclusion of all probes (perfectly matching and mismatching to the targets) resulted in a good correlation of predicted and actual target quantities, as demonstrated by the high R^2 of the composite linear regression of the mixtures (Fig. 3). The importance of this finding is that the inclusion of both perfect match and mismatch probes simplifies microarray experiments because one does not need to know the exact sequence of the target (*i.e.*, occasional Ns in the sequences are permitted) to determine actual target quantities.

It should be mentioned that alternative approaches for quantifying targets in mixtures using oligonucleotide microarrays have been proposed. The Marcelino et al. (2006) approach estimates nonspecific target binding by "training" experiments and then determining the quantities of specific targets using a complex algorithm (2006). This approach might indeed provide a more suitable way to minimize the effects of nonspecific binding, but the utility of this approach with high-density microarrays and short oligonucleotides (*e.g.*, 20-mers, which have a high specificity, Suzuki et al., 2007) has not been demonstrated. A potential drawback of the Marcelino et al. (2006) approach is that it requires extensive training of the system with targets that might be present in a sample in order to assess the levels of cross-hybridization of each probe, which is not possible for most natural samples. Our approach might also be affected by natural samples, particularly those having a very high diversity of targets (*i.e.*, tens of thousands of targets), because it relies on the establishment of a library of fingerprints that represent those targets. Nevertheless, we envision an approach that accounts for unknown targets, which is discussed below.

It should be noted that the Marcelino et al. (2006) approach does not provide statistical confidence of the quality of target quantifications. In contrast, in our approach, the R^2 of the numeric solution is a reliable measure of the quality of the quantification as shown in the experiment that assessed the effects of unknown targets on quantification (*i.e.*, Fig. 5).

Palmer et al. (2006, 2007) proposed a quantification technique where a sample under investigation was labeled with one dye and co-hybridized with a sample of known composition labeled with another dye. Ratios of signals coming from these dyes on the same probe reflect the relative amount of a target in the sample under investigation. Because their studies ignore the effects of nonspecific target binding, it is difficult to ascertain the extent of false-positive or false-negative results, even in light of the comparisons with clone sequence frequencies. There is no objective control to ascertain if the signals are coming from specific or nonspecific targets. As their studies employed a strict probe(s)–target relationship, nonspecific targets hybridizing to these probes could result in an inaccurate quantification of the target microbe. We have previously shown that nonspecific target binding is extensive in high-density arrays in mixed target samples (see Fig. 3 in Pozhitkov et al., 2007a). As acknowledged by the authors (Palmer et al., 2007), some microbial species predicted to be present in microarray results were not corroborated by sequence analysis of the samples.

In our approach, such a relationship between probes and a target does not exist. In principle, nonspecific targets hybridizing to several (out of many) probes should minimally affect the determined concentrations of the targets because we are dealing with all (selective) probes on the arrays — not a specific probe–target set. Furthermore, we would know if unknown targets had cross-hybridized to too many probes because it would not be resolved by the numeric solution (*i.e.*, its R^2 will be <0.75). Two additional problems of the Palmer et al. (2006, 2007) approach are: (i) it relies on the design of specific probes, which have been shown to be at best only partially predictive (Pozhitkov et al., 2006), and (ii) it depends on PCR amplification of the targets which introduces biases in the quantification.

4.3. Effects of unknown targets on quantification

The key to successful quantification of specific targets in mixtures is having recorded fingerprints in the library before deconvolution of a mixture into its component parts. However, the presence of unknown targets in the mixture, which bind well to microarray probes (see Results), does affect quantification of known targets, when they exceed more than 50% of the known targets in a mixture. We speculate that unknown targets in a biomedical or environmental sample will not strongly hybridize to the probes, which were derived from the desired targets. Further, if these unknown targets do hybridize, the signal from the targets will add subtle noise to the system, which will be reflected in R^2 of the numeric solution. If the R^2 of the numeric solution is not sufficiently high (i.e., <0.75), one should be cautious about the accuracy of all quantifications in the mixtures.

A possible approach to deal with unknown targets in a sample is to remove the known targets from a subsample and then to compare hybridization patterns of the subsample to those obtained from another subsample containing both the known and unknown targets. That is, one subsample can be subjected to an *in vitro* knockout (Mei et al., 2003) of the targets being quantified and then hybridized to the microarray, while the other subsample will undergo standard hybridization to the microarray. The pattern of the unknowns (no known targets) must then be added to the library of fingerprints in order to quantify the known targets in the sample. By doing so, quantities of known targets could then be accurately determined. Ongoing studies are testing this approach.

4.4. New probe selection approach and stability of the analytical approach

Prior to this study, we thought that many probes were needed to ensure the stability of the analytical approach in order to yield accurate quantification of targets in mixtures. However, reanalysis of the data from our previous study revealed that the analytical approach was very robust even for different random picks of the data. Our finding that accurate quantification was possible using about 1000 probes opened up the possibility of using other microarray platforms, which lead us to Febit arrays used in this study. We were surprised to discover that the number of probes needed for accurate quantification could be further reduced using a selectivity measure. The new approach offers an objective way for deciding which probes should be retained on the microarray for accurate quantification. We emphasize that this new approach is very different from the concept of probe design, optimization, and validation that is commonly suggested in microbiology (*e.g.*, see review of Loy and Bodrossy, 2006). Research aimed at determining the optimal selectivity for a given data is currently in progress.

4.5. Deviations from actual concentrations

One can argue that Figs. 3 and 8 reveal rather large deviations for the predicted concentration of some targets from their actual concentration. It is certainly true, but so far, no array approach has provided better quantifications. Similar to mass spectrometry (MS) analysis which employed the same fingerprint approach at the beginning, the MS method used to be semi-quantitative (Gillette, 1959). Progress in the field of MS drastically improved the methodology, retaining the

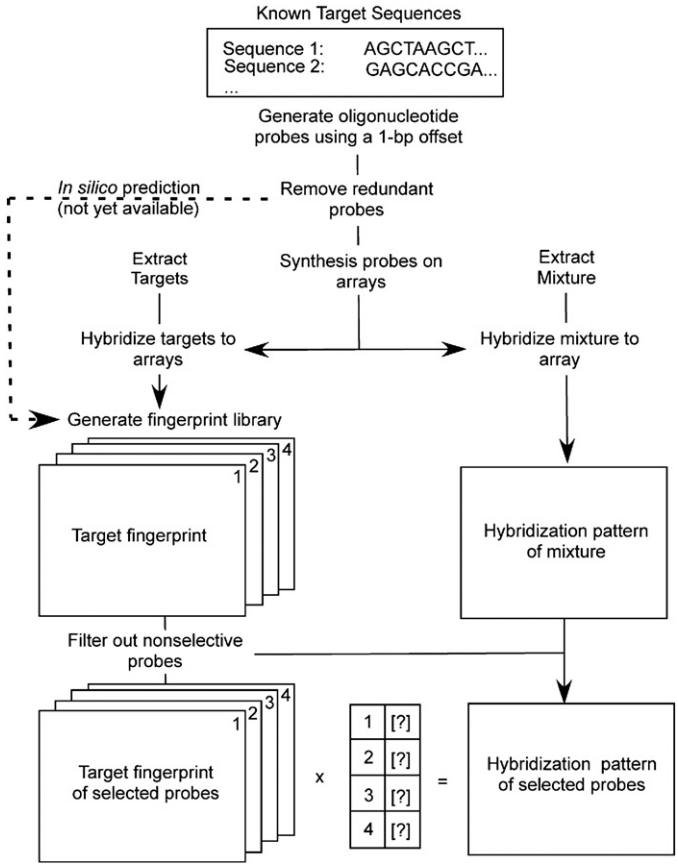


Fig. 9. Cartoon showing putative standard operating procedures for our approach using 4 hypothetical targets and one mixture. [?], determined concentration.

fingerprint-based idea as its core. We believe that the approach described in this paper is equally promising.

A second obstacle in our analytical approach is determining the optional exposure settings for the CCD camera on the Febit platform, which was found to vary from 1560 to 2970 ms for different experiments. The CDD camera, as well as PMT devices, has a nonlinear response to signal from the arrays (Shi et al., 2005; Noble and Pozhitkov, unpublished results). The way to circumvent this problem is to explicitly calibrate the detector, which is part of our ongoing research aimed at minimizing variability in array output. It is conceivable that a properly calibrated system containing thousands of probes would alleviate the variability problems encountered in this study as well as others. Exposure differences explain the fact that the slope of the regression line between predicted and actual target concentration is not 1.

4.6. Future directions

Similar to quantification of rRNA targets, mRNA (gene expression) profiles may benefit from the approach being proposed as well. For example, a library of fingerprints for a certain set of genes could be recorded, and the mRNA sample might be accurately quantified for the presence of these genes. Since most genes of one microbe are more diverse in terms of sequence than rRNA genes of closely related microbial species, one would expect much less influence of the rest of the transcriptome that are not included into the library of fingerprints.

Based on the results shown in this study, we propose the following standard operating procedures (Fig. 9): (i) design (*in silico*) a tiling array from sequences of targets under investigation (20 or 25-mers with 1 base-pair offset), remove redundant probes, and then *in situ* synthesize the probes on the array; (ii) generate a fingerprint library by experimentally recording the fingerprints of targets or using *in silico* programs that accurately predict fingerprints from sequence (not yet available); (iii) filter out nonselective probes, and (iv) quantify the sample.

5. Conclusion

In summary, accurate quantification of nucleic acid targets by oligonucleotide microarrays has been severely hampered by nonspecific target binding which affects interpretation of microarray results. Since no effective way has been shown to minimize or reduce nonspecific hybridization, we used this phenomenon as a source of information. Hybridization patterns (fingerprints) of targets can be accurately quantified in complex target mixtures at varying concentrations as demonstrated by the three Latin-square designed experiments. The approach offers tremendous potential for the multiplexed quantification of nucleic acid targets in biomedical and environmental studies.

Acknowledgements

We thank Kyle Bailey and Rebecca Rule for their critical comments on earlier versions of the manuscript. This research was supported in part by a grant from the US National Oceanic and Atmospheric Administration (NAO3NOS4260216), and by the Royalty Research Fund, and the Provost Bridge Funding Program to P.A.N.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.mimet.2008.05.013.

References

Binder, H., Preibisch, S., 2005. Specific and nonspecific hybridization of oligonucleotide probes on microarrays. Biophys. J. 89, 337–352.

Burden, R.L., Faires, J.D., 2000. Numerical Analysis, 7th ed. Cole, Brooks. 0-534-38216-9.

DeSantis, T.Z., Stone, C.E., Murray, S.R., Moberg, J.P., Andersen, G.L., 2005. Rapid quantification and taxonomic classification of environmental DNA from both prokaryotic and eukaryotic origins using a microarray. FEMS Microbiol. Lett. 245, 271–278.

Gillette, J.M., 1959. Mass spectrometer data reduction program for IBM 650. Anal. Chem. 31, 1519.

Halperin, A., Buhot, A., Zhulina, E.B., 2006. On the hybridization isotherms of DNA microarrays: the Langmuir model and its extensions. J. Phys., Condens. Matter 18, S463–S490 Sp. Iss.

He, Z., Wu, L., Li, X., Fields, M.W., Zhou, J., 2005. Empirical establishment of oligonucleotide probe design criteria. Appl. Environ. Microbiol. 71, 3753–3760.

Li, X., He, Z., Zhou, J., 2005. Selection of optimal oligonucleotide probes for microarrays using multiple criteria, global alignment and parameter estimation. Nucleic Acids Res. 33, 6114–6123.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., Lockhart, D.J., 1999. High-density synthetic oligonucleotide arrays. Nat. Genet. 21, 20–24.

Loy, A., Bodrossy, L., 2006. Highly parallel microbial diagnostics using oligonucleotide microarrays. Clin. Chim. Acta 363, 106–119.

Marcelino, L.A., Backman, V., Donaldson, A., Steadman, C., Thompson, J.R., Preheim, S.P., Lien, C., Lim, E., Veneziano, D., Polz, M.F., 2006. Accurately quantifying low-abundant targets amid similar sequences by revealing hidden correlations in oligonucleotide microarray data. Proc. Nat. Acad. Sci. 103, 13629–13634.

Matveeva, O.V., Shabalina, S.A., Nemtsov, V.A., Tsodikov, A.D., Gesteland, R.F., Atkins, J.F., 2003. Thermodynamic calculations and statistical correlations for oligo-probes design. Nucleic Acids Res. 31, 4211–4217.

Mei, R., Hubbell, E., Bekiranov, S., Mittmann, M., Christians, F.C., Shen, M.M., Lu, G., Fang, J., Liu, W.M., Ryder, T., Kaplan, P., Kulp, D., Webster, T.A., 2003. Probe selection for high-density oligonucleotide arrays. Proc. Natl. Acad. Sci. USA 100, 11237–11242.

Miller, S.M., Tourlousse, D.M., Stedtfeld, R.D., Baushke, S.W., Herzog, A.B., Wick, L.M., Rouillard, J.M., Gulari, E., Tiedje, J.M., Hashsham, S.A., 2008. In situ-synthesized virulence and marker gene biochip for detection of bacterial pathogens in water. Appl. Environ. Microbiol. 74, 2200–2209.

Noble, P.A., Bidle, K.D., Fletcher, M., 1997. Natural microbial community compositions compared by a back-propagating neural network and cluster analysis of 5S rRNA. Appl. Environ. Microbiol. 63, 1762–1770.

Palmer, C., Bik, E.M., Eisen, M.B., Eckburg, P.B., Sana, T.R., Wolber, P.K., Relman, D.A., Brown, P.O., 2006. Rapid quantitative profiling of complex microbial populations. Nucleic Acids Res. 34, e5.

Palmer, C., Bik, E.M., Digiulio, D.B., Relman, D.A., Brown, P.O., 2007. Development of the human infant intestinal microbiota. PLoS Biol. 5, e177.

Pozhitkov, A., Chernov, B., Yershov, G., Noble, P.A., 2005a. Evaluation of gel-pad oligonucleotide microarray technology using artificial neural networks. Appl. Environ. Microbiol. 71, 8663–8676.

Pozhitkov, A., Stemshorn, K., Tautz, D., 2005b. An algorithm for the determination and quantification of components of nucleic acid mixtures based on single sequencing reactions. BMC Bioinformatics 6, 281.

Pozhitkov, A., Noble, P.A., Domazet-Loso, T., Staehler, P., Beier, M., Tautz, D., 2006. Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. Nucleic Acids Res. 34, e66.

Pozhitkov, A., Bailey, K.D., Noble, P.A., 2007a. The development of a statistically robust quantification method for microorganisms in mixtures using oligonucleotide microarrays. J. Microbiol. Methods 70, 292–300.

Pozhitkov, A., Tautz, D., Noble, P.A., 2007b. Oligonucleotide arrays: widely applied – poorly understood. Brief. Funct. Genomic Proteomic 6, 141–148.

Pozhitkov, A.E., Stedtfeld, R.D., Hashsham, S.A., Noble, P.A., 2007c. Revision of the nonequilibrium dissociation and stringent washing approaches for microbial identification studies using oligonucleotide DNA arrays. Nucleic Acids Res. 35, e70.

Pozhitkov, A.E., Rule, R.A., Stedtfeld, R.D., Hashsham, S.A., Noble, P.A., 2008. Concentration-dependency of nonequilibrium thermal dissociation curves in complex target samples. J. Microbiol. Methods. doi:10.1016/j.mimet.2008.03.010.

Rouillard, J.-M., Herbert, C.J., Zuker, M., 2002. Oligoarray: genome-scale oligonucleotide design for microarrays. Bioinformatics 18, 486–487.

Shi, L., Tong, W., Su, Z., Han, T., Han, J., et al., 2005. Microarray scanner calibration curves: characteristics and implications. BMC Bioinformatics 6, S11.

Stedtfeld, R.D., Wick, L.M., Baushke, S.W., Tourlousse, D.M., Herzog, A.B., Xia, Y., Rouillard, J.M., Klappenbach, J.A., Cole, J.R., Gulari, E., Tiedje, J.M., Hashsham, S.A., 2007. Influence of dangling eds and surface-proximal tails of targets on probe-target duplex formation in 16S rRNA gene-based diagnostic arrays. Appl. Environ. Microbiol. 73, 380–389.

Suzuki, S., Ono, N., Furusawa, C., Kaskiwagi, A., Yomo, T., 2007. Experimental optimization of probe length to increase the sequence specificity of high-density oligonucleotide microarrays. BMC Genomics 8, 373.

Urakawa, H., Noble, P.A., ElFantroussi, S., Kelly, J.J., Stahl, D.A., 2002. Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses. Appl. Environ. Microbiol. 68, 235–244.

Urakawa, H., El Fantroussi, S., Noble, P.A., Kelly, J.J., Stahl, D.A., 2003. Single-base-pair mismatch discrimination using oligonucleotide DNA microarrays and melting profiles. Appl. Environ. Microbiol. 69, 2848–2856.

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

Wilson, W.J., Strout, C.L., DeSantis, T.Z., Stilwell, J.L., Carrano, A.V., Andersen, G.L., 2002b. Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. Mol. Cell. Probes 16, 119–127.

Wu, C., Carta, R., Zhang, L., 2005. Sequence dependence of cross-hybridization on short oligo microarrays. Nucleic Acids Res 33, e84.

Zhang, Y., Hammer, D.A., Graves, D.J., 2005. Competitive hybridization kinetics reveals unexpected behavior patterns. Biophys. J. 89, 2950–2959.