

Archived version from NCDOCKS Institutional Repository <http://libres.uncg.edu/ir/asu/>



## Limitations And Benefits Of ARISA Intra-Genomic Diversity Fingerprinting

By: Radu Popa, Rodica Popa, Matthew J. Marshall, Hien Nguyen, Bradley M. Tebo, & **Suzanna Brauer**

### Abstract

Monitoring diversity changes and contamination in mixed cultures and simple microcosms is challenged by fast community structure dynamics, and the need for means allowing fast, cost-efficient and accurate identification of microorganisms at high phylogenetic resolution. The method we explored is a variant of Automated rRNA Intergenic Spacer Analysis based on Intra-Genomic Diversity Fingerprinting (ARISA-IGDF), and identifies phylotypes with multiple 16S–23S rRNA gene Intergenic Transcribed Spacers. We verified the effect of PCR conditions (annealing temperature, duration of final extension, number of cycles, group-specific primers and formamide) on ARISA-IGD fingerprints of 44 strains of *Shewanella*. We present a digitization algorithm and data analysis procedures needed to determine confidence in strain identification. Though using stringent PCR conditions and group-specific primers allow reasonably accurate identification of strains with three ARISA-IGD amplicons within the 82–1000 bp size range, ARISA-IGDF is best for phylotypes with  $\geq 4$  unambiguously different amplicons. This method allows monitoring the occurrence of culturable microbes and can be implemented in applications requiring high phylogenetic resolution, reproducibility, low cost and high throughput such as identifying contamination and monitoring the evolution of diversity in mixed cultures and low diversity microcosms and periodic screening of small microbial culture libraries.

Popa, Radu & Rodica, Marshall, M., Nguyen, H., Tebo, M., & **Brauer, S.** (2009). Limitations and Benefits of ARISA Intra-genomic Diversity Fingerprinting. *Journal of Microbiological Methods*. Volume 78, Issue 2, 2009. Pages 111–118. <https://doi.org/10.1016/j.mimet.2009.06.005>. Publisher version of record available at: <http://www.sciencedirect.com/science/article/pii/S0167701209001821>

# Limitations and benefits of ARISA intra-genomic diversity fingerprinting

Radu Popa <sup>a,\*</sup>, Rodica Popa <sup>a</sup>, Matthew J. Mashall <sup>b</sup>, Hien Nguyen <sup>a</sup>, Bradley M. Tebo <sup>c</sup>, Suzanna Brauer <sup>c,1</sup>

<sup>a</sup> Department of Biology, Portland State University, 1719 SW 10th Ave., SB2 Room 246, Portland, OR, 97201, USA

<sup>b</sup> Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, 99352, USA

<sup>c</sup> Department of Environmental and Biomolecular Sciences, Oregon Health and Science University, 20000 NW Walker Rd., Beaverton, OR, 97006, USA

## A B S T R A C T

Monitoring diversity changes and contamination in mixed cultures and simple microcosms is challenged by fast community structure dynamics, and the need for means allowing fast, cost-efficient and accurate identification of microorganisms at high phylogenetic resolution. The method we explored is a variant of Automated rRNA Intergenic Spacer Analysis based on Intra-Genomic Diversity Fingerprinting (ARISA-IGDF), and identifies phylotypes with multiple 16S–23S rRNA gene Intergenic Transcribed Spacers. We verified the effect of PCR conditions (annealing temperature, duration of final extension, number of cycles, group-specific primers and formamide) on ARISA-IGD fingerprints of 44 strains of *Shewanella*. We present a digitization algorithm and data analysis procedures needed to determine confidence in strain identification. Though using stringent PCR conditions and group-specific primers allow reasonably accurate identification of strains with three ARISA-IGD amplicons within the 82–1000 bp size range, ARISA-IGDF is best for phylotypes with  $\geq 4$  unambiguously different amplicons. This method allows monitoring the occurrence of culturable microbes and can be implemented in applications requiring high phylogenetic resolution, reproducibility, low cost and high throughput such as identifying contamination and monitoring the evolution of diversity in mixed cultures and low diversity microcosms and periodic screening of small microbial culture libraries.

Keywords:  
ARISA-IGDF  
ITS  
Molecular fingerprinting  
qPCR

## Contents

1. Introduction . . . . .	111
2. Materials and methods . . . . .	112
3. Results . . . . .	113
3.1. RISA profiles . . . . .	113
3.2. Optimizing the PCR conditions . . . . .	113
3.3. Constructing fingerprints . . . . .	115
3.4. Confidence in phylotype identification . . . . .	116
4. Discussions . . . . .	117
5. Conclusions . . . . .	117
Acknowledgements . . . . .	118
References . . . . .	118

## 1. Introduction

Molecular approaches that aim to monitor changes in community structure in mixed cultures and low diversity microcosms are

challenged by considerations related to cost, small 16S rRNA gene sequence differences between otherwise distinct phylotypes, limited phylogenetic resolution of some fingerprinting methods, and potentially large changes in community structure over time. For monitoring mixed communities at subspecies-level resolution cloning/sequencing and metagenomic sequencing are impractical, and fingerprinting methods are sought that are fast, low cost and offer high phylogenetic resolution. Automated rRNA Intergenic Spacer Analysis (ARISA) is a high resolution repetitive PCR fingerprinting method based on the size variability of the

\* Corresponding author. Tel.: +1 503 725 9503; fax: +1 503 725 3888.

E-mail address: [rpopa@pdx.edu](mailto:rpopa@pdx.edu) (R. Popa).

<sup>1</sup> Present address: Department of Biology, Appalachian State University, 572 Rivers Street, Boone, NC 28608, USA.

16S–23S Intergenic Transcribed Spacers (ITSs) (Garcia-Martinez et al., 1999). To date, ARISA, and its simpler variant RISA, were applied to studying overall changes in diversity and to some lesser extent for microbial identification (Bergamo et al., 2004; Grassi et al., 2006; Heuer et al., 1997; Hewson et al., 2003; Jensen et al., 1993; Ni et al., 2007; Siering and Ghiorse, 1996; Torzilli et al., 2006). A number of modifications implemented in ARISA (such as qPCR, sequencing, group-specific primers, fluorescent densitometry, and statistical means to describe diversity and similarity) led to sharp electrophoretic peaks, high throughput, and high phylogenetic resolution (Acinas et al., 1999; Fisher and Triplett, 1999; Hewson et al., 2003; Offre et al., 2007; Ranjard et al., 2003; Robleto et al., 1998). Still, objective technical constraints restrict the usefulness of this method in very complex natural samples (Hewson and Fuhrman, 2006; Loisel et al., 2006). The most notable limitations of ARISA are: (a) use of taxonomically-ambiguous entities (*i.e.* Operational Taxonomic Units or OTUs); (b) hard to describe fingerprints of uncultured microbes; (c) numerous peaks when “universal primers” are used on complex samples; and (d) uneven confidence in identification across the phylogenetic spectrum. Because the concept of species remains ambiguous in *Bacteria*, and because more work is needed to define ARISA’s phylogenetic resolution, we prefer using the term *phyloptype* (rather than species, strain or OTU) to discuss the entities identified by this method. By extension, *phylogroup* represents a group of phylotypes revealed by a fingerprinting approach (*e.g.* specific PCR conditions, specific primers, specific numerical analysis for identifying amplicons, etc.). In this study the genus *Shewanella* was used as a model phylogroup, and the different strains of *Shewanella* showing distinct ARISA fingerprints are defined as phylotypes.

Our hypothesis was that an ARISA-based protocol can be developed that will allow fast, high throughput and low cost identification of microbial phylotypes from the *Shewanella* phylogroup at species and subspecies-level resolution. The variant of ARISA we explored identifies phylotypes based on the Intra-Genomic Diversity (IGD) of ITSs, hence IGD Fingerprinting (IGDF). We studied requirements, limitations and benefits of this approach and verified the effect of PCR conditions (*i.e.* different pairs of primers, duration of the final PCR extension, annealing temperature, number of PCR cycles and presence of formamide) on the PCR amplification yield, the formation of “ghost” peaks, resolution, peak shape and fingerprint reproducibility. We measured errors in estimating the size of amplicons and present an algorithm for quantifying the confidence in strain identification. We introduce a digitization protocol for ARISA-IGDF libraries that will facilitate the automation of fingerprint comparison.

## 2. Materials and methods

The genus *Shewanella* was selected because the genomes of the *Shewanella* strains sequenced to date revealed sufficiently high diversity of intra-genomic ITSs, and because *Shewanella*-specific 16S rRNA gene primers exist (Todorova and Costello, 2006). A total number of 44 *Shewanella* strains were obtained from four culture collections (Table 1). Genomic DNA was obtained by direct heat extraction from cell pellets (modified after: Cook and Meyers, 2003; Emerson and Moyer, 1997; Mazza et al., 2003; Siering and Ghiorse, 1996). After 24–48 h of growth in liquid LB medium in aerobic conditions cells were centrifuged (2 min; 7500 g; 20 °C), and washed twice in 500 µl TE buffer (10 mM Tris HCl; 1 mM EDTA; pH 8). After vortexing for 2 min with 2 mm glass beads the genomic DNA was extracted by heating for 10 min on a 100 °C water bath. The cellular debris was pelleted by centrifugation (3 min; 7500 g; 20 °C) and ~70–80% of the supernatant was recovered and used to quantify the dsDNA using the PicoGreen method (Molecular Probes, CA).

The DNA extracts were diluted to 1 µg/ml in ddH<sub>2</sub>O and 3.6 µl of each template solution was used in 10 µl PCR reactions (Fermentas Inc., MD) with 0.4 µl of each 10 µM primer. We used the primers: ITSF (5’GTCGTAACAAGGTAGCCGTA); ITSReub (5’GCCAAGGCATCCACC)

**Table 1**

Strains used in this study (nomenclature after Fredrickson et al., 2008).

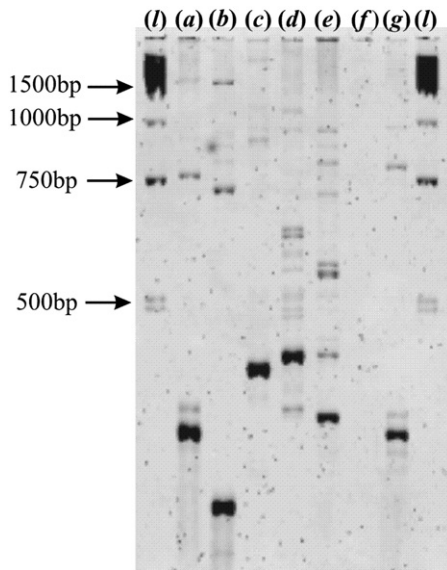
A	B	C	A	B	C
1	<i>Shewanella oneidensis</i> MRI	a	27	<i>Shewanella</i> sp. HRCR1(1A)	b
2	<i>Shewanella amazonensis</i> SB2B	a	28	<i>Shewanella</i> sp. HRCR2(1B)	b
3	<i>Shewanella baltica</i> OS155	a	29	<i>Shewanella</i> sp. HRCR3(1C)	b
4	<i>Shewanella frigidimarina</i> NCIMB400'	a	30	<i>Shewanella</i> sp. HRCR4(2A)	b
5	<i>Shewanella denitrificans</i> OS217	a	31	<i>Shewanella</i> sp. HRCR5(2B)	b
6	<i>Shewanella loihica</i> PV4	a	32	<i>Shewanella</i> sp. HRCR5(2B)	b
7	<i>Shewanella putrefaciens</i> 200'	a	33	<i>Shewanella</i> sp. HRCR7(3A)	b
8	<i>Shewanella putrefaciens</i> CN32	a	34	<i>Shewanella</i> sp. HR R8(3B)	b
9	<i>Shewanella</i> sp. MR7	a	35	<i>Shewanella</i> sp. HRCR9(3C)	b
10	<i>Shewanella</i> sp. PV4	a	36	<i>Shewanella</i> sp. HRCR10(4A)	b
11	<i>Shewanella frigidimarina</i> NCIMB400''	a	37	<i>Shewanella</i> sp. HRCR11(4B)	b
12	<i>Shewanella baltica</i> OS195	a	38	<i>Shewanella</i> sp. HRCR12(4C)	b
13	<i>Shewanella</i> sp. MR4	a	39	<i>Shewanella</i> sp. HRCR13(5A)	b
14	<i>Shewanella</i> sp. L1250	a	40	<i>Shewanella</i> sp. HRCR14(5B)	b
15	<i>Shewanella</i> sp. HRCR15(5C)	a	41	<i>Shewanella</i> sp. HRCR15(5C)	b
16	<i>Shewanella</i> sp. W3-18-1	a	42	<i>Shewanella</i> sp. HRCR18(4D)	b
17	<i>Shewanella</i> sp. ANA-3	a	43	<i>Shewanella</i> sp. HRCR19(5D)	b
21	<i>Shewanella</i> sp. HRCR21	b	44	<i>Shewanella</i> sp. HRCR20(5E)	b
22	<i>Shewanella</i> sp. HRCR22	b	45	<i>Shewanella</i> sp. OGI(A)	c
23	<i>Shewanella</i> sp. HRCR23	b	46	<i>Shewanella</i> sp. OGI45	c
24	<i>Shewanella</i> sp. HRCR24	b	47	<i>Shewanella</i> sp. OGI	c
25	<i>Shewanella</i> sp. HRCR25	b	48	<i>Shewanella</i> sp. OGI 104	c
26	<i>Shewanella</i> sp. HRCR26	b	50	<i>E. coli</i> CFT073	d

(A) Strain number in the PSU clone library. (B) Species and strain according to the source culture collection. (C) Source: (a) University of Southern California; (b) Pacific Northwest National Laboratory; (c) OGI School of Science and Engineering at Oregon Health and Sciences University; (d) ATCC. Strains # 34 and 50 do not belong to the genus *Shewanella* and were used as controls.

and Shew1259F (5'TACAGAGGGTTGCAAAGCC). The ITSF and ITSReub primers were proposed earlier as “bacterial universal” RISA primers (Cardinale et al., 2004) (3). Shew1259F is a putative “*Shewanella*-specific” primer redesigned after Todorova and Costello (2006). ITSReub was fluorescently labeled with Cy3, FAM or HEX; HEX has the advantage of non-overlapping fluorescence with SYBR green (used during qPCR amplifications), or with ROX (present in the ladder used for electrophoretic separation). Non-labeled primers were obtained from IDT DNA (Integrated DNA Technologies, IA); fluorescently-labeled primers were obtained from Operon Biotechnologies (AL). Formamide (Sigma, F9037) was used at a final concentration of 1.5%. The following range of PCR conditions was tested: denaturation (15 min at 95 °C); 35 to 50 cycles of amplification (30 s at 94 °C; 30 s annealing between 45 °C and 65 °C; and 1 min at 72 °C); and final extension (between 0 and 59 min at 72 °C). For common amplifications we used a GeneAmp 2400 thermocycler (Perkin Elmer) and a Fermentas PCR kit (Fermentas, CA); for temperature gradient PCR we used a Veriti Thermal Cycler (Applied Biosystems, CA) and the Fermentas PCR kit; for qPCR reactions we used a Stratagene MX3005P instrument and a SYBR Green PCR kit (QIAGEN, CA).

All PCR amplifications were separated by electrophoresis in 0.7% agarose/TAE gels (100V; 20 min), stained with ethidium bromide and visualized at 302 nm. For RISA profiles via polyacrylamide gel electrophoresis (PAGE) we used 5% polyacrylamide gels with 44:1 acrylamide:bis molar ratio (100 V; 5 C; 11 h), visualization by laser scanning fluorescence densitometry on a Typhoon 9200 Variable Mode gel imager using the program Typhoon scan; the images were analyzed with ImageQuant and adjusted for contrast, brightness and intensity.

For obtaining ARISA profiles 1 µl of each PCR product (or dilutions of it) was mixed with 12.1 µl HiDi formamide and 0.4 µl ROX1000 ladder (Applied Biosystems, CA), heat-denatured (5 min; 95 °C), cooled fast at –40 °C and injected in triplicates in a ABI 310 capillary sequencer with POP4 resin (Applied Biosystems, CA). Controls were also used based on different replicates and dilutions of ladder and *E. coli* amplicates. The electrophoretic separation occurred at 15 kV. Peak resolution decreased with fragment size and became >0.28 pb



**Fig. 1.** Electropherogram of RISA profiles of isolated strains. This analysis was used to verify a potential mislabeling in the *Shewanella* library. For PCR amplification we used ITSF/Cy3-ITSReub primers and 63 °C annealing temperature. Except for the lane (f) where 1:10 dilutions were used, all lanes were loaded with 5 µl of PCR products. (a) *S. oneidensis* MR1. (b) *S. amazonensis* SB2B. (c) *S. baltica* OS155. (d) *S. putrefaciens* CN32. (e) *S. putrefaciens* CN32. (f) *S. putrefaciens* CN32 and (g) *E. coli* CFT073. (l) Ladder.

above 500 bp; yet, because in some *Shewanella* strains some important diagnostic peaks are >500 bp long we analyzed peaks as large as 1000 bp. The electropherograms were downloaded in GeneScan v2.5 (ABI), analyzed with Peak Scanner v1.0 (ABI) to obtain fragment sizes, range and fluorescence, and the numerical data was processed in Microsoft Excel. We used the following cutoffs: (1) eliminated all peaks smaller than 82 bp (*i.e.* the sum between the size of the primers and the remaining parts of the 16S rRNA and 23S rRNA genes); (2) eliminated all peaks larger than 1000 bp; (3) eliminated all peaks with fluorescence <5 times the background (Hewson et al., 2003); and (4) eliminated all peaks with a fluorescence <0.1% from the total fluorescence of all larger peaks; and (5) analyzed ≤20 largest peaks for each strain. All PCR amplifications were produced in at least triplicates, and each PCR product was read in at least two dilutions.

Strain-diagnostic peaks were identified based on co-occurrence across replicates. For a specific peak to be validated as diagnostic it had to be present in 100% of ≤4 replicates, or ≥80% of ≥5 replicates. To determine the fluorescence of peaks saturating the detector we repeated electropherograms using diluted samples. Verifying if peaks from different electropherograms represent the same amplicons required determining the relationship between fragment size (in bp) and Standard Deviation (*SD*) of fragment size (Eq. (1)), and the relationship between peak abundance and the *SD* of abundance (Eq. (2)).

$$SD \text{ of fragment size (in bp)} = 0.000315 \cdot \text{Fragment size (in bp)} + 0.126 \quad (1)$$

(based on five replicate readings of one sample of *S. amazonensis* SB2B)

$$SD \text{ of peak abundance (in \%)} = 0.0684 \cdot \text{Peak abundance (in \%)} + 0.934 \quad (2)$$

(based on five independent PCR amplifications of *S. amazonensis* SB2B)

Two peaks, *x* and *y*, from two electropherograms were considered putatively dissimilar at 97% confidence, if the  $x \pm SD_x$  range did not overlap the  $y \pm SD_y$  range.

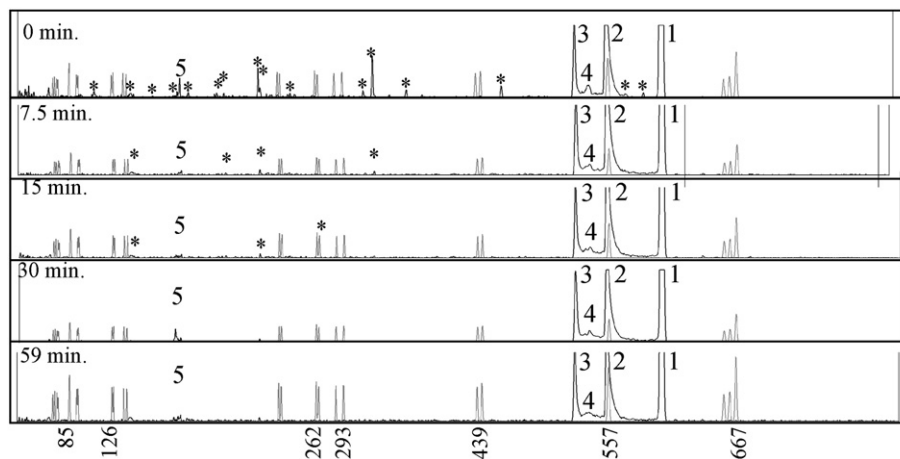
### 3. Results

#### 3.1. RISA profiles

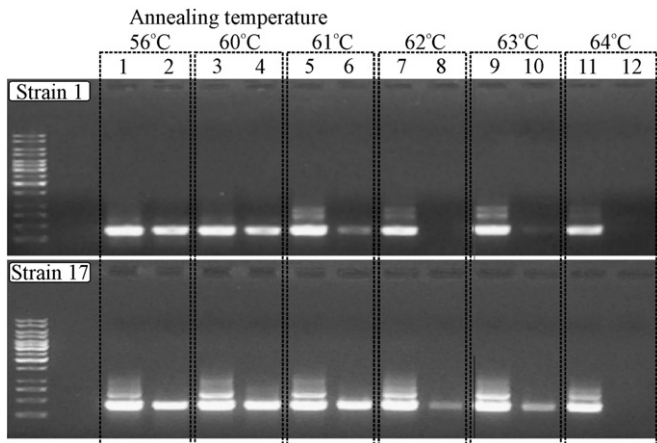
All strains were characterized first by RISA profiling using PAGE. Though this method is less precise than ARISA and adds ghost bands due to heteroduplexes and homoduplexes (Daffonchio et al., 2003), it has the advantage of being fast, allowing large scale and cost effective screening, identifies cross-contamination and mislabeled strains in microbial libraries (Fig. 1), and helps in formulating dilution series for sequencer-based fingerprinting.

#### 3.2. Optimizing the PCR conditions

Increasing the duration of the final 72 °C extension in PCR reactions helps in minimizing peak shoulders caused by TAQ polymerase-induced artifacts (Torzilli et al., 2006). Expecting that too short final



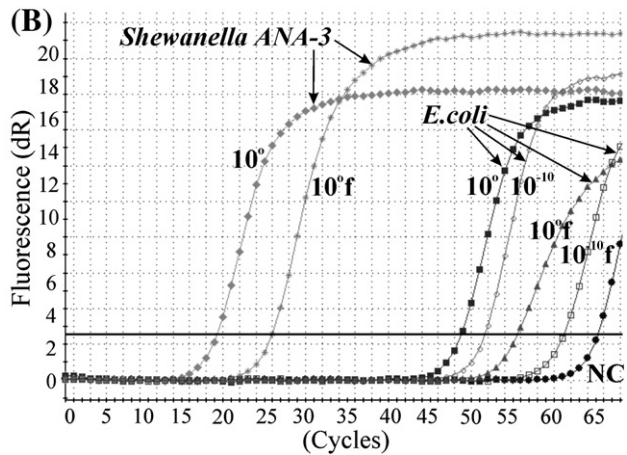
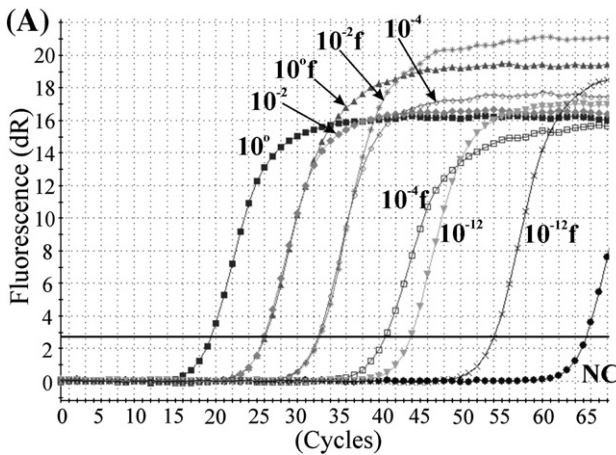
**Fig. 2.** The effect of the duration of final extension at 72 °C on ARISA-IGD ghost peaks. The electropherograms present PCR products of *S. amazonensis* SB2B amplified with the Shew1259F/FAM-ITSReub primers at five 72 °C extension times (0 min; 7.5 min; 15 min; 30 min; and 59 min). The numbered peaks (1–5) are the ARISA-IGD peaks that are diagnostic for this strain and combination of primers. The unlabeled peaks represent the ladder (ROX1000). The peaks labeled with (\*) are the ghost peaks. The values on the x axis represent sizes (in bp) for some of the ladder peaks. Electropherograms were stretched vertically for peak #3 to reach 100% and thus allow image comparisons.



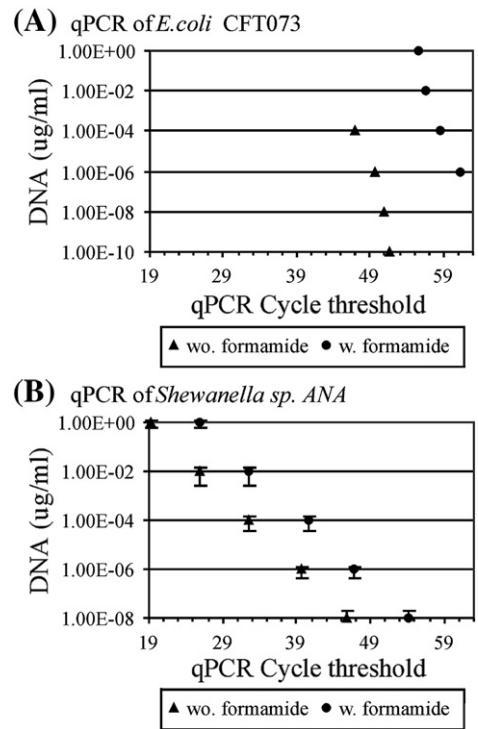
**Fig. 3.** The effect of annealing temperature and 1.5% formamide on the PCR amplification in *S. oneidensis MR1* (Strain #1) and *Shewanella sp. ANA* (Strain #17). The ladder, at the far left, was 1 kb IDNA (Fermentas). The sample gel load was 5  $\mu$ l of PCR products. Odd numbers represent the no-formamide treatments; even numbers represent the formamide treatments. The electropherograms were obtained on 0.7% agarose/TAE gels, (120 V; 30 min), stained with ethidium bromide and visualized at 302 nm.

extension will produce false positives (*i.e.* “ghost” peaks), we verified the effect of five extension times on ARISA-IGDF patterns on DNA from *S. amazonensis SB2B* (Fig. 2, Q2404, p45). We found no significant differences in peak shape between the different extension times (results not shown), and that *ghost* peaks existed in all <30 min extension treatments. A final extension of 45 min at 72 °C was used in all subsequent PCR amplifications.

Formamide was reported to increase primer specificity during PCR reactions (Grassi et al., 2006; Sarkar et al., 1990; Zhang and Albert, 1996), and thus may help in implementing group-specific primers in ARISA-IGD fingerprinting. However, formamide may also inhibit the PCR reaction when appropriate primers are used on target phylogroups. We verified the effect of 1.5% formamide on the maximum annealing temperature on eight strains of *Shewanella* (#1, #2, #3, #7, #8, #9, #11 and #17; see: Table 1) using the primers Shew1259F/ITSReub (two examples are given in Fig. 3). Based on these eight strains the best annealing temperature for *Shewanella* was 64 °C, and dropped by ~3–4° when formamide was used. Peak patterns (mainly relative abundances of different peaks) were also affected by formamide (results not shown). To verify the effect of formamide on the overall efficiency of amplification with the primers Shew1259F/

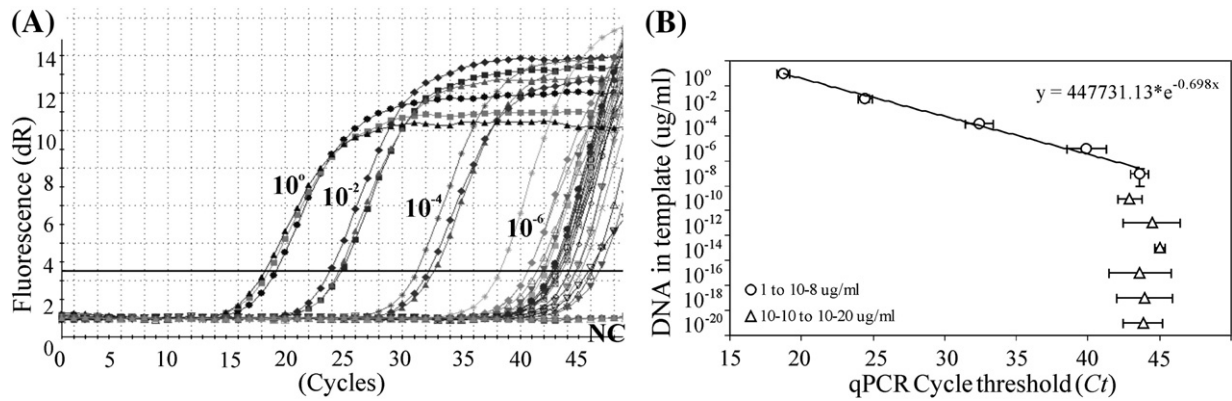


**Fig. 4.** qPCR results showing the effect of 1.5% formamide (f) on the amplification of genomic DNA from two strains using the primers Shew1259F/HEX-ITSReub at 60 °C annealing temperature. (A) = *Shewanella sp. ANA-3*. (B) = *E. coli CFT073* and two treatments with *Shewanella sp. ANA-3* used for comparison. NC = negative control. Fluorescence (on the y axis) is given in relative units (= fluorescence/1000). Seven dilutions were verified for each strain starting from 1  $\mu$ g/ml in  $10^{-2}$  decrements. To avoid overcrowding the graphs, only a few relevant amplification plots are shown.



**Fig. 5.** The efficiency of qPCR amplification of genomic DNA from *Shewanella sp. ANA-3* with the primers Shew1259F/HEX-ITSReub, at 60 °C annealing temperature with and without 1.5% formamide. The y axis indicates the concentration of DNA in template solutions. Error bars are 1SD of DNA concentrations calculated from the variance of the cycle threshold (Ct).

ITSReub we produced serial dilutions ( $10^0$  to  $10^{-12}$  of genomic DNA; where  $10^0 = 1 \mu\text{g DNA ml}^{-1}$  in the template solutions), and amplified by qPCR. All qPCR amplifications were done in triplicate. Formamide inhibited the amplification of *Shewanella* DNA about 100 times (Figs. 4A and 5), with no notable changes in ARISA patterns and variability (discussed below). Although Shew1259F is supposed to be “*Shewanella*-specific” and the annealing temperature was high (60 °C), amplification occurred eventually in a control strain of *E. coli* as well if enough cycles were applied (Fig. 4B). This may partly be due to using the “bacterial universal” ARISA reverse primer (ITSReub), which lowers the stringency of amplification. Overall, the efficiency of amplifying *E. coli* DNA under these conditions was very low ( $<10^{-7}$ ) relative to *Shewanella*, the variability of the cycle threshold (Ct) in *E.*



**Fig. 6.** Variance (A) and detection limits (B) based on qPCR amplification of a mixture of ten strains of *Shewanella* (#1, #2, #3, #7, #8, #9, #11, #12, #14 and #15; see Table 1), with the primers Shew1259F/ITSReub, at 60 °C annealing temperature without formamide. The DNA concentration in the template was 1 µg/ml for the 10<sup>0</sup> dilution, all strains were present in equal DNA amounts and all amplifications were performed in triplicate. In (A) the unlabeled plots represent dilutions > 10<sup>6</sup> µg/ml. NC = negative control. Data for all dilutions and replicates were used to construct (B). Although amplification occurred at all concentrations good correlation between Ct and DNA concentration was found only in the range of 1 to 10<sup>-8</sup> µg/ml DNA. The error bars in (B) represent 1SD of Ct.

*coli* amplifications was very large (results not shown) and the DNA concentration vs. Ct dependency was steep. Therefore the uncertainty when quantifying the concentration of *E. coli* DNA from Ct remains large. Formamide had a stronger inhibitory effect on the amplification of the control *E. coli* DNA than on *Shewanella* DNA. Based on these results, all PCR amplifications used to construct the *Shewanella* ARISA-IGDF library were made at 60 °C annealing without formamide.

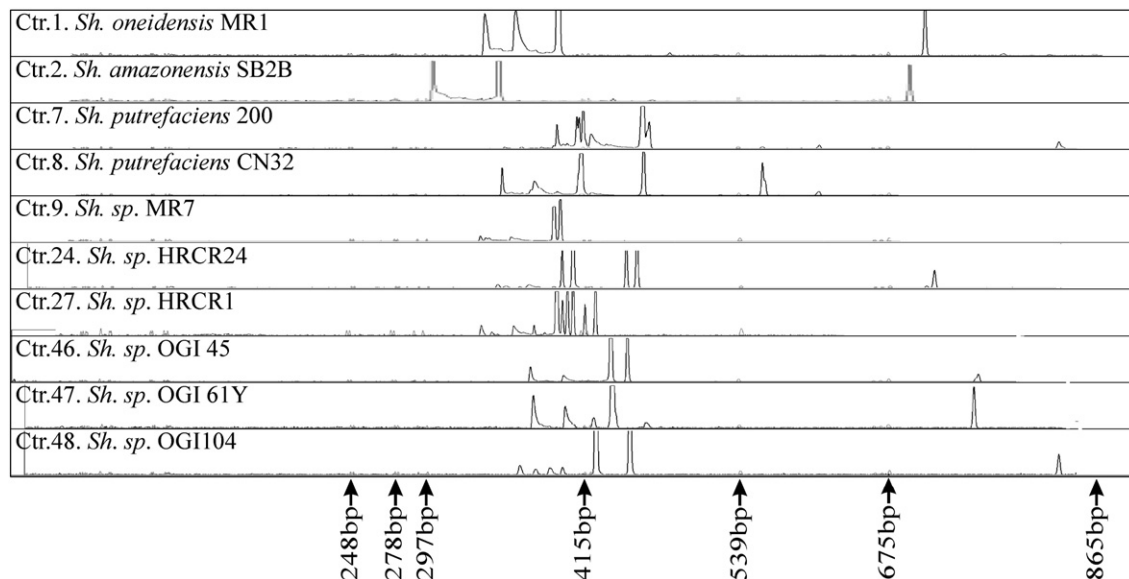
To define the relationship between the concentration of DNA template and the optimal number of PCR cycles in ARISA-IGDF, we used qPCR and measured the overall efficiency of amplification of *Shewanella* DNA using a mixture of equal proportions of genomic DNA from ten strains of *Shewanella* (Fig. 6). The optimal range for quantifying genomic DNA as a function of the qPCR's cycle threshold (Ct) results was 10<sup>-8</sup>–1 µg/ml, with a Ct ranging between 17 and 40. At lower template concentrations the variance of Ct became too large to allow confident DNA quantification. With an appropriate number of replicates such results can be used to determine errors when quantifying DNA using specific combinations of phylogroup, primers and PCR conditions.

Last but not least, the logarithmic nature of the PCR amplification makes the number of PCR cycles influence the relative fluorescence of the different peaks. Based on the results presented above we have

established that obtaining consistent results in ARISA-IGDF requires rigorous and stringent PCR conditions; for the construction of our *Shewanella* library we used no formamide, 35 PCR cycles, 60 °C annealing temperature, made amplifications in independent triplicates and fingerprinted the PCR products in two dilutions (1:1 and 1:7).

### 3.3. Constructing fingerprints

In ARISA the electrophoretic bands represent polymers and thus, in the ideal case, averages can be rounded to closest integer values (Hewson and Fuhrman, 2006). However, in most electropherograms peak position precision remains hard to achieve, an uncertainty that increases significantly above 500 bp. Mistakes in assigning integer values to DNA amplicons, and thus in phylotype identification, made in early steps of analysis have large consequences on fingerprint comparisons, and are very time consuming to correct afterward. In the 82–500 bp range this problem can be addressed by increasing the number of replicates and by using ladders with denser peaks; yet, this approach becomes cost prohibitive, especially for large DNA fragments where the intrinsic errors of the method are large. We chose using curve



**Fig. 7.** Analogic ARISA-IGD profiles in the 100–1000 bp range for ten strains of *Shewanella* from PCR amplifications with ITSF/FAM-ITSReub without formamide. Ctr.#s correspond to PSU strain numbers from Table 1. The PCR cycle consisted of: (1) thermal activation of TAAQ and denaturation (15 min; 95 °C); (2) amplification (35 cycles; 30 s at 95 °C; 30 s at 63 °C; and 20 s at 72 °C); and (3) final extension (45 min; 72 °C). Numbers on the x axis represent the positions of selected peaks from the ladder.

**Table 2**  
Digital fingerprint of *Shewanella baltica* OS155 obtained with the primers ITSF/ITSReub.

Peak number	Average peak size (bp)	SD of peak size (bp)	Mean peak abundance (%)	SD of peak abundance (%)
1	438.743	0.539	0.212	0.206
2	163.423	0.330	0.214	0.206
3	318.715	0.448	0.219	0.206
4	143.603	0.315	0.415	0.206
5	426.527	0.529	0.454	0.206
6	524.339	0.604	0.529	0.207
7	863.019	0.860	0.608	0.207
8	774.529	0.793	0.661	0.207
9	387.718	0.500	1.064	0.207
10	772.015	0.791	1.404	0.207
11	406.322	0.514	6.333	0.211
12	375.817	0.491	7.734	0.212
13	796.581	0.810	8.221	0.212
14	447.056	0.545	11.091	0.215
15	408.696	0.516	11.976	0.215
16	377.369	0.492	12.935	0.216
17	445.937	0.544	13.345	0.216
18	448.300	0.546	22.583	0.223

based estimates of peak sizes and variances obtained from three or more replicates and not assigning integer values to the different DNA fragments. Because in some applications only one or two electropherograms may be practical, or because when comparing peaks between two different electropherograms intrinsic values variance are not available, we used the errors in interpreting peak as a function of fragment size (Eqs. (1) and (2)) to obtain variance estimates.

A systemic difficulty in this analysis was the asymmetry of peaks (*i.e.* peak shoulders), which can be revealed based on differences between mean and median within the fluorescence data distribution for each peak. Measuring peak height rather than peak area will lower the quantitative determination of fragment abundance, but will help limit the contribution of the surrounding peaks. One approach we have verified to correct peak asymmetry (*i.e.* peak fronting and peak tailing), was to use peak-deconvolution algorithms to deconstruct asymmetric peaks into a set of smaller peaks. Peak deconvolution is a legitimate approach for analyzing polymers that are different by discrete size intervals, and has the benefit of producing normal peak distribution, thus allowing parametric analysis. Though peak deconvolution is time consuming and impractical in large scale fingerprinting, for very small data sets it may help resolve some of the partly overlapping peaks. We discourage this analysis in ARISA-IGDF, because in this method peak variability and peak asymmetry may be produced by many hard to define and reproduce artifacts, and will result in a number of questionable peaks.

Analogical fingerprints are graphs showing peak abundances relative to fragment sizes. Obtaining them was facilitated by prior description of “*Shewanella*-specific” primers, and by the sufficiently numerous and variable ITS amplicons in the 82–1000 bp range for this

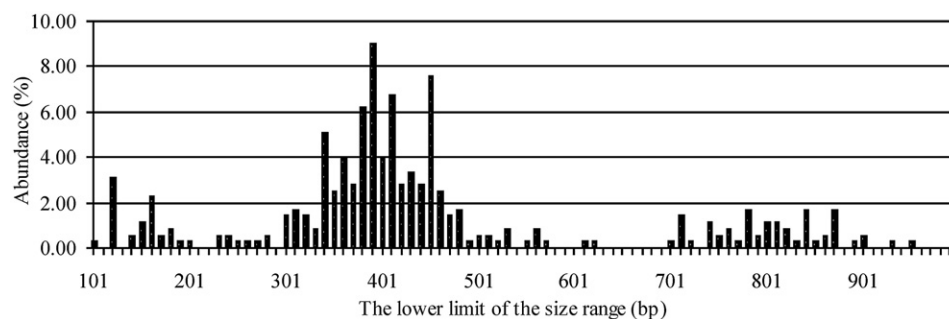
phylogroup. Some strains of *Shewanella* have a couple of ITS peaks that are >1000 bp large (Fig. 1), and in some phylogroups (such as *Rhizobia*) ITSs fragments larger than 1 kb are frequent. Such large fragments are more difficult to measure because the electrophoretic separation becomes too long and errors increase significantly above 1000 bp. If ARISA-IGDF is restricted to appropriate phylogroups ( $\geq 3$  peaks within the 82–1000 bp range) sufficiently high confidence in strain identification can be obtained (see below). Examples of analogical *Shewanella* ARISA-IGD fingerprints illustrating the diversity of this phylogroup and the phylogenetic resolution of this method are given in Fig. 7.

Digitizing ARISA-IGD fingerprints has a number of benefits: (1) facilitates applying a variety of statistical procedures; (2) makes data compatible with other libraries created under comparable conditions; (3) facilitates comparisons with *in silico* ARISA-IGD fingerprints predicted from genomic analysis; and (4) will help in developing software-based screening of fingerprint libraries. One of our goals was to establish a numerical convention that will help formulate a minimalist set of data for digital ARISA-IGD fingerprinting. The size of different ARISA fragments was determined using the GeneScan v2.5 software and validated with Peak Scanner v1.0. We eliminated some of the peaks using the cutoff criteria given in Materials and methods. As different peaks vary in fluorescence considerably, in most samples determining peak abundances required combining data from electropherograms obtained from different dilutions, and remains semi-quantitative. We used Eqs. (1) and (2) to assign SD to individual peaks, and cross-compared all peaks from different electropherograms of independent replicates of the same strain (as shown in Materials and methods) to identify similar peaks. The peaks validated for each strain were used to construct a digital fingerprint; one example of such fingerprint is given in Table 2.

We analyzed digital ARISA-IGD fingerprints of 44 *Shewanella* strains (Table 1) using two pairs of primers (ITSF/ITSReub and Shew1259F/ITSReub) and the following PCR conditions: 0.34 mg template DNA ml<sup>-1</sup> PCR reaction; denaturation (15 min; 95 °C); 35 cycles (30 s at 94 °C; 30 s at 60 °C; 1 min at 72 °C); and final extension (45 min at 72 °C). Differences between the abundances of some ARISA-IGD peaks were very large (*e.g.*  $\geq 50:1$ ); hence, in most cases peak abundance ratios could not be established from a single dilution. Some peaks were too large and saturated the detector (in undiluted 1:1 samples); others were too small to be identified in diluted samples ( $\geq 10:1$  dilution). For each phylotype and for each pair of primers obtaining ARISA-IGD fingerprints requires six sequencer readings (*i.e.* three independent PCR amplifications and two 1:7 dilutions for electropherograms).

### 3.4. Confidence in phylotype identification

Confidence in identification using ARISA-IGDF depends on parameters such as: the number of peaks validated for each phylotype; the maximum number of peaks distinguishable in a electropherogram



**Fig. 8.** The size distribution of the ARISA-IGD peaks for the *Shewanella* phylogroup based on the primers ITSF/ITSReub. The graph shows the distribution of 256 peaks from 36 strains. For this analysis 10 bp size ranges were used and we calculated the  $p_i$  values for each size range.

and the peak position variability; the total number of phylotypes that can be amplified by a specific set of primers and under specific PCR conditions; the diversity of peak sizes for the phylotypes belonging to a certain phylogroup; and the relative abundance of peaks in each electropherogram. In the simplest case, if the SD of the fragment size is  $<1$  bp (Eq. (1)), up to  $\sim 300$  distinct peaks may be possibly discriminated in the 82–1000 bp range, with  $\sim 97\%$  confidence (i.e. non-overlapping range for the average  $\pm 1SD$  between two adjacent peaks). The probability of an incorrect identification, i.e. the probability for two phylotypes to share a similar number ( $n$ ) of ARISA bands of similar length and by chance alone is:

$$p = 1 / 300^n. \quad (3)$$

For evolutionary reasons the distribution of the ARISA-IGDF amplicons within a phylogroup is not even across the 82–1000 bp size range (Fig. 8). This distribution will probably vary significantly among different phylogroups and with different primers. After analyzing fingerprints of 36 strains of *Shewanella* from Table 1 we divided this range in 10 bp size intervals, accounted all peaks and calculated the probability for a peak to be present in each specific size interval. Using these  $p_i$  values the probability of an incorrect identification becomes:

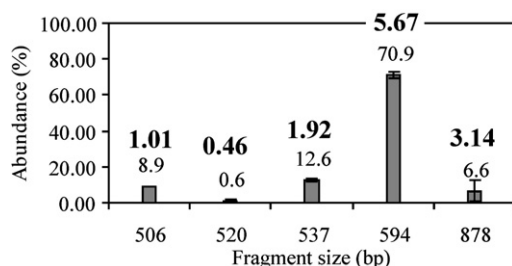
$$p(n) = \frac{1}{\prod_{i=1}^m ip_i} \quad (4)$$

where:

- $n$  the total number of validated peaks of a phylotype;
- $m$  the total number of size range intervals;
- $p_i$  the probability for a peak to belong to a specific size range.

Using Eqs. (3) and (4) it can be shown that if  $\geq 4$  ARISA-IGDF peaks are present (belonging to four different size ranges), the chance for two different phylotypes of *Shewanella* to share similar patterns is  $\sim 1/300^4$  (i.e. one in  $8.1 \cdot 10^9$ ). No estimates exist regarding the total number of *Shewanella* phylotypes in nature, or the number of microbes that amplify with these primers at 60 °C; but it is highly unlikely for this number to be anywhere near  $10^9$ . Hence, phylotypes having dissimilar ARISA-IGDF patterns consisting of  $\geq 4$  peaks in the 82–1000 bp range can be confidently identified by this method.

When two microbial clones produce amplicons of similar sizes, verifying that they have identical ARISA-IGD fingerprints is helped by ranking peaks based on relative abundances. Fig. 9 gives an example of graphical representation of peak relative abundances. When using relative abundances to compare fingerprints of different strains we recommend using Shannon's index ( $H'$ ) rather than Simpson's index



**Fig. 9.** Relative abundances of five amplicons from an ARISA-IGD profile of *S. amazonensis* SB2B based five independent replicates and using the primers ITSF/ITSReub. The non-bold font numbers in the graph are the relative abundance of each peak relative to the total fluorescence of all peaks. Error bars are 1SD of relative abundances; the SD values are given in bold font.

(D), which would introduce more biases due to the most fluorescent peaks.

#### 4. Discussions

We explored the capacity of ARISA-IGDF to distinguish microorganisms from a target phylogroup by analyzing 44 strains from the genus *Shewanella*. The stringency of primer/phylogroup recognition improved by increasing the annealing temperature and by adding 1.5% formamide to the PCR mixture. Formamide had larger inhibition outside the target phylogroup, produced small changes in fingerprint patterns (e.g. some minor peaks were harder to identify and required less stringent cutoffs to reveal), inhibited ( $\sim 100$  times) the amplification of the target phylogroup (*Shewanella*) and decreased the annealing temperature by  $\sim 3\text{--}4$  °C. The number of PCR cycles was shown to influence the relative abundance of different amplicons (Daffonchio et al., 2003) and is a key requirement for the reproducibility of this method. Increasing the duration of final extension (72 °C) to  $\geq 30$  min eliminated false peaks and made fingerprints more reproducible. High confidence in identification ( $\sim 8 \cdot 10^{-9}$ ) can be obtained when using phylotypes with  $\geq 4$  distinct ARISA-IGD amplicons in the 82–1000 bp range.

ARISA-IGDF retains a couple of limitations. Amplification of DNA HRCR-8 and *E. coli* which do not belong to the *Shewanella* phylogroup occurred with the primers Shew1259F/ITSReub (Shew1259F is supposed to be *Shewanella*-specific). In such instances using two primers that are group-specific, and increasing the stringency of the PCR conditions by increasing the annealing temperature and adding 1.5% formamide is recommended. Even if 16S rRNA gene sequence information is available, patterns obtained with the Shew1259F/ITSReub primers can only be useful to a limited extent to predict patterns obtained with ITSF/ITSReub primers, and very little for fragment abundance analysis. This is a primarily a mispriming problem and may be caused by minute sequence differences in the priming areas, differences in optimal annealing conditions between primers, and the presence of the 5'-dye. ARISA-IGD fingerprinting is warped toward culturable microbes, while most species of microbes cannot be isolated in pure cultures. Unresolved peaks from complex samples will be difficult to assign without sequencing the flanking 16S rRNA genes as well. Even if good group-specific primers are available, the efficiency of amplification of different strains within a target phylogroup may still remain uneven, and the fingerprinting of microbes with 1–2 ITSs per genome (such as: *Sulfolobus*, *Magnetospirillum*, and *Magnetococcus*) will require ARISA/tRFLP combinations. More work is needed to compare the performance of ARISA-IGDF with other methods of molecular fingerprinting of high (SSCP, ITS-HHP, Multi Locus Sequence Typing, LH-PCR, Rep-PCR and RAPD) and low phylogenetic resolution (DGGE, TGGE and tRFLP).

#### 5. Conclusions

We propose that ARISA-IGDF is a low cost high resolution approach to identify the presence of specific microbes from phylogroups of culturable microorganisms for which group-specific ARISA primers are available. If future work can verify the predictability of ARISA-IGD fingerprints from *in silico* analysis of genomic data, (and also because once a fingerprint was established there is little need for future sequencing), ARISA-IGDF can be used to complement metagenomic studies in mixed cultures and low diversity microcosms. This method can be applied to monitor the evolution of diversity and the spatial distribution and dynamics of different microorganisms in a little complex community. Other potential applications of ARISA-IGDF include: tracking specific target microbes; identifying sources of microbial contamination; identifying contamination in axenic cultures and routine screening of libraries of microbial cultures.



## Acknowledgements

This work was supported by a research grant from a NASA Astrobiology Program (NNH07ZDA001N-EXOB) and by a 2008 Portland State University Faculty Enhancement Grant. We thank Dr. Mitch Cruzan and Trieste Dobberstein (PSU) for support with using the ABI 310 sequencer, Dr. Niles Lehman (PSU) for access to the Typhoon gel reader, Dr. James K. Fredrickson (PNNL) for *Shewanella* strains from the PNNL culture collection, Dr. Ken Stedman for access to the gradient PCR instrument and Dr. Kenneth Nealson, Dr. Ana Obratsova and Shana Rapoport for strains from the USC culture collection.

## References

- Acinas, S.G., Anton, J., Rodriguez-Valera, F., 1999. Diversity of free-living and attached bacteria in offshore western Mediterranean waters as depicted by analysis of genes encoding 16S rRNA. *Appl. Environ. Microbiol.* 65, 514–522.
- Bergamo, R.F., Novo, M.T.M., Verissimo, R.V., Paulino, L.C., Stoppe, N.C., Sato, M.I.Z., Manfio, G.P., Prado Jr., P.I., Garcia, O., Ottoboni, L.M.M., 2004. Differentiation of *Acidithiobacillus ferrooxidans* and *A. thiooxidans* strains based on 16S–23S rDNA spacer polymorphism analysis. *Res. Microbiol.* 155, 559–567.
- Cardinale, M., Brusetti, L.P., Quatrini, S., Borin, A.M., Puglia, A., Rizzi, E., Zanardini, C., Sorlini, C., Corselli, B., Daffonchio, D., 2004. Comparison of different primer sets for use in automated ribosomal intergenic spacer analysis of complex bacterial communities. *Appl. Environ. Microbiol.* 70, 6147–6156.
- Cook, A.E., Meyers, P.R., 2003. Rapid identification of filamentous actinomycetes to the genus level using genus-specific 16S rRNA gene restriction fragment patterns. *Int. J. Syst. Evol. Microbiol.* 53, 1907–1915.
- Daffonchio, D., Cherif, A., Brusetti, L., Rizzi, A., Mora, D., Boudabous, A., Borin, S., 2003. Nature of polymorphisms in 16S–23S rRNA gene intergenic transcribed spacer fingerprinting of *Bacillus* and related genera. *Appl. Environ. Microbiol.* 69, 5128–5137.
- Emerson, D., Moyer, C., 1997. Isolation and characterization of novel iron-oxidizing bacteria that grow at circumneutral pH. *Appl. Environ. Microbiol.* 63, 4784–4792.
- Fisher, M.M., Triplett, E.W., 1999. Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Appl. Environ. Microbiol.* 65, 4630–4636.
- Fredrickson, J.K., Romine, M.F., Beliaev, A.S., Auchtung, J.F., Driscoll, M.E., Gardner, T.S., 2008. Towards environmental systems biology of *Shewanella*. *Nature* 455, 592–603.
- Garcia-Martinez, J., Acinas, S.G., Anton, A.I., Rodriguez-Valera, F., 1999. Use of the 16S–23S ribosomal genes spacer region in studies of prokaryotic diversity. *J. Microb. Methods* 36, 55–64.
- Grassi, M., Volpe, E., Colizzi, V., Marian, F., 2006. An improved, real-time PCR assay for the detection of GC-rich and low abundance templates of *Mycobacterium tuberculosis*. *J. Microb. Methods* 64, 406–410.
- Heuer, H., Krsek, M., Baker, P., Smalla, K., Wellington, E.M.H., 1997. Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. *Appl. Environ. Microbiol.* 63, 3233–3241.
- Hewson, I., Fuhrman, J.A., 2006. Improved strategy for comparing microbial assemblage fingerprints. *Microb. Ecol.* 51, 147–153.
- Hewson, I., Vargo, G.A., Fuhrman, J.A., 2003. Bacterial diversity in shallow oligotrophic marine benthos and overlying waters: effects of virus infection, containment, and nutrient enrichment. *Microb. Ecol.* 46, 322–336.
- Jensen, M.A., Webster, J.A., Straus, N., 1993. Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms. *Appl. Environ. Microbiol.* 59, 945–952.
- Loisel, P., Harmand, J., Zemb, O., Latrille, E., Lobry, C., Delgenes, J.P., Godon, J.J., 2006. Denaturing gradient electrophoresis (DGE) and single-strand conformation polymorphism (SSCP) molecular fingerprintings revisited by simulation and used as a tool to measure microbial diversity. *Environ. Microbiol.* 8, 720–731.
- Mazza, P., Monciardini, P., Cavaletti, L., Sosio, M., Donadio, S., 2003. Diversity of *Actinoplanes* and related genera isolated from an Italian soil. *Microb. Ecol.* 45, 362–372.
- Ni, Y.Q., Yang, Y., Bao, J.T., He, K.Y., Li, H.Y., 2007. Inter- and intraspecific genomic variability of the 16S–23S intergenic spacer regions (ISR) in representatives of *Acidithiobacillus thiooxidans* and *Acidithiobacillus ferrooxidans*. *FEMS Microbiol. Lett.* 270, 58–66.
- Offre, P., Pivato, B., Siblot, S., Gamalero, E., Corberand, T., Lemanceau, P., Mougel, C., 2007. Identification of bacterial groups preferentially associated with mycorrhizal roots of *Medicago truncatula*. *Appl. Environ. Microbiol.* 73, 913–921.
- Ranjard, L., Lejon, D.P.H., Mougel, C., Schehrer, L., Merdinoglu, D., Chaussod, R., 2003. Sampling strategy in molecular microbial ecology: influence of soil sample size on DNA fingerprinting analysis of fungal and bacterial communities. *Environ. Microbiol.* 5, 1111–1120.
- Robbleto, E.A., Borneman, J., Triplett, E.W., 1998. Effects of bacterial antibiotic production on rhizosphere microbial communities from a culture-independent perspective. *Appl. Environ. Microbiol.* 64, 5020–5022.
- Sarkar, G., Kapelner, S., Sommer, S.S., 1990. Formamide can dramatically improve the specificity of PCR. *Nucleic Acids Res.* 18, 7465.
- Siering, P.L., Ghiorse, W.C., 1996. Phylogeny of the *Sphaerotilus–Leptothrix* group inferred from morphological comparisons, genomic fingerprinting, and 16S ribosomal DNA sequence analyses. *Int. J. Syst. Bacteriol.* 46, 173–182.
- Todorova, S.G., Costello, A.M., 2006. Design of *Shewanella*-specific 16S rRNA primers and application to analysis of *Shewanella* in a minerotrophic wetland. *Environ. Microbiol.* 8, 426–432.
- Torzilli, A.P., Sikaroodi, M., Chalkley, D., Gillevet, P.M., 2006. A comparison of fungal communities from four salt marsh plants using automated ribosomal intergenic spacer analysis (ARISA). *Mycologia* 98, 690–698.
- Zhang, W., Albert, B., 1996. Direct PCR sequencing with denaturants (formamide). *Methods Mol. Biol.* 65, 137–147.