BASELINE

SELECTION AND IMPROVING OF FIT-FOR-PURPOSE SAMPLING PROCEDURES FOR SPECIFIC FOODS AND RISKS

PROJECT DELIVERABLE

Collaborative project	Large-scale integrating project
Project acronym	BASELINE
Project title	Selection and improving of fit-for-purpose sampling procedures for specific foods and risks
Grant Agreement number	222738
Date of latest version of Annex I	22/01/2013

Del. No.	Deliverable name	WP no.	Lead partecipant	Nature	Dissemination Level	Due delivery date from Annex I
D.2.3	A prototype of DNA-array method for detection and typing of selected *Salmonella* serotypes	2	UCPH	P	P	43

Delivery Date 08/05/2013

Project co-ordinator

Prof. Gerardo Manfreda

Alma Mater Studiorum – Università di Bologna

Tel: +39 051 20 9 785
E-mail: gerardo.manfreda@unibo.it

Project website address www.baselineeurope.eu

TABLE OF CONTENTS

SUMMARY and PURPOSE OF STUDY 3

METHODS 4

The array platform 4

The protocol 4

Data analysis 6

RESULTS 7

DISCUSSION 10

Project funded under the Food, Agriculture and Fisheries, and Biotechnology theme (KBBE)

SEVENTH FRAMEWORK PROGRAMME

SUMMARY and PURPOSE OF STUDY

This report describes the development of a diagnostic method meant for serotype specific separation of *Salmonella* serotypes. A DNA array with 15000 random sequences is used. DNA from bacteria is labeled and hybridized to the array, and recognition is performed based on pattern analysis.

While species specific identification was shown to be possible by this method, it was not possible to separate *Salmonella* serotypes in a statistically valid way.

The delivery is an evaluation of a universal DNA micro array prototype with random probe sequences for discrimination of *Salmonella* serovars.

Project funded under the Food, Agriculture and Fisheries, and Biotechnology theme (KBBE)

SEVENTH FRAMEWORK PROGRAMME

METHODS

The array platform

The array is a custom Agilent 8x15.000 micro array. Agilent produces the slides on demand.

On each slide are printed eight arrays each consisting of 15.000 probes. The probes have a length of 20 base pairs plus linker DNA.

On the array are included two internal controls:

1. The bacterial 16s rRNA eub-338 probe with decreasing homology.
2. Viral Hemorrhagic Septicaemia(VHS) DNA with decreasing homology.

The eub-338 probe will anneal to the added fragmentized chromosomal DNA and the VHS probe will anneal to added 101mer synthetically produced DNA. The latter allows quantification of mis-annealing signal strength.

The remaining probes are designed to consist of randomized nucleotides with a melting temperature of 60°C and a GC level 35-60%.

The protocol

For the identification of bacterial species, typing and community analysis with the array the following protocol is used.

Purification of chromosomal DNA from pure cultures.

Purification is performed with the Qiagen - DNeasy Blood & Tissue Kit, using a slightly modified protocol.

The supplemental protocol “Pretreatment for Gram-Positive Bacteria” is modified in step 4 and 5 to the following:

- 25 µl Proteinase K is added followed by incubation at 56°C for at least 30 min.
- 4 µlRNase A (100 mg/ml) is added and the solution is incubated for 10 min at room temperature.
- 200 µl Buffer AL (without ethanol) is added followed by vortex mixing.

The protocol “Purification of Total DNA from Animal Tissues (Spin-Column Protocol)” is followed from step 4, with the following modifications.

- In step 6, the AW2 buffer is exchanged with fresh 80% ethanol. After the first centrifugation the collection tube is emptied and the spin column is placed back in the tube. An extra centrifugation at 20.000 x g for 2 minutes is performed.
- In step 7, 200 µl elution buffer (AE) is added to the spin column followed by incubation at 65°C for 5 minutes. Centrifugation is performed as described in the protocol.

The DNA quantity and quality is investigated using the NanoDrop ND-1000 UV-VIS spectrophotometer and by running samples on a 1% agarose gel containing ethidium bromide.

Amplification of chromosomal DNA from pure cultures and more complex samples.

This protocol has been used with pure cultures and also with milk samples.

Project funded under the Food, Agriculture and Fisheries, and Biotechnology theme (KBBE)

SEVENTH FRAMEWORK PROGRAMME

Amplification is performed with the GenomePlex® Complete Whole Genome Amplification (WGA2) kit from Sigma-Aldrich. The amplification is performed following the manufacturer's instructions with minor changes.

- The random fragmentation step in the protocol was omitted, instead a 1 ml sample is spun down and the pellet resuspended in 300 µl MilliQ followed by heat treatment at 95 °C for 10min.
- The debris is spun down and 10 µl supernatant used for the OmniPlex Library Preparation and Whole Genome Amplification following the manufacturer's instructions.

The DNA quantity and quality is investigated using the NanoDrop ND-1000 UV-VIS spectrophotometer and by running samples on a 1% agarose gel containing ethidium bromide.

Go to step 3 in the protocol.

Fragmentation of the chromosomal DNA with enzymes supplied by Ferments.

- For each strain 2 µg chromosomal DNA is divided into three tubes. In each tube the DNA is restricted by 1 µl (10 units) of either RsaI, AluI or Sau3AI in a total volume of 30 µl. The restrictions are performed in Tango restriction buffer.
- The tubes are incubated at 37°C for 30 minutes followed by inactivation at 65°C for 20 minutes. Finally the three restrictions are collected in one tube.
- 7 µl of the fragmentized DNA is added loading dye and loaded on a 1% agarose gel containing ethidium bromide. The gel is inspected to verify successful fragmentation.

Purification of fragmentized/amplified DNA with the Qiagen - MinElute Reaction Cleanup Kit, with the following modifications.

- After step 6, an additional wash is performed by the addition of 500 µl fresh 80% ethanol to the spin column followed by centrifugation as specified in the protocol.
- In step 9, the elution buffer (EB) is exchanged by MilliQ, pH 8. 10 µl MilliQ is added and the spin column is left at room temperature for 5 min before centrifugation. After centrifugation an additional 10 µl MilliQ is added followed by an additional incubation at room temperature for 5 min. The sample is collected in the same eppendorf tube.

The DNA quantity and quality is investigated using the NanoDrop ND-1000 UV-VIS spectrophotometer. Because the quality of the DNA is very important the A_{260}/A_{280} ratio should be 1.8-2.0 and the A_{260}/A_{230} should be >2.0.

DNA labeling and general micro array processing

The DNA labeling and general micro array processing (hybridization, wash and data extraction) is performed as described in the "Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis (ULS Labeling for Blood, Cells, Tissues or FFPE)" starting from page 29, with the following modifications.

- When preparing the 8 µl sample containing 250 ng fragmentized DNA on page 30 include 250 ng 101mer synthetically produced VHS DNA in the tubes.
- Preparation of labelling master mix on page 33 is performed for ULS-Cy3 using the 4x microarray protocol. The ULS-Cy3 master mix is added to the mixture of fragmentized chromosomal DNA and VHS DNA as described on page 34.

Project funded under the Food, Agriculture and Fisheries, and Biotechnology theme (KBBE)

SEVENTH FRAMEWORK PROGRAMME

- When preparing hybridization master mix for 8x microarray on page 40, exchange the Cot-1 DNA with nuclease free water.
- Step 7 and 8 on page 41 is changed to
    - Transfer sample tubes to a circulating water bath or heat block at 95°C. Incubate at 95°C for 10 minutes.
    - Immediately transfer sample tubes to ice. Incubate for 10 min.
- The hybridization is performed at 52°C for 24 hours.
- The washing procedure is modified to:
- Disassembly in OligoaCGH Wash Buffer 1 at 37°C.
- 1st wash in OligoaCGH Wash Buffer 1 at 37°C for 5 minutes.
- 2nd wash in OligoaCGH Wash Buffer 2 at 37°C for 1 minute.
- 3rd wash in acetonitrile at room temperature for 1 minute.
- 4th wash in Stabilization and Drying Solution at room temperature for 30 seconds.
- The microarray slides were scanned with a Molecular Devices GenePix 4000B microarray scanner. The data was extracted with GenePix Pro 6.0 and analyzed with MatLab.

Data analysis

On each microarray slide, Hamming distance is used to measure the difference between each pair of samples and a distance matrix is constructed. Hierarchical Cluster Analysis (HCA) and Principle Coordinate Analysis (PCoA) is used to analyse the distance matrix. The dendrograms obtained by HCA and the scores plots obtained by PCoA, is examined to identify possible separations between different strains of bacteria or microbial communities.

The next step is to combine all microarrays together and perform PCoA on the combined data in a similar way as on each individual microarray. The scores plot is examined for continued separation between strains or communities.

To test the robustness of the separation shown in the scores plot, support vector machines is used to build a predictive model. The model is built by using a subset of samples as the training set and the remaining samples are used to test the accuracy of the prediction of the model. This procedure is repeated 10.000 times, each time with a different selection of training and test samples (based on boot raping methodology).

Project funded under the Food, Agriculture and Fisheries, and Biotechnology theme (KBBE)

SEVENTH FRAMEWORK PROGRAMME

RESULTS

Micro array analysis has been performed on pure cultures of a range of bacterial species, listed in Table 1. From most of the species the type strain, reference strains and isolates from either clinical cases or the food industry has been analysed.

The averaged overall prediction accuracy was 96% and most strains obtained 90-99% of prediction accuracy as listed in Table2**.** The low score of *Enterococcus faecium* is probably caused by only testing one strain in few replicates at this point. For comparison only one strain has been tested of *Bacillus cereus,* but in several replicates.

Table 1: Bacterial species analyzed using the Universal DNA micro array method.

Species	Likelihood of correct identification [%]
Bacillus cereus	95.8
Enterococcus faecalis	96.36
Enterococcus faecium	69.22
Listeria monocytogenes	99.81
Staphylococcus aureus	97.76
Staphylococcus pseudointermedius	90.24
Streptococcus agalactiae	99.26
Streptococcus equi	92.67
Streptococcus suis	92.84

For evaluation of the DNA array for discrimination of *Salmonella* Serovars, 4 strains of each of the 5 serovar for which there is a specific FSO in Europe (not present in laying hens) listed in Table 2 was used.

Table 2: *Salmonella* Serovars analyzed to investigate if discrimination of them is possible

"O"-group	Serovar	"O"-antigens	Phase 1 (motile) "H" antigens	Phase 2 (non-motile) "H" antigens
B	*S*. Typhimurium	1,4,5,12	i	1,2
C_1	*S*. Infantis	6,7	r	1,5
D	*S*. Enteritidis	1,9,12	g,m	no antigen
C_2	*S*. Hadar	6,8	z10	e,n,x
C_1	*S*. Virchow	6,7	r	1,2

Project funded under the Food, Agriculture and Fisheries, and Biotechnology theme (KBBE)

SEVENTH FRAMEWORK PROGRAMME

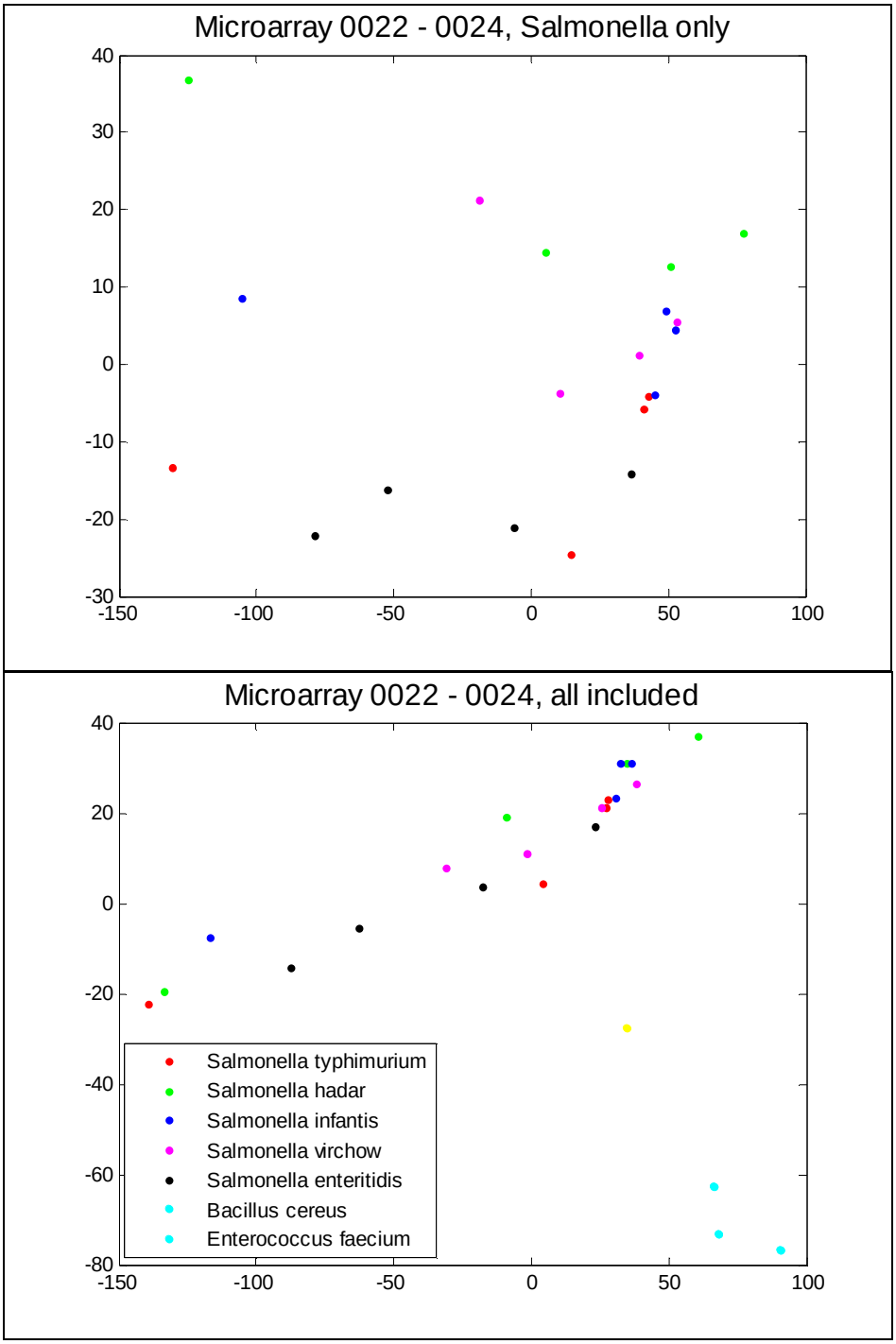

Fig 1. Identification of strains of *Salmonella* by universal DNA array. Signals from strains of *Salmonella* clustered distinctly in PCA analysis compared to control strains from two other species (bottom). However, signal from the individual serotypes could not be separated (top).

Project funded under the Food, Agriculture and Fisheries, and Biotechnology theme (KBBE)

SEVENTH FRAMEWORK PROGRAMME

Analysis of the results showed that *Salmonella* can be specifically identified, however the analysis unfortunately showed that it was not possible to discriminate between the Salmonella serovars (Fig 1).

Project funded under the Food, Agriculture and Fisheries, and Biotechnology theme (KBBE)

SEVENTH FRAMEWORK PROGRAMME

DISCUSSION

The DNA array platform developed can be used for species specific diagnosis in the same way as PCR and Malditof. In the set up used here it is more cumbersome than both these methods, however, each of the steps can be automated. The signal from all strains from each species, except *Enterococcus faecalis*, were clearly separated. In statistical terms, the similarity analysis gave confidence intervals between 92-99 %, however, no misdiagnosis on strain level was observed.

The purpose of the investigation in this project was to evaluate the possibility to specifically identify the five serovars of *Salmonella* that are mentioned in the FSO for laying hens. This analysis showed that *Salmonella,* like the other species previously tested, gave specific signals. However, the sensitivity of the assay was not sufficient to be able to separate the signals from the five serovars from each other. Genome sequences are not available for all the serovars, and it is not possible to control the result in silico. The result means that the sequence differences in the genomes corresponding to our random probes are not sufficiently big to overcome the inherent sensitivity level of the method caused by random noise.

At the mid-term review of the BASELINE project, the review committee judged this method to be less suitable for industrial use than PCR. Based on the disappointing results and this judgment by the review committee, no further work has been performed in BASELINE on this method.

Project funded under the Food, Agriculture and Fisheries, and Biotechnology theme (KBBE)

SEVENTH FRAMEWORK PROGRAMME