Aquavalens Project

"Protecting the health of Europeans by improving methods for the detection of pathogens in drinking water and water used in food preparation."

Grant agreement number: 311846

Deliverable D5.2

Data set on the distribution of different chemical and microbial MST parameters in surface waters with known faecal pollution origin and different geographic regions

Author: Andreas Tiehm, Claudia Stange

TZW, Germany.

Table of Contents

Andreas Tiehm
Claudia Stange

1. Literature critical review and selection of indicators

A literature revision and critical review to support the selected parameters (task 5.1) was performed (April to September 2013). It was supporting the decision on the microbial and chemical parameters to be measured although most of them had already been identified in previous studies.

Several criteria for selection of MST indicators among those proposed were determined:

1. Standardized or established (robust) methods to measure indicators already available

2. Previous knowledge on their host-specificity, prevalence in environmental waters, seasonality and environmental die-off of the indicator

3. Appropriate environmental detection limits of methods beyond the analytical sensitivity of methods.

4. Cost / effectivity and easy performance for routine analyses.

These activities allowed defining and writing a booklet of methodologies and protocols for internal use of participants (deliverable 5.1). This booklet was used for the sampling campaign that would allow the development of the data set on the distribution of different chemical and microbial MST parameters in surface waters with known faecal pollution origin and different geographic regions.

The selected parameters to be analysed were:

- *E.coli* (EC)
- Faecal Enterococci (microbial and molecular detection, FE and FEqPCR)
- *Clostridium perfringens* (CP)
- Somatic coliphages (SomPhg)
- Total bifidobacteria (BifTotal) and sorbitol fermenting bifidobacteria (BifSorb).
- *Bacteroides* host-specific phages (human HMBactPhg), porcine (PGBactPhg), bovine (CWBactPhg), poultry (PLBactPhg).
- Human bifidobacteria (HMBif), bovine bifidobacteria (CWBif), porcine *Neoscardovia* (PGBNeo), poultry bifidobacteria (PLBif).
- Total *Bacteroidetes* (AllBac), human *Bacteroidetes* (HF183TaqMan) and ruminant *Bacteroidetes* (BacR), pig *Bacteroidetes* (Pig-2-Bac).
- Human mitochondrial (HMMit), bovine mitochondrial (PGMit), porcine mitochondrial (PGMit) and poultry mitochondrial (PLMit)
- Adenoviruses (Adeno)
- Noroviruses (NoV)
- Acesulfame
- Cyclamate
- Saccharin
- Sucralose

2. Implementation of methods and setting operating principles

Partners agreed the use of international standard (ISO, CEN) protocols when available. Other protocols of new indicators were written up and joined with those standardized in a booklet of standard operating procedures (SOP) for the internal use. Partners agreed to perform the quality control of results by performing a verification test with blind samples instead of doing a two weeks training session. The verification test for traditional microbial parameters was performed among the participants in October 2013. Two raw urban sewage (high and low faecal concentration) were sent (October 21st) by UB to all partners. The participant partners analysed these "blind samples" at the same day (October 28th) for *E. coli*, enterococci, *C. perfringens*, somatic coliphages and total and fermenting-sorbitol bifidobacteria, following the agreed SOP. Specific instructions were sent by UB to the partners how to proceed. Results (enumerations) were sent to UB for statistical comparison of results. No significant differences were observed between the results of partners, though some higher variance was observed for certain parameters.

3. Measuring faecal load in polluted waters (point source) for the development of a data set

This task constituted the first sampling campaign for the participants in order to obtain the data set on the distribution of different chemical and microbial MST parameters in surface waters with known faecal pollution origin and different geographic regions. This data set is the training set of results to be later used on inductive learning machine methods (task 5.4).

Samples have been taken by partners from October 2013 to September 2014, using the established SOP and following the next criteria:

Criteria and activities:

1. Samples of wastewater (point source) of known faecal origin. If collective (population) faeces were taken as substitutive for wastewater, at least 25 individual faeces were integrated in one sample. In case this number of individual faeces was not constituting the integrated sample, it should be registered and communicate to the partners.
2. Samples representative of populations.
3. Samples were identified by codification as agreed:

    a. 1 character for partner code related to respective country: E for UB, A for TU WIEN, D for DVGW, P for IST, F for UH

    b. 2 character for number of sample: 01, 24, 56, …

c. 2 characters for faecal source: HM for human, PG for pigs, CW for cow, PL for poultry. Any other origin was indicated to the partner and codification was communicated.

d. "_" plus 1 character for season: W for winter, P for spring, S for summer F for Fall

Example: P37PG_S

4. A total of 25 samples by participant being at least 5 samples from each faecal origin. A total of 125 samples constituting the training matrix for modelling. Each partner distributed sampling covering the four seasons.

5. Each partner measured locally the following parameters in their respective samples: *E. coli*, enterococci, *C. perfringens*, somatic coliphages and total and fermenting-sorbitol bifidobacteria.

6. Each partner delivered one aliquot of 50 ml (for microbiological or eukaryotic analyses) and one aliquot of 100 ml (for chemical analyses) of each fresh and recently taken sample to each the partners during the 24 h of sampling. Additionally one more aliquot of 50 ml was send to UB for somatic coliphages and host-specific *Bacteroides* bacteriophages analyses. Samples were shipped at 4°C with private courier or any other 24 h delivery service. Recipient laboratories preserved or analysed the samples right after their arrival at their laboratories.

7. Each partner was responsible of analysing at least one of the MST host-specific indicators. Distribution of molecular targets:

a. qPCR specific bifidobacteria plus total bifidobacteria: UB

b. *Bacteroides* host-specific phages: UB

c. qPCR specific *Bacteroidetes* plus total *Bacteroidetes*: TU WIEN

d. Chemical compounds (Acesulfame, Cyclamate, Saccharin, Sucralose): DVGW (TZW)

e. qPCR mitochondrial: IST

f. qPCR adenoviruses: IST

g. qPCR norovirus: UH

h. HF183 (US-EPA last protocol): TU WIEN

8. The die-off regression in the environment and the detection limits for each measured MST indicator was provided by responsible partner (based on experimental assays or in literature when not available from own experimental assays).

9. A frame (Excel file) was provided by UB to introduce results along the analyses on Task 5.3. Results were sent to UB, where all results were summarized. It was essential to avoid missing values – otherwise the whole sample will be omitted for training the inductive learning methods.

4. Results and discussion

A sampling calendar was established for delivery of samples by each partner to the rest of participants. Each partner communicated the estimated date for delivering of samples.

After revising and harmonizing the results, a total of 118 out of 125 samples constituted finally the data set. They were distributed as indicated:

	Human	Porcine	Bovine	Poultry	Horse	Others	TOTAL
UB	6	9	5	6			**26**
TU WIEN	8	4	4	2	2	4 (bird, goose, cat, dog)	**24**
IST	7	5	5	6		2 (cat, dog)	**25**
DVGW (TZW)	8	4	5	5	5	1 (rabbit)	**25**
UH	7	2	4	5			**18**
TOTAL	**33**	**24**	**23**	**24**	**7**	**7**	**118**

An additional parameter (faecal enterococci by qPCR, FEqPC) was analysed with the rest of planned parameters. The final overview of measured indicators is shown below:

Faecal indicators	
E. coli Enterococci Total bifidobacteria Total *Bacteroides* *Clostridia* Somatic phages	8

MST				
	Culture		**Molecular**	
Bacteria	Sorbitol fermenting bifidobacteria	1	Host-specific bifidobacteria Host-specific *Bacteroides*	7
Virus	*Bacteroides* phages	4	Adenovirus Norovirus	2
Eukaryotic			Mitochondrial DNA	4
Chemical	Acesulfam, Cyclamate, Saccharin, Sucralose			4

At present, the developed data set is for internal use of the consortium under terms of confidentiality. It is being used for the numerical analyses planned on Task 5.4 for the selection of subsets of best

parameters to distinguish main faecal sources in water (human, bovine, porcine and poultry). This data set is planned to be used as training set for the machine learning analyses in order to develop predictive models with highest accuracy based on the most robust parameters when effects of environmental dilution and aging are considered.