# C:\Users\la40019\Desktop\HAP-E logo_JohnBrown.jpghttp://connect.hutton.ac.uk/Organisation/Comms/Scottish%20Government%20Logo/SG_Dual_linear_CMYK.jpghttp://connect.hutton.ac.uk/Organisation/Comms/PublishingImages/logos/JHI_STRAP_Print.png

# Review of approaches for microbial source tracking (MST) in waters.

### Report for the Scottish Government: RESAS Theme 1 RD 1.2.1

### Eulyn Pagaling and Lisa Avery - May 2017

# Summary:

# Microbial source tracking (MST) techniques are intended to allow practitioners and researchers to identify the source or sources of faecal pollution in waters, however there are no defined “standard” methods for MST in the UK. This document reviews the current best available microbial source tracking approaches using a rapid literature review and a survey of expert opinion. This will inform the application of source tracking methods to DNA archived from the National Waters Inventory of Scotland (NWIS)1. There was no clear consensus on the “best” methods for source tracking among experts and practitioners contacted and a variety of methods were discussed in the literature, with no clear “best” approach. The importance of using multiple markers was highlighted several times and a pragmatic and practical approach appeared to be the adoption of a “toolbox” of different source tracking markers to be applied in conjunction with some knowledge of the catchment characteristics. It was decided that this approach would be applied to the NWIS samples with evaluation of the findings after application of generic, ruminant specific and human specific markers.

### Background:

### Microbial source tracking (MST) techniques are intended to allow practitioners and researchers to identify the source or sources of faecal pollution in surface waters, ground waters or drinking waters. There are no standard methods for MST in the UK, although some MST approaches are already in use by regulators such as SEPA. MST techniques can be applied to fresh or archived samples, depending on the selected method and how the samples have been archived. The James Hutton Institute generated a DNA archive from its Scotland-wide “National Waters Inventory” survey. This is described in a previous report1.

Aim:

The aims of this review was primarily to provide a general overview of the current best available microbial source tracking approaches and to therefore identify a suitable method for microbial source tracking to apply to the NWIS DNA archive1 (see RESAS WP 1.2.1 O2.2 D D2.2i).

Methodology:

A two-fold approach was applied to carry out this review, comprising an evaluation of available literature and consultation with relevant experts/practitioners in the field.

Literature review:

First of all, a high level scan of the literature was completed and information from a selection of key papers was summarised. There is a substantial body of literature on microbial source tracking approaches including several recent review papers. To avoid replicating existing reviews, we utilised these and some of the key papers in the literature to identify current practical approaches for source tracking. We aimed to identify tests with good specificity (true negative rate) and sensitivity (true positive rate) which would be practical to implement. The focus was primarily (although not exclusively) on library-independent methods which could be applied direct to environmental DNA.

Expert Opinion Survey:

Scientists or stakeholder staff with an interest in/expertise in microbial source tracking/source apportionment from the following 12 institutions were consulted by email in reference to the above query: James Hutton Institute, SEPA, Scottish Water, CREH, University of Brighton, University of Barcelona, University of Stirling, University of Chicago, Water Technology Centre (TZW), Agrisearch New Zealand, ESR NZ, Argonne National Laboratory (USA).

The questions posed were as follows:

* What do you consider to be the current best practical approaches for source tracking (microbial or otherwise) of faecal pollution?
* Could you give some reasons for the answer above
* What do you think are the most appropriate ways of applying the above techniques?
* Are there any up-and-coming techniques that you think show promise? If so what are they and why are they not currently in use?

### Results:

### Findings from the literature review are summarised in Table 1. There was a 75 % response rate in the survey of expert opinion with 58 % providing information. Answers are summarized in Table 2.

|  |  |
| --- | --- |
| **Source Literature** | **Summary** |
| 2Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes’ Harwood et al (2014) FEMS Microbiol Rev 1-40. | Focus of review: Q-PCR approaches.**Human-associated *Bacteroidales***: End-point PCR: HF183F forward primer designed against human-associated *Bacteroides*, paired with general reverse primer (Bac708R). This was followed by a nested PCR. Considered the best in terms of sensitivity and specificity. Limit of detection (LOD) was 10-12 g DNA or 105 gene copies. The HF183F was a pre-cursor to all the QPCR tests that were developed thereafter.**16S rRNA gene of Bacteroidales**: Many tests. They all amplify the 600bp region of 16S rRNA gene of *B. dorei*. They all vary in sensitivity and specificity. Some produce false positives due to cross-reactivity amongst species. None of the tests available (in this paper) appear reliable.**Other *Bacteroidales* gene targets**: TaqMan assay for *B. thetaiotomicron* α-mannase was 100% sensitive and 100% specific when tested against sewage and faeces. LOD is 9.3 gene copies. TaqMan assay for hypothetical protein HumM2 was 100% sensitive, but showed cross-reactivity. TaqMan assay for *B. fragilis* *gryB* was 100% sensitive, but also showed cross-reactivity.***Bifidobacterium***: Assays developed to test for *B. adolescentis*, with the aim of tracking human faecal pollution. However, subsequent analysis showed that *B. adolescentis* is not confined to the human gut. TaqMan assay (ADO) showed 90% sensitivity for faeces and 100% sensitivity for sewage. It showed 94.5% specificity, generating false positives from cow and bird samples. Possibility of using this assay as a general screen for fecal pollution if *B. adolescentis* is detected in a range of hosts. ***Enterococcus faecium* *esp* gene**: specific for human. SYBR Green QPCR assay developed with forward primer specific for *Ent. faecium* and reverse primer that targeted both *Ent. faecium* and *Ent. faecalis*. Specificity was high, but sensitivity was not. Assays detecting HF183 are more sensitive.***Methanobrevibacter smithii***: specific for human. Assay developed to target the *nifH* gene. Highly specific, but its low sensitivity does not allow the assay to detect fecal pollution. ***Lachnospiraceae***: Second most abundant bacterial group in human faeces, therefore is human specific. No sensitivity or specificity testing, but was found to correlate well with human pathogen presence.***E. coli***: mostly used in library-dependent methods i.e. antibiotic resistance, ribotyping or DNA fingerprinting. No QPCR assays available, but on-going work will probably identify human-specific biomarkers.**F+ RNA coliphage genotyping**: Multiplex QPCR assays developed to target RNA coliphages, which are associated with faeces. No sensitivity or specificity testing has been done. The designated coliphage groups associated with human and animal waste does not seem to be correct i.e. human coliphages have been detected in animal waste, therefore it is not an appropriate test to detect hosts, but could be used as a general screen for fecal pollution.**Pepper mild mottle virus**: Plant pathogen that is a major component of the metagenome in human faeces, therefore human specific assay. Shows good persistence through wastewater treatment, but does not survive well in environmental waters, therefore, this would not be an appropriate test for water-related source tracking (e.g. NWIS).**HPyVs**: specific to human and highly prevalent. All assays developed have only been tested on sewage, faeces and urine; it has not been tested on field samples. Therefore, this is not a good choice for NWIS.**Human mitochondria**: Multiplex QPCR developed (Caldwell et al 2007 – see below), but due to false-positives from mixed sources, adequate DNA concentrations are required, so this may not perform well for environmental waters. |
| *3‘Performance of forty-one microbial source tracking methods: a twenty-seven lab evaluation study’* Boehm et al (2013) Wat Res 6812-6828. | * Best quantitative methods were HF183Taqman1 and BacH2 (human), Rum2Bac and BacR (ruminant), LeeSeaGull (gull) and Pig2Bac (pig).
* Blind test was done where methods were tested on sewage, septage and fecal samples. Some faecal samples were tested as doubletons as well as singletons.
* However, some of these tests were only tested in one lab, so further field testing may be required.
 |
| *4‘Molecular indicators used in the development of predictive models for microbial source tracking’* Balleste et al (2010) Appl Environ Microbiol 1789-1795. | * Tested a range of biomarkers for MST, but decided that no individual test enabled 100% source identification. Instead, a ‘tool box’ approach was recommended to raise sensitivity and specificity.
* One tool box used a combination of ADO (*Bifidobacterium adolescentis* – to detect human, although there is cross-reactivity as noted by Harwood *et al.* above), Bomito (cattle) and Pomito (pigs). This enabled correct identification of 75.7% of samples, differentiating between human, swine, bovine and poultry.
* The more biomarkers in the tool box, the better the test. The tool box containing 5 biomarkers of HF134 (*Bacteroides*), ADO (human, with cross-reactivity), DEN (*Bifidobacterium dentium* - human), Bomito (cattle) and Pomito (pigs) provided 90.1% correct classification.
 |
| 5‘*Mitochondrial multiplex real-time PCR as a source tracking method in fecal-contaminated effluents’* Caldwell et al (2007) Environ Sci Technol 3277-3283. | * Mulitplex QPCR for mitochondrial DNA (mtDNA) developed to detect human, bovine and swine faeces, with an internal amplification control (IAC).
* Assay was tested on effluent, reporting no false positives or cross-reactions. However, the test was not carried out on field samples where concentrations of mitochondria are likely to be substantially less than in effluent.
* Assay had a detection limit of 102 copies per reaction. This translated to 0.2 g faeces per 100 ml of effluent.
 |
| 6‘*Development of new host-specific Bacteroides qPCRs for the identification of fecal contamination sources in water’* Gomez-Donate et al (2016) Microbiology 5:83-94 | * TaqMan PCR assay developed to target the 16S gene in *Bacteroides* to discriminate between human, cattle, pig and poultry. They did this by using DGGE to identify host-specific bands as potential targets.
* Tested on faecal slurries.
* The cattle assay did not work, but they claim that the assay for human, pig and poultry outperform all other assays to date, including HF183 and Pig2Bac.
* Despite high specificities (97-100%), they had low sensitivities (45-73%)
 |
| 7‘*A novel microbial source tracking microarray for pathogen detection and fecal source identification in environmental systems*’ Li et al (2015) EST 49: 7319-7329 | * Microarray developed to detect pathogens, microbial source tracking and AMR genes.
* Tested against whole genome amplification (WGA) of DNA from sewage and animal faeces. Sensitivity was low (21-33%), but specificity was high (83-90%).
 |
| 8‘*Analysis of human and animal fecal microbiota for microbial source tracking*’ Lee et al (2011). ISME J 5:362-365 | * Used 454 pyrosequencing to identify bacterial species specific for human and animal faeces. Concluded that these bacterial species could be used as biomarkers for MST.
 |
| 9‘*Use of Barcoded Pyroseqeuencing and shared OTUs to determine sources of fecal bacteria in watersheds*’ Unno et al (2010) EST 44:7777. | * Pyrosequencing reads from faeces were compared to pyrosequencing reads from watercourses.
* Human and domesticated animal faeces predominantly contained operational taxonomic units (OTUs) belonging to *Bacteroides* and Firmicutes, whereas wild goose faeces predominantly contained OTUs belonging to the Proteobacteria.
* When compared to OTUs from watercourses, they identified that they were contaminated with human and swine faeces.
* Quantification of contamination was determined by density of the reads.
 |

### Table 1: Summary of key literature on microbial source tracking techniques

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Organisation type** | **Best practical approach to MST** | **Reasons** | **How to apply** | **Up and coming approaches** | **Sources of relevant information** |
| Research | Modelling based approaches |  | To identify source and potential pathways of faecal entry to watercourses |  | 10Crowther, J., Kay, D., Anthony, S., Gooday, R., Burgess C.,and Douglass, J. (2016), *CREW REPORT CRW2015\_1* |
| Research | Combination of different markers needed to make the best prediction is needed | One marker usually insufficient | The further from the source, the more markers may be needed – dilution, mixing and aging effects.  |  | 6Gomez-Donata *et al*. (2015) Microbiology Open . doi: 10.1002/mbo3.31311Jofre *et al.*  (2014). Water Research 55; 1-11. |
| Practitioner | FC:FS ratio – used over past 50 years | Crude clues especially if close to source | Initial indication - further from source different die-off rates affect ratios. |  |  |
| Practitioner | Q-PCR (“Allbact” generic Bacteroides primers then “HuBac” (human) and “BacR” (ruminant) markers | Analysis can be completed within the time FIOs are incubated and whether to run MST can be decided based on early indication from culture plates. | Many temporal samples needed at a given site to differentiate sources | Adding genomic markers for gulls, dogs, waterfowl and more human and ruminant sequences |  |
| Research | Probably species specific viruses (Rusinol) and/or the EA approach based on *Bacteriodes* (Stapleton) | Species specifc viruses - only approach seen tested quantitatively / good source apportionment. Bacteroides widely used worldwide but remains qualitative | Extreme caution; be very circumspect in assuming quantitative species contributions | Genomic profiling is being developed but remains a library dependent method to the best of my knowledge | 12Rusinol *et al.*(2014); Water Research 59; 119-129.13Stapleton *et al.* (2009). Water Research 43; 4888-4899. |
| Practitioner | Bacteriophage – but only speaking from experience of clean water tracking. | Sensitive and specific | Spike at point of source and sample along expected course of water or end point. |  |  |
| Research | Phage lysis of a human-specific gut bacterium |  |  | Risk assessments and predictive models based on microbe and environmental data within agro-hydrodynamic models |  |
| EA Starcross | Nucleic acids, routinely detected via PCR, (not necessarily qPCR. A lot could be taken from presence/absence if the sampling were considered further (e.g. similar approach to occupancy models). | The wide range of possible targets that can be looked for from the one DNA prep. | Use PCR in the lab, with thoughtful data analysis. Using the data – take more “other” measurements e.g. rainfall over catchment response-specific sampling periods, tide, turbidity, nitrogen species. Sampling regimes need more thought than they are sometimes given; generally need to compare baseline vs response/incident. FIO plate counts show huge variability when sampled repeatedly and these are the “correct” answers MST is expected to explain. | Chemical methods deserve more attention; One area of potential must be the possibility of determining the age of faecal pollution, which would give useful context for the rest of the data. |  |

### Table 2: Summary of “Expert Opinion” responses from Research and Stakeholder (Practitioner) organisations

### Discussion:

#### Literature Review:

#### The main focus of this review of approaches for microbial source tracking was methods suitable for application to archived DNA. Broadly, “library independent” methods of source tracking focus on DNA (or organisms) extracted direct from environmental samples with no requirement for a “library” of sequence or typing information from potential sources against which to compare. Library dependent methods typically require the generation of a culture collection of a particular source tracking organism with a suite of isolates from each potential source. Isolates cultured from the samples of interest are then compared against this library.

For archived DNA (such as NWIS) none of the library-dependent methods (e.g. typing of FIO isolates) can be employed as this involves a cultivation step. This limits the method of choice to library-independent methods that target a specific feature or species. The trend in recent years has been to use Real-time or Quantitative PCR (QPCR), presumably due to its relatively low cost, low labour intensity and relatively good time efficiency leading to same day results and can provide quantitative information. Terminal restriction fragment length polymporphism (TRFLP) and denaturing gradient gel electrophoresis (DGGE) – based approaches appear to have fallen out of favour as stand-alone MST methods; the last papers using these methods were published in early 2000. Currently, these methods are used to identify host-specific sequences so that novel QPCR assays can be developed (Gomez-Donate et al 2016). Microarrays have not been sufficiently tested on field samples (Li et al 2015) and are currently likely to be prohibitively expensive beyond research applications. Some have reported using pyrosequencing for MST (Lee et al 2011; Unno et al 2010). Particularly as it is not a currently widely developed, this approach would probably require a reference sample to be sequenced for comparison of operational taxonomic units (OTUs; Unno et al 2010).

Most MST methods involving PCR have been designed to be specific to certain hosts. Assays targeting human faeces are better developed than those targeting animals (farm, domestic or wild) because contamination with human faeces is regarded as a critical health issue. Several QPCR assays have been developed to target *Bifidobacterium*, *Enterococcus*, *Methanobrevibacter*, various phages and mitochondrial DNA. However, the majority of assays reported in the literature target different genes within *Bacteroides* genus (Harwood et al 2014 and references theirin). A common problem of all of these assays is that they rely on high concentrations of targets in faeces so that they can be detected once diluted in watercourses. In this respect *Bacteroides* and *Bifidobacterium* are good targets. The target also needs to have a long half-life in environmental waters. Bacteroides survive for a few days in water but it is thought that their DNA persists for longer (Gomex-Donate *et al.* 2016).

 Some recommend a ‘tool box’ approach because currently, no single assay is perfect (Harwood, et al 2014; Balleste et al 2010). A further recommendation is to choose assays according to the likely sources of contamination in the environment being tested (Harwood et al 2014). Therefore, there are two possible approaches:

1. Targeted approach based on land use data: Simultaneously applying a ‘tool box’ of assays selected according to land use data. For example, urban-impacted waters could be tested for human and domestic animals, upland waters could be tested for sheep and deer, waters near wetlands can be tested for wild birds, etc.
2. Initial screen followed by more in-depth analysis: Choose a general presence/absence assay such as a generic Bacteroides assay for an initial screen of archive DNA. This is based on the assumption that there is no point in trying to detect host-specific Bacteroides (and possibly other markers) if few or no Bacteroides are detected in the first place. Based on these results, select samples for quantitative host-specific assays

When the library independent MST assays of choice have become well-established, it should not be necessary to screen potential sources (e.g. host faeces). However, trialling chosen assays against host-specific faecal DNA may be a wise control while methods continue to be developed.

#### Expert Opinion Survey:

#### There was no clear consensus on the “best” methods for source tracking among experts and practitioners contacted. The importance of using multiple markers was highlighted several times and virus or phage-based approaches were also mentioned. Modelling aspects were also noted as important and it was suggested that the importance of appropriate sampling regimes and measurement of other catchment variables is often neglected.

Up-coming approaches with potential that were highlighted included the potential for determining the age of faecal pollution, the addition of molecular markers for additional species/groups to “toolbox” type approaches, genomic profiling, chemical methods and predictive modelling/risk assessment approaches based on environmental data.

### Conclusions:

### Based on current knowledge, it seems that a common-sense approach to application of source tracking information to archived DNA such as the NWIS archive is to take a “tool box” approach, applying a general screen (e.g. generic *Bacteroides*), followed by specific markers selected based on the catchment or general archive characteristics. For a mixed catchment, a suggested approach would be

**Other catchment-relevant markers selected from: gulls (e.g. LeeSeaGull), pigs (Pig2Bac; Pomito), cattle (Bomito)** **sheep, deer (tbc)**

**Ruminant markers**

**(e.g. Rum2Bac and BacR)**

**Generic Bacteroides e.g. Allbac**

**Human markers**

**(e.g. HF183; BacH)**

### Application to NWIS:

### For NWIS we will:

* Screen the archive using Generic Bacteroides “Allbac” primers
* Screen Allbac positive samples with Human and Ruminant Markers HF183 and BacH, RumBac and BacR.
* Review findings and decide whether to screen with additional markers.
* Assess pyrosequencing data already sequenced for microbial diversity/AMR) for source tracking markers (Sequence some source references i.e. faecal DNA from cow, sheep, human, duck, goose, chicken – which we have – and deer/other wild birds). Not a proven approach but interesting to compare with the other approaches we’re using.

### References:

1. Avery, L.M., Pagaling, E. and Wyness, A. (2017). ‘Exploring water DNA: our strategy for exploiting the National Waters Inventory for Scotland DNA archive.’ Technical report for the Scottish Government, delivering for RESAS RD 1.2.1.
2. Harwood, V.J., Staley, C., Badgley B.D., Borges, K., Korajkic, A. (2014). Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. FEMS Microbiol Rev 38: 1-40.
3. Boehm, A.B., Van De Werfhorst, L.C., Griffith, J.F., Holden, P.A., Jay, J.A., Shanks, O.C., Wang, D.and Weisberg S.B. (2013). Performance of forty-one microbial source tracking methods: a twenty-seven lab evaluation study. Wat Res 47: 6812-6828.
4. Balleste, E., Bonjoch, X., Belanche, L. and Blanch, A (2010). Molecular indicators used in the development of predictive models for microbial source tracking. Appl Environ Microbiol 76:1789-1795.
5. Caldwell, J.M., Raley, M.E. and Levine J.F. (2007). Mitochondrial multiplex real-time PCR as a source tracking method in fecal-contaminated effluents. Environ Sci Technol 41: 3277-3283.
6. Gomez-Donate, M., Casanovas-Massana, A., Muniesa and M., Blanch, A.R. (2016). Development of new host-specific Bacteroides qPCRs for the identification of fecal contamination sources in water. Microbiology 5:83-94.
7. Li, X., Harwood, V.J., Nayak, B., Staley, C., Sadowsky, M.J. and Weidhaas, J. (2015) A novel microbial source tracking microarray for pathogen detection and fecal source identification in environmental systems. Environ Sci Technol 49: 7319-7329.
8. Lee, J.E., Lee, S., Sung, J. and Ko, G. (2011). Analysis of human and animal fecal microbiota for microbial source tracking. ISME J 5:362-365.
9. Unno, T., Jang, J. Han, D., Kim, J.H., Sadowsky, M.J., Kim, O.S., Chun, J. and Hur, H.G. (2010). Use of Barcoded Pyroseqeuencing and shared OTUs to determine sources of fecal bacteria in watersheds. Environ Sci Technol 44: 7777-82.
10. Crowther, J., Kay, D., Anthony, S., Gooday, R., Burgess C. and Douglass, J. (2016), *CREW REPORT CRW2015\_1*
11. Jofre, J., Blanch, A., Lucena, F. and Muniesa, M. (2014). Bacteriophages infecting Bacteroides as a marker for microbial source tracking. Water Research 55; 1-11.
12. Rusinol,. M., Fernandez-Cassi, X., Hundesa, A., Vieira, C., Kern, A., Eriksson, I., Ziros, P., Kay ,D., Miagostovich, M., Vargha, M., Allard, A., Vantarakis, A., Wyn-Jones, P., Bofill-Mas, S. and Girones, R. (2014). Water Research 59; 119-129.
13. Stapleton, C.M., Kay, D., Wyer, M.D., Davies, C., Watkins, J., Kay, C., McDonald, A.T., Porter, J., Gawler, A. (2009)Evaluating the operational utility of a bacteroidales quantitative PCR-based MST approach in determining the source of faecal indicator organisms at a UK bathing water. Wat Res 43: *4888-4899.*