

**Dr Lisa Avery, Dr Eulyn Pagaling, Dr Adam Wyness March 2017**

**Exploring water DNA: Our strategy for exploiting the National Waters Inventory for Scotland DNA archive**

**Summary:**

Building on existing SEPA sampling from primarily end-of catchment sites across Scotland, the James Hutton Institute added value through the application of a range of additional and novel analyses. One key aspect of the work including enumeration of faecal indicator organisms (FIOs) which are not normally measured on these samples, and of particular novelty, extraction and archiving of DNA for subsequent interrogation by molecular techniques. DNA quantity did not correlate with FIO counts. This is not particularly surprising as DNA quantity reflects microfauna from healthy environments as well as more polluted waters. Archived DNA quality and quantity have been evaluated and the majority of samples are deemed appropriate to carry out the following initial analyses, in order of priority: Microbial source tracking with initial analyses based on Bacteroidales; detection of selected antimicrobial resistance genes, next generation sequencing of the 16-S gene to explore microbial diversity and finally detection of key bacterial pathogens likely to include toxigenic *E. coli*, *Campylobacter*, *Salmonella*, Clostridia and *Mycobacterium avium paratuberculosis* (causative agent of Johne’s disease).

**A brief introduction to NWIS:**

National waters inventory Scotland provides a national baseline study of the state of Scotland’s water resource which is required to understand resilience to the accumulating catchment pressures of delivering national objectives for food and renewable energy production, water supply for people, ecology and industry. Scientists from the James Hutton Institute built upon existing SEPA freshwater sampling and analysis, combining powerful analytical tools to build a picture of the state of our water resource. We have corresponding catchment and land use data and analyses focused on basic to advanced water physico-chemistry and microbiology.

**Microbiological and Molecular Biological aspects of NWIS**

A subset of NWIS samples were analysed for faecal indicator organisms (FIOs - total coliforms and *E. coli*) using the IDEXX colilert MPN method. Samples were analysed as soon as possible on receipt at the James Hutton Institute, however usually there were significant delays in the order of several days. This was unavoidable given the nature of the sampling programme, therefore some practical work was carried out to allow us to predict the original numbers of FIOs from those measured several days later at the time point when samples were analysed. Broadly, there was little change over ~ 4 days of sample storage but notable declines beyond this period. Modelled data as well as raw data are retained within our NWIS databases.

DNA was also extracted from a subset of NWIS samples (all of which had already been analysed for FIOs). To date, 207 samples have been extracted. These samples were selected to fall within a storage time cut-off of ~ 4 days and where possible from sites that a number of samples to give temporal variability. The chosen sites provide a good coverage of different land uses, although they do not provide particularly good geographical coverage of the whole country (due to the fact that more southern sites generally had a greater storage time and were therefore excluded). The extracted DNA forms the “NWIS DNA archive”. Raw water samples have not been discarded to date, however and remain in frozen storage. For the archived samples, DNA quality was determined by the NanoDrop 1000 (Thermo Fisher Scientific Inc., UK) and quantified using the Qubit dsDNA High Sensitivity kit (Invitrogen, UK). DNA was eluted in 100 µl. A small volume was required for quantification and quality analysis and a small volume inevitably irrecoverable. DNA concentrations have not been normalised therefore this would likely be required prior to any analysis. The remainder has been archived and for most samples, 4 x 20 µl aliquots are currently stored at -80˚C.

DNA degrades over a relatively short time period even when frozen, therefore it is important to utilise these samples as soon as possible (ideally within the early part of the current RESAS programme). However, they are a limited resource and as such decisions on how to maximise their usefulness form a key early deliverable for WP 1.2.1 (RESAS SRP 2016-21).

A preliminary analysis of the DNA quality and quantity data yielded the following summary information:

DNA quantity: The lower end of the concentration range is < 1 ng/ µl. Nine samples fall into this category and are potentially too low a DNA concentration to yield a PCR product i.e. may be unsuitable for further analysis. The remaining 198 samples can be taken forward for further analysis but we may experience difficulties in amplifying particular gene products in low end concentrations. We note that 10 ng DNA is commonly required for Q-PCR analyses. This would typically be achieved by using 5 µl of a concentration of 2 ng/ µl. We note that 17 of the samples in the archive have less than 2 ng/ µl and are likely to be problematic for Q-PCR and therefore may be excluded from AMR, source tracking and pathogen analyses. Whether or not these samples can be used for sequencing depends on the platform used as some require a minimum of 5 ng DNA.

DNA quality: This was determined based on DNA sample absorbance ratios at 260/280 nm and 260/230 nm. A 260/280 ratio of ~1.8 is generally accepted as “pure” for DNA and a lower ratio commonly indicates the presence of contaminants. The 230/280 ratio is a secondary measure of purity, commonly around 2, but this is influenced by pH. Environmental samples typically have a much wider range of ratios. Over 60% of the extracted samples had a 260/280 ratio between 1.5 and 2, indicating good (environmental) DNA quality. The ratios of 230/280 were generally very low. This indicates presence of humic acids, proteins and polysaccharides as contaminants which are common in environmental samples. A common approach is to “dilute out” these contaminants when performing PCR amplification. Visualisation of ethidium bromide-stained DNA run on agarose gels was carried out for all samples. Where concentrations were high enough to visualise (EtBr is not particularly sensitive), there was usually a degree of smearing which can indicate degradation of DNA (some is expected due to sample freezing) but most showed a substantial amount of high molecular weight DNA which indicates sufficient intact DNA for subsequent testing.

 1 2 3 4 5 6 7 8 9 10

**An example gel from NWIS DNA archive. Strong bands at the top of lanes 2-9 indicate high MW DNA.**

There was no significant relationship between either *E. coli* or coliforms with DNA quantity. This is perhaps not surprising, since there will be some trade-off between inputs of DNA from sources of pollution (e.g. farm run-off, wastewaters, soil erosion) and healthy populations of indigenous micro (and macro) flora in less polluted waters.

**Proposed Analyses:**

The proposed analyses are as follows, in order of priority:

1. Microbial source tracking test based on Bacteroidales (as per review of “best practice” source tracking methodologies (D2.2i) (Generic Bacteroides Q-PCRs run on all extracted samples demonstrate a generally high rate of detection).
2. Detection of antimicrobial resistance genes. This is particularly topical and we anticipate that AMR genes, because they are present across a wide range of bacterial species, are likely to be sufficiently abundant for detection by Q-PCR provided we are not attempting to detect rare AMR genes. This work links with detection and modelling of AMR in the environment in WP 2.2.6 and 3.1.3 (where work focusses more on soils and organic amendments) and to ensure continuity, the final selection of AMR genes to detect and Q-PCR methodology will be defined by methods developed in these work packages (by Karen Scott, RINH). However, the intention is to select representatives of two key classes of AMR gene such as B lactams (primarily human usage) and Tetracyclines (primarily animal usage).
3. 16S rRNA gene “shallow” Next Generation Sequencing. This will allow us to evaluate microbial diversity across the archive samples, which can provide important ecological and functional information but critically also underpins the AMR detection work above. AMR are present in pathogens and non-pathogens and this sequencing approach, while likely not to account for rare species, would provide information relating to the presence of particular groups and species which provides excellent background information against which to understand AMR prevalence. All extracted samples have been prepared and submitted for sequencing.
4. Key pathogens. If sufficient DNA is leftover after the previous analyses, we will select samples containing medium to high numbers of FIOs to screen for pathogens. This is justified in that where FIO loads are low, few pathogens are likely to be present and are likely to be more difficult to detect. Seventy-six of the extracted samples had >500 *E. coli*/100 ml (which, for context, is the bathing water standard for inland waters; Bathing Water Regulations, 2013). Forty-one extracted samples were >1000 *E. coli* /100 ml (limit of “good” standard for inland bathing waters). If medium to high FIO samples yield useful data we will analyse the 76 >500 *E. coli* samples for pathogens.

We plan to focus on a range of pathogens of differing environmental survival characteristics and providing a range of human and animal health issues. These include: toxigenic *E. coli* (stx genes – pathogenic *E. coli* demonstrated to behave differently to non-pathogenic strains; complements work in 2.2.6/3.1.3); *Clostridium* species (range of human and animal diseases; *C. perfringens* also used as an FIO; limited data on Clostridia in the environment, spore formers therefore robust survival); *Campylobacter* and *Salmonella* as key zoonotic pathogens with environmental transmission routes (complements work in WP 2.2.6 and 3.1.3); *Mycobacterium avium paratuberculosis* (causes Johne’s disease in cattle; complements WP 2.2.3 detection of MAP in soils). Sampling approach was not suited to Protozoa, however there would be potential to detect viruses. Norovirus and HepA which are pertinent to wastewater sources are RNA viruses and cannot be detected within a DNA based archive. Furthermore, viral DNA may degrade more readily outside a host, therefore we consider that the most effective use of the DNA archive for pathogen detection will be based around bacterial pathogens.

The overarching research questions to be addressed through the analysis of the NWIS DNA archive include:

1. how land use characteristics and other water characteristics influence i) the spatial and (to a degree) temporal distribution of AMR genes and pathogens; ii) microbial diversity
2. Whether molecular source apportionment approaches reflect land use characteristics of source catchments
3. How source apportionment findings relate to pathogen and AMR prevalence and microbial diversity
4. How standard FIO measurements relate to land use, pathogens, AMR genes, microbial diversity and molecular source tracking methods.