National Laboratory Service

Science Service Solutions

Investigation into the bacterial communities in bathing waters and associated sites around Scarborough South Bay, 2016.

Report produced for the Integrated Environment Planning Team, Yorkshire Area

Jonathan Porter,
National Laboratory Service,
Staplake Mount,
Starcross,
Devon.
EX6 8PE.

October 2017.

Investigation into the bacterial communities in bathing waters and associated sites around Scarborough South Bay, 2016.

Summary.

Microbiological water quality at Scarborough South Bay is cause for concern. Data over the last 4 years suggest water quality is variable over time, with occasional peaks in microbiological contamination. Multiple possible sources of bacteria could affect the water quality.

The aims of this study were to determine:
i. Is an industrial effluent a factor in the intestinal enterococci (IE) and *Escherichia coli* counts at the designated sampling point in South Bay?
ii. If so, how much of an impact does this industrial effluent have?
iii. Are there other factors, and how much do they contribute?

Samples collected around the area were analysed using standard bacterial counts to measure the levels of microbiological pollution, microbial source tracking (MST) to determine the original host animal group of the microbiological pollution, next-generation sequencing (NGS) to characterise the bacterial communities within the various samples, and a form of DNA-fingerprinting (rep-PCR) to compare IE bacteria isolated from the samples.

The results suggest:

1. The MST data gives strong evidence that pollution from sources other than the industrial effluent are present in South Bay.
2. The NGS data gives strong evidence that the bacterial community from the industrial effluent has an influence in South Bay occasionally.
3. The rep-PCR data suggest IE isolates most similar to that of the WwTW are the most widely distributed across all the sites sampled in this study. IE isolates most similar to those from seabirds and those from the industrial effluent are less widely distributed. Although not proven, one straightforward explanation for this is that the movement of water between the sites sampled is typically north to south, with less movement south to north. This matches observations from the NGS data and does not contradict the MST data.
4. The other pollution sources in South Bay are human, seabird and dog. Typically, a combination of these sources contribute to the total pollution.

Introduction and rationale.

There is an intrinsic relationship between the microbiological quality of water and the associated risk to health for recreational activities such as bathing in that water. Established methods for assessing the microbiological quality of bathing waters rely on the measurement of concentrations of the bacterial species *Escherichia coli* and the bacterial group, the intestinal enterococci (IE). These two bacterial types are often grouped together as faecal indicator organisms (FIO). Monitoring bathing waters for FIO according to the procedures set out in the bathing water directive enable regulation and management.

The epidemiological studies that underpin the bathing water directive use a dose-response relationship between the concentrations of the FIO and reported rates of human illness. The relationship is conceptually very straightforward; the higher the concentration of the faecal bacteria, the more likely the water is to contain pathogens, and the greater the risk of a bather becoming ill. A key assumption of the relationship is that the bacteria being assessed are of faecal origin. This assumption is generally accepted, although the possibility of environmental reservoirs of both *E. coli* and IE has received attention (Whitman *et al*., 2003, Kinzelman *et al*., 2004).

The epidemiological assumptions used in bathing water assessment could be questioned if it were demonstrated that a non-faecal source of either (or both) of the FIO contributed at least some of the bacteria found in bathing water samples. This scenario is likely to be restricted to specific sites, where local factors may be suspected of confounding the overall understanding of a bathing water. The line of thought could follow the line of: FIO that are non-faecal in origin do not indicate the presence of viral or bacterial pathogens such as norovirus (genogroups GI and GII) or *Salmonella* species. Should the non-faecal indicator bacteria contribute to the overall FIO in a water body, it is possible that the risk of infection to a bather could be over-estimated. Regardless of any flaws within this chain, the bathing water directive does not contain any clauses that would allow a reconsideration of the data to allow for any such over-estimation of risk. However, an understanding of the sources of bacterial pollution impacting a bathing water would help water quality management planning and provision of information.

Although historically well established as indicators of pathogens, developments in microbiological monitoring have begun to move past the measurement of FIO concentrations to try and gain more information on water quality. An example of this is the use of microbial source tracking (MST). Most *E. coli* isolated in the laboratory are identical in appearance, regardless of the original host animal species - that is, an *E. coli* isolate from a cow looks the same as an isolate from a sheep, or from a human. MST attempts to extract further information from a sample to help identify the sources of the faecal pollution. A variety of different approaches have been proposed and tested for MST, with varying degrees of success. Several review articles are available for further detail on the scope and performance of the most common applications (Stewart *et al.*, 2013; Harwood *et al.*, 2014). Target species used for MST have included the actual FIO, but have in general used different species or organisms to try and obtain source-tracking information. Information from MST data is becoming more widely established and adopted for environmental monitoring, though it is not used directly for regulation.

The monitoring programme and associated analytical laboratory work used to collect data to fulfil the requirements of the bathing water directive stipulates the use of FIO and the methods used to assess their concentrations. These methods use culturable bacteria; that is, bacteria that are able to grow under defined laboratory conditions. However, the majority of bacteria found in the majority of environments have proven difficult to impossible to culture under standard laboratory conditions. (This includes many gut and faecal bacteria, including those used for MST.) Assessment of non-culturable bacteria requires very different laboratory approaches, often using molecular biological methods to examine the nucleic acids found in the non-culturable bacteria.

In order to detect, quantify and compare two bacterial species when neither can be cultured in the laboratory, it is standard practice to use a gene known to be present in both bacterial species. Several genes involved in essential biological functions have been used to examine non-culturable bacteria, with the most commonly used gene(s) being those involved with the steps of taking the information encoded in DNA to produce a biologically active protein; the ribosomal RNA genes. One such gene, the small subunit RNA gene (commonly called the 16S rRNA gene or 16S rDNA) has been used extensively in studies of non-culturable bacteria, and indeed, ribosomal RNA gene sequences now form the entire base of the classification of all living organisms (the so-called Tree of Life). Ribosomal RNA gene sequences have some key features which help greatly in such studies; the genes will be present in some form in all cellular life forms; there are highly conserved sections within the gene (allowing definitive detection); there are variable sections between the conserved sections (providing species-specific information);

Bacterial ribosomal RNA gene sequences can be assessed in a variety of ways, including detection of a known sequence through an approach such as the polymerase chain reaction (PCR – which may be quantitative – qPCR) or through an approach using conserved regions of the gene to obtain a large number of ribosomal RNA gene sequences from a sample. The sequences can then be analysed to obtain information about the composition of the bacterial community in that sample. The use of next generation sequencing (NGS) is often used to obtain these sequences. NGS has the potential to produce millions of sequences in parallel; although each sequence tends to be relatively short, the information is suitable for bacterial community assessment – also called bacterial community metagenomics, of which ribosomal RNA gene profiling is but one part (Zhou *et al*., 2015).

The use of a sequence-based approach through NGS is a relatively new phenomenon; however, microbiologists have attempted to characterise bacteria for many decades, in both clinical and environmental studies. Bacterial isolates have been profiled in many different ways, often with the aim of determining whether an isolate is the same, or different, from other isolates. In the clinical setting, this may confirm the

identity of an isolate and/or determine whether a known or an unknown strain is responsible for an infection. Profiling bacterial isolates using DNA is well established, and often utilises specific features within the chromosome. The chromosome is a large DNA molecule that is the result of millions of generations, during which the DNA will have repeatedly acquired (and possibly lost or changed) a variety of mobile genetic elements; small pieces of DNA that can readily move around both within and between cells. They may be seen both as a burden and a benefit to bacteria, but once established, offer a potential method to profile the host. Again, a wide range of methods have been developed to enable profiling. One such method which has been studied in bacteria of the genus *Streptococcus* uses a mobile genetic element called BOX. This element may be found in various sites within the chromosome, and may be repeated multiple times. Using PCR, it is possible to copy DNA across BOX elements within a bacterial isolate. The number of repeats and the distance between the BOX elements directs the resulting sizes of the PCR products, which can then be used to profile that bacterium. The general approach of repetitive element PCR (rep-PCR) is well established, with different repetitive elements being targeted in different assays. Use of the BOX element is well established in bacteria such as the streptococci (it has also been used for profiling many other bacterial species; Versalovic *et al*., 1994).

Briefly, then, water quality is monitored using FIO concentrations through the bathing water directive to protect public health. The possibility that non-faecal bacteria could be detected using the specified methods, thereby over-estimating a risk to public health, does exist, but allowing for this is not possible under the bathing water directive. Culturing FIO provides data on their concentrations, but gives no information on the source of that bacterium, although the isolates can be profiled. Assessing bacteria to access further information often requires a molecular biological approach to work with the uncultured bacterial majority. The molecular biological tools include detection and quantification of gene sequences associated with known host animals that may be a source of pollution (MST). The tools also include metagenomics through NGS to assess the numerically abundant members of a community.

Bathing water quality at Scarborough South Bay is a cause of concern for the Environment Agency. Bathing water classification data for the previous few years are available on the gov.uk website: https://environment.data.gov.uk/bwq/profiles/data-samples.html?_search=Scarboro&bw=uke2206-07400#2016

In recent years, water quality at Scarborough South Bay has been classified as poor (2013, 2016) and sufficient (2014, 2015). The data show that the bathing water is subject to intermittent episodes of increased FIO concentrations, in addition to a baseline concentration that is cause for concern. The “spikes” in the data have impacted on the classification within the last 4 years.

Rationale.

A set of samples (including water and sediment) that aimed to represent the conditions around the Scarborough South Bay bathing water was collected. Analyses included standard FIO counts, microbial source tracking of faecal pollution, next-generation sequencing to characterise the bacterial communities within the various samples, and rep-PCR of IE isolated from the samples.

Materials and Methods.

Sample collection.

Samples were collected from sites around Scarborough. Sample types included water, soil and sediment. Occasionally, when sampling the bathing waters, a routine sample was collected and the bottle was sealed, after which the sampler then actively disturbed the sand/sediment on the sea bed and collected a separate sample. Samples were collected from different places within the industrial processing plant; namely the raw water used during processing (taken from a borehole on site); the soil from potatoes stored on site; and from the industrial effluent sampled on site. Effluent samples were also taken from Scarborough WwTW, after U.V. treatment. One faecal sample from a gull was collected from the beach at Scarborough South Bay, close to the designated sample point for the bathing water. Sample points are detailed in Table 1, with more general areas of interest further illustrated in Figure 1.

Samples were collected by Environment Agency staff in accordance with standard sample collection procedures and delivered to the Starcross laboratory (Exeter) where they were processed within 24h of being taken.

Sampling took place over the period August to September 2016. Not every sample point and matrix was sampled on every occasion, and not every possible test was carried out on every sample. A full list of samples and associated data points is presented in Appendix 2.

For clarity, only an overview of the laboratory methods is provided here, with a more detailed description of the laboratory procedures and associated data analysis presented in Appendix 1.

Samples were processed for standard FIO counts, for MST, for bacterial community analysis using ribosomal RNA genes through NGS, and for rep-PCR of intestinal enterococci isolates. Samples were also processed for chemical analysis, especially for the herbicide chlorpropham, which is used to inhibit sprouting in potatoes.

Site descriptions (notes)	Location (NGR)
Industrial processor, final effluent (sampled on site)	TA 05067 83657
Industrial processor, raw process water (from borehole, on site)	TA 05067 83658
Scarborough WwTW (final effluent, after U.V.)	TA 02593 92221
[a]Scarborough South Bay (compliance point)	TA 04577 88566
Whitby Rd pumping station (CSO outfall, Jackson's Bay)	TA 03671 91292
[a]Scarborough North Bay (compliance point)	TA 03769 89807
Scarborough South Cliff (Wheatcroft beach)	TA 05943 86276
Industrial processor potato store	TA 05067 83656

Table 1. Sampling points used for this study. [a] Water and sediment samples were collected (also a bird faecal sample). Not every site was sampled for every matrix at each time point. Some samples collected during this project were for chemical analysis only (no bacteriological analysis). Samples were collected over the period August to September 2106. A full list of the samples is presented in Appendix 2.

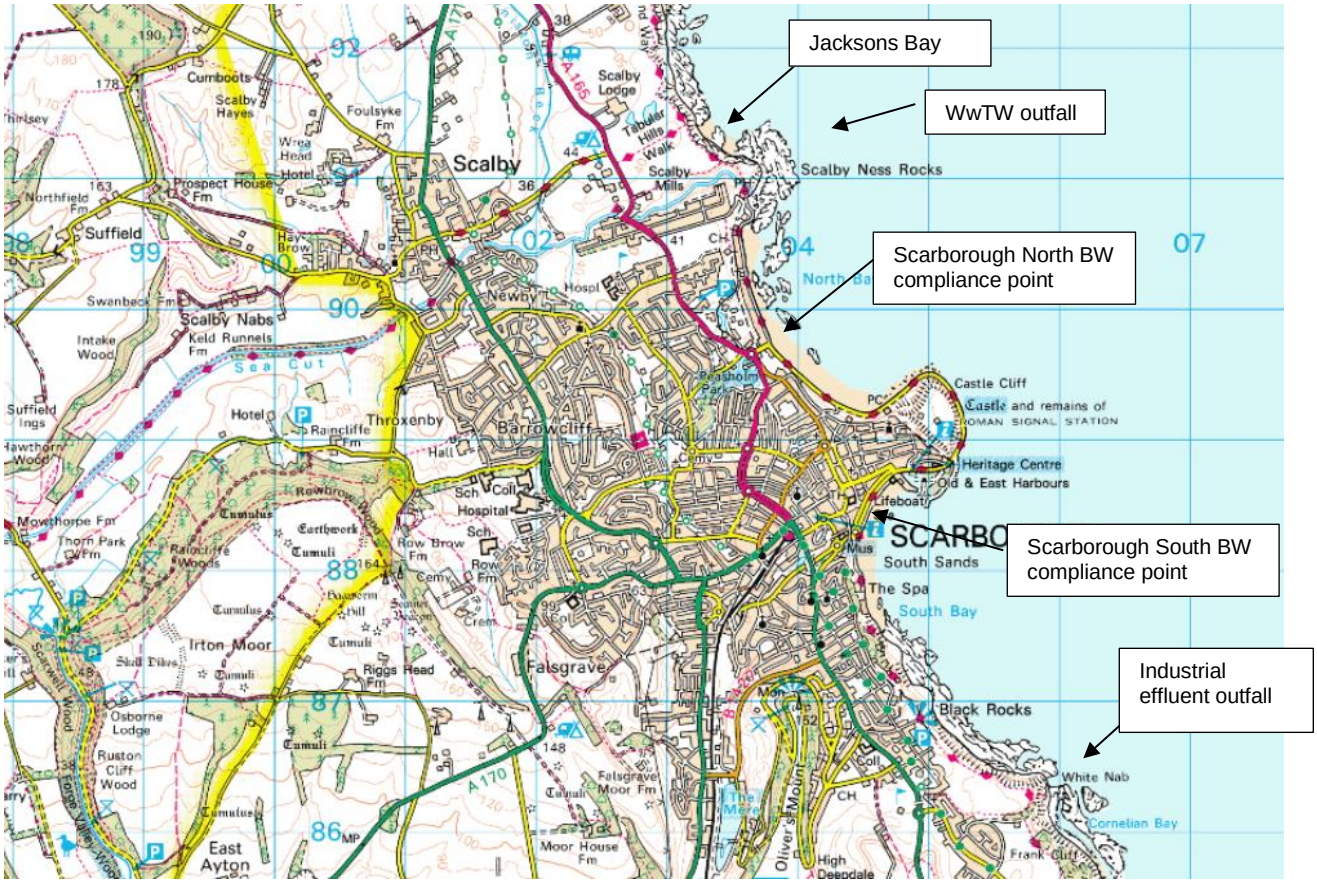

Figure 1. Map of the sampling area highlighting points of interest within the overall sampling regime.

Results and Discussion.

Enumeration of FIO.

Summary statistics on the FIO concentrations in this study are presented in Table 2. FIO concentrations from some individual samples are presented in Table 3, alongside the relevant concentrations of the MST markers. A full set of data is presented in Appendix 2. FIO counts for the solid matrices (Table 2) show that per gram wet weight, bird faeces contains the highest FIO numbers. Counts for the water matrices show the WwTW and the industrial effluents contained the highest FIO concentrations, with the industrial effluent containing the highest numbers of IE (Table 2). Figures for the 90th and 95th percentile values are also presented in Table 2 to provide some context for bathing water classification. It must be noted that this context is not directly applicable to the bathing water directive for solid matrices. Data for the water and for the solid matrix samples were ranked in general terms (not a formal statistical ranking) according to concentrations of *E. coli* and concentrations of IE. For the *E. coli* data in water, the ranking by concentrations followed the general order:

Industrial effluent > WwTW effluent > South Bay > Whitby Road PS ≈ North Bay.

However, this general ordering is not absolute, with 2 of the 8 industrial effluent samples showing intermediate concentrations of *E. coli* and 2 of 8 showing low concentrations of *E. coli* (Table 3).

For the *E. coli* data in solid samples, the ranking by concentrations followed the order:

Bird faeces >> Potato store >> South Bay sediment ≈ North Bay sediment

When ranking the sites according to concentrations of IE in the water samples, the order was:

Industrial effluent >> WwTW effluent > South Bay > Whitby Road PS ≈ North Bay

Again, the order was not absolute, as some exceptions were noted as for the *E. coli* data.

For the IE data in solid samples, the ranking by concentrations followed the series:

Bird faeces > Potato store >> South Bay sediment > North Bay sediment

	Number of samples	*E. coli*				IE			
		Mean	SD	90%tile	95%tile	Mean	SD	90%tile	95%tile
Potato store	9	4.06	0.88	1.53×10^5	3.22×10^5	5.27	0.69	1.41×10^6	2.53×10^6
North Bay sediment	9	0.11	0.32	3	4	0.14	0.32	4	5
South Bay sediment	13	0.13	0.29	3	4	0.68	0.63	30	52
Bird faeces	1	7.17				6.88			
South Cliff sediment	1	Not detected				Not detected			
Raw process water	10	0.58	0.60	22	37	0.91	1.02	165	390
Industrial effluent	10	3.93	2.52	1.45×10^7	1.22×10^8	5.72	3.54	1.79×10^{10}	3.60×10^{11}
North Bay water	13	1.10	0.21	23	28	1.00	0.00	10	10
South Bay water	13	1.63	0.51	192	296	1.69	0.66	348	610
South Cliff water	4	1.00	0.00	10	10	1.11	0.22	24	29
WwTW effluent	8	3.78	0.60	35415	59024	3.10	0.69	9637	17342
Whitby Rd pumping stn	10	1.26	0.50	79	121	1.21	0.42	56	79

Table 2. Descriptive statistics for FIO counts for solid (/g wet weight) and water (/100 ml) matrices. Averages are expressed as $\log_{10}$ CFU. Only a sub-set of samples analysed for FIO were subject to further (molecular biological) analysis; see Appendix 2.

It was noted above that the bathing water quality at South Bay has been subject to occasional “spikes” in FIO counts over the last 4 years. This was also observed during this study (data not shown). The FIO concentrations in the industrial effluent were also noted to be highly variable, covering the range 10^1 to 3×10^6 *E. coli* /100 ml and 10^1 to 3×10^9 IE /100 ml (Figure 2).

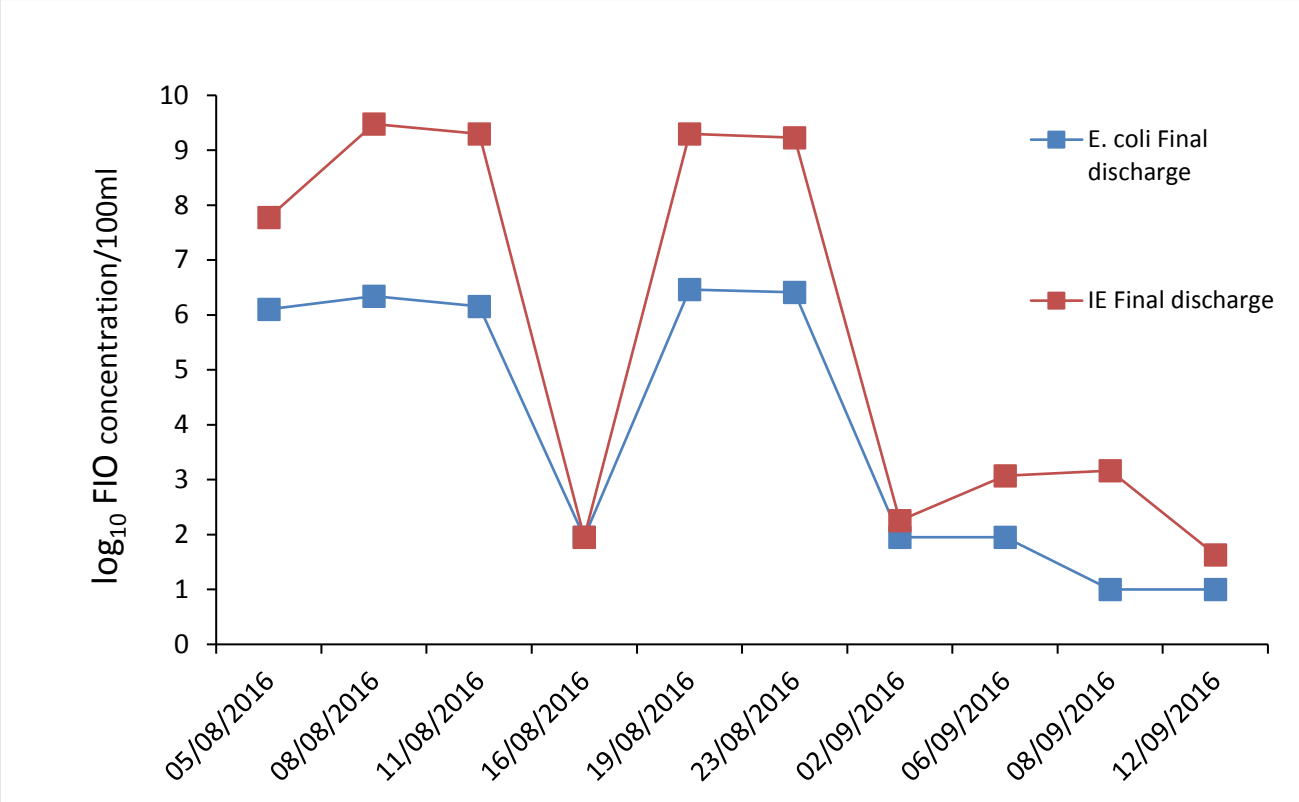

Figure 2. Concentrations of *E. coli* and IE from the industrial effluent over the time of this study.

Microbial Source Tracking.

Only those samples with the higher FIO counts were selected for MST analysis. (The same caveat applies to the samples processed for NGS.) MST data are reported in gene copies (gc) per unit volume. The ruminant bacterial marker was not detected at any site, or any sample tested. Dog mitochondrial DNA was found in two water samples only, both from the same site (Scarborough South Bay $\log_{10}$ 1.83 gc/100 ml on 11th August and $\log_{10}$ 2.28 gc/100 ml on 8th September). No other samples from any other sites (including the wastewater treatment effluent and the industrial effluent) showed the presence of dog DNA.
The known gull faecal sample from South Bay was strongly positive for the seabird marker at a concentration of approximately 10^{11} gene copies /gram wet weight faecal material. Measuring such a high concentration is difficult and inaccuracies will be introduced from dilutions etc. Other workers have reported similar estimates from gull faeces (Green *et al*., 2012; Ryu *et al*., 2012).

Data for the human and seabird marker are presented in Table 3. When the water samples were ranked according to concentration of the human-specific marker, the order was:

WwTW effluent > industrial effluent > Whitby Road PS > South Cliff ≈ North Bay > South Bay

Only two South Cliff and one North Bay samples were analysed for MST, therefore the positioning of these sites within the overall ranking must be treated with caution.

When ranked by the concentration of the seabird marker, the order of the water samples was:

South Bay > South Cliff = North Bay = Whitby Rd pumping station > WwTW effluent > potato store > industrial effluent

As above, only two South Cliff and one North Bay samples were analysed, therefore the positioning of these sites within the overall ranking must be treated with caution.

Very high numbers of the seabird marker were found in the seabird faecal sample (Table 3); however, it is difficult to incorporate the gene copies from a solid sample (gc /g wet weight) with the water (gc /100 ml) samples.

The seabird marker was found in samples taken from the plant at low levels only (Table 3).

Associations between the concentrations of the FIO and the MST markers were assessed using correlations of both $\log_{10}$ transformed data and of ranked data. The strongest overall correlations noted were between concentrations of *E. coli* and IE, with R-sq values of 0.81 and 0.79 (parametric and ranked correlations respectively) across all data points. When zero data points were removed from the data set, the R-sq values between the concentrations of the human marker and *E. coli* concentrations were 0.45 and 0.52 (parametric and ranked correlations respectively). No other associations were noted in the overall data. Correlations between the concentrations of the microbial parameters were also examined within each site. At some sites, too few data points were available for both FIO and MST. However, correlations were noted between the FIO and the seabird marker at South Bay (R-sq values of 0.67 and 0.61 *E. coli*/seabird and IE/seabird parametric correlations respectively, Figure 3) and between the concentrations of the FIO and the human marker for the WwTW effluent samples (R-sq values of 0.67 and 0.86 *E. coli*/human and IE/human parametric correlations respectively, Figure 3).

The human-specific MST marker was consistently detected in the industrial effluent. The bacterium used for the assay comes from the Bacteroidetes; a group of bacteria known to be tolerant or demanding of anaerobic conditions. The (few) well characterised bacterial species from this group have also been suggested to be effective users of carbohydrate from their environment. To investigate whether the human-specific bacterial marker was likely to have come from (recent) human faecal material, or from growth of an established population within the infrastructure, those samples were analysed further for human mitochondrial DNA (the rationale being that mitochondria cannot establish or replicate in the environment; therefore the presence of human faecal bacteria and human mitochondrial DNA together would indicate recent addition of the material; the presence of the bacteria without the mitochondrial DNA leaving considerable room for doubt). Human mitochondrial DNA was detected in 6 out of 8 samples, suggesting recent pollution from human sources was present in the industrial effluent.

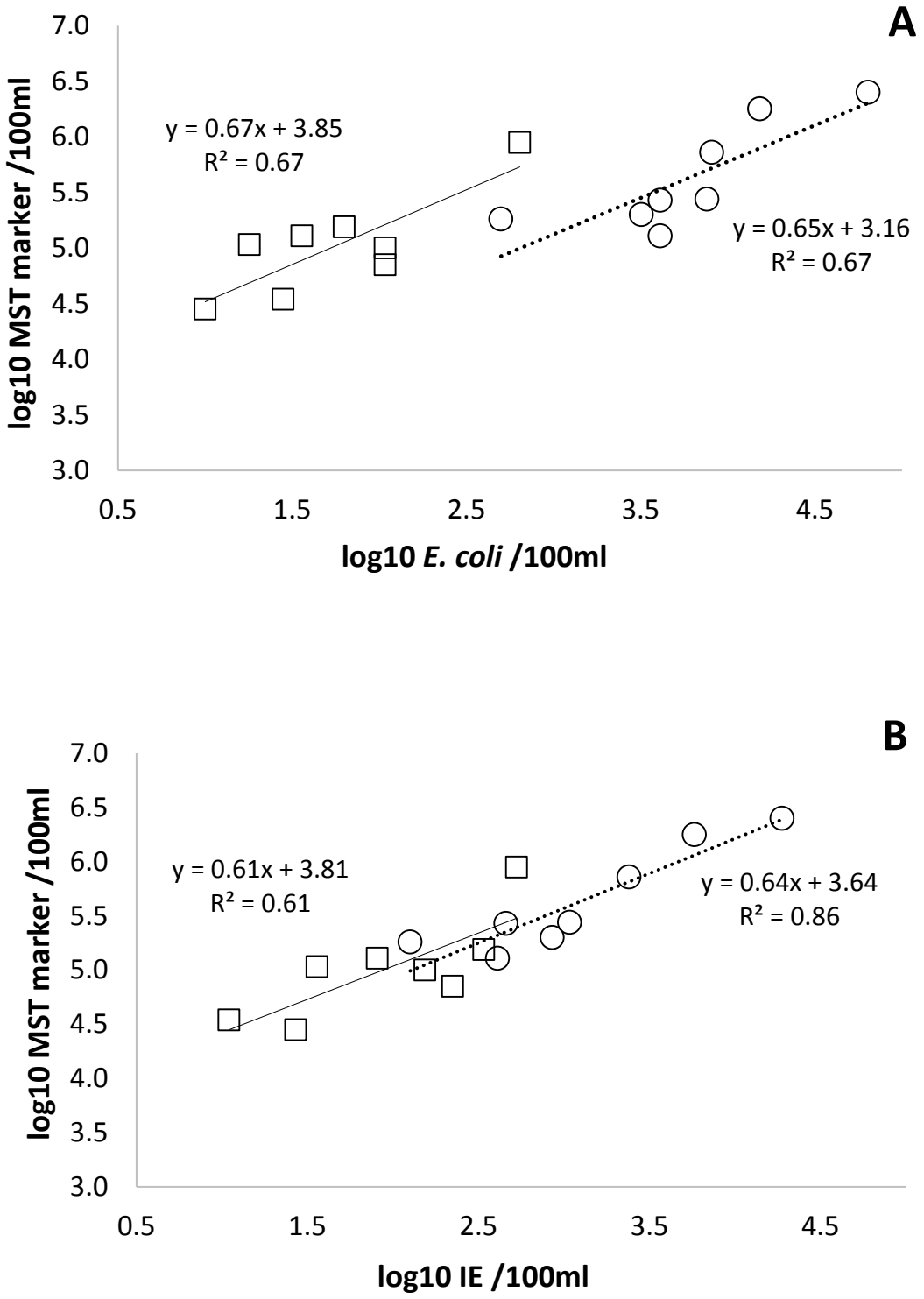

Figure 3. Significant correlations between concentrations of MST markers and FIO. A; *E. coli*; B; IE. Squares and solid lines; Seabird MST marker in Scarborough South Bay water. Circles and dashed lines; Human MST marker in wastewater treatment work samples. Adjacent equations describe fitted line for each dataset.

Sample site	Date	Time	IE	*E. coli*	Human	Seabird
Industrial effluent	08/08/16	09:50	9.48	6.34	5.32	0.00
Industrial effluent	11/08/16	13:00	9.30	6.16	4.56	0.00
Industrial effluent	23/08/16	11:03	9.23	6.41	4.23	0.00
Industrial effluent	05/08/16	11:16	7.78	6.10	3.45	2.08
Bird faeces	19/08/16	12:45	6.88	7.17	0.00	11.02
Potato store	23/08/16	11:07	6.23	5.65	0.00	0.00
Potato store	02/09/16	11:25	5.95	4.34	0.00	0.00
Potato store	08/08/16	10:00	5.65	4.21	0.00	2.70
Potato store	08/09/16	10:52	5.49	3.96	0.00	0.00
Potato store	11/08/16	13:50	5.08	3.28	0.00	2.91
Potato store	12/09/16	10:50	4.86	3.96	0.00	0.00
Potato store	06/09/16	11:07	4.62	3.16	0.00	0.00
WwTW effluent	12/09/16	09:45	4.28	4.81	6.40	2.62
Potato store	19/08/16	11:00	4.03	2.95	0.00	0.00
WwTW effluent	11/08/16	14:05	3.76	4.19	6.25	3.40
Raw process water	11/08/16	13:05	3.40	1.73	0.00	0.00
WwTW effluent	02/09/16	13:30	3.38	3.91	5.86	2.56
Industrial effluent	08/09/16	10:55	3.16	1.00	2.92	0.00
Industrial effluent	06/09/16	10:54	3.07	1.95	2.41	0.00
WwTW effluent	16/08/16	12:15	3.03	3.88	5.44	0.00
WwTW effluent	08/09/16	11:30	2.93	3.51	5.30	0.00
South Bay	23/08/16	11:25	2.72	2.81	3.20	5.95
WwTW effluent	05/08/16	12:41	2.66	3.61	5.43	2.14
WwTW effluent	06/09/16	12:50	2.61	3.61	5.11	0.00
South Bay	05/08/16	12:57	2.53	1.80	3.78	5.19
South Bay	19/08/16	12:30	2.35	2.03	2.74	4.85
Industrial effluent	02/09/16	11:15	2.26	1.95	4.22	2.76
Whitby Rd pumping stn	05/08/16	11:40	2.23	2.61	4.07	3.65
South Bay	19/08/16	12:30	2.19	2.03	2.77	5.00
WwTW effluent	23/08/16	11:55	2.10	2.70	5.26	3.24
South Bay	16/08/16	11:55	1.91	1.56	2.66	5.11
Whitby Rd pumping stn	08/08/16	11:57	1.65	1.00	2.13	3.44
Industrial effluent	12/09/16	10:55	1.63	1.00	2.83	0.00
South Bay	11/08/16	13:00	1.56	1.26	0.00	5.03
South Bay	08/08/16	13:15	1.43	1.00	0.00	4.45
South Cliff	08/08/16	10:47	1.43	1.00	2.58	4.17
Raw process water	02/09/16	11:20	1.41	0.60	0.00	0.00
Whitby Rd pumping stn	23/08/16	10:40	1.26	1.00	3.46	3.38
South Bay	08/09/16	11:42	1.04	1.45	2.61	4.54
South Cliff	05/08/16	13:40	1.00	1.00	3.36	3.75
North Bay	06/09/16	11:42	1.00	1.00	2.58	3.87

Table 3. Concentrations of MST markers (gene copy/100 ml) and associated FIO (cfu/100 ml) taken this study. All data are $\log_{10}$ X /100 ml or g wet weight. Data are arranged in decreasing order of concentrations of IE. Data values of 0.00 should be taken as below limit of detection (rather than zero).

FIO and MST concentration data were examined for any possible associations with the solar U.V. index, tidal state, prior rainfall, issue of a short-term pollution warning and known CSO spill. Data on these parameters was provided by Environment Agency staff. The sampling regime did not target known environmental conditions (that is, samples were collected on planned days, rather than in response to rainfall or tidal state), and thus this dataset was too limited to be of use. However, during the sampling period, 3 CSO spill events were recorded that coincided with samples being collected within 24 to 72h of the spillage. Spills on the 4th August occurred at assets in Scarborough north and Scarborough south; spills on the 22nd August and the 3rd

September only occurred at assets located in Scarborough north. FIO and MST marker concentrations were elevated threefold or more at sampling dates after a spill event (Table 4).

Hours since spill	FIO			MST		
	n	*E. coli*	IE	n	Human	Seabird
24	3	163	324	2	3090	3.72×10^5
No spill	10	28	28	6	63	6.76×10^4
72	4	94	136	n/a	n/a	n/a
No spill	9	30	31	n/a	n/a	n/a

Table 4. Geometric mean concentrations of FIO and MST markers (/100 ml) in South Bay samples taken soon (24 to 72 h) after a spill event, and those with no recent spill event.
n; number of samples, n/a; not available (no samples were processed for MST in the appropriate time period).

As an addendum to this project work, the possibility of faecal pollution from donkeys working on the beach at South Bay was questioned. An MST marker for bacteria found in large numbers in horses is available; this marker was tested on 10 separate donkey faecal samples from physically separated, healthy donkeys (samples kindly provided by the Donkey Sanctuary, Sidmouth, Devon). Half (5 out of 10 samples) tested as positive for the horse bacterial marker. The South Bay DNA extracts were then tested for the presence of the horse marker; all samples were negative. It is understood that 2 concessions for donkeys were working on the South Bay beach, with 4 donkeys for each concession (8 donkeys in total). Although the data from the known samples from the Donkey Sanctuary suggest only a sub-set of donkeys carry the horse bacterial MST marker, the available evidence would suggest donkeys on the beach did not have a detectable impact on water quality during the sampling period.

Overall, the MST data suggested that human and seabird pollution markers were frequently found at the South Bay sample site, with strong associations between FIO concentrations and the seabird marker. Dog mitochondrial DNA was also found on two occasions, indicating dogs were present on the beach, and therefore a possible source of faecal pollution. A very limited dataset suggested pollution from donkeys was not a factor for South Bay.

Bacterial community profiling using next-generation sequencing.

DNA from the same extractions used for MST was then used for bacterial community profiling using next generation sequencing. Sequences were analysed and compared in several ways. This targeted the numerically most abundant bacterial ribosomal RNA genes in a culture-independent approach for the samples. Within this context, and within the constraints of the level of resolution available from the sequence data, detection of *E. coli* and IE within the data was not expected, and indeed, sequences directly associated with these bacteria were not found. However, the data were used rather to assess bacterial community diversity with the hypothesis that effluent (whether industrial or from the WwTW) would be likely to show a reduced community diversity that would be distinct from the communities present in natural seawater samples. Data were analysed by time and by site. No patterns were noted when data were analysed over time; but clear groupings of the data were shown by site. Data for the 4 sites with the greatest processed sample numbers are presented in Figure 4. The proportion of the total community diversity represented by different bacterial phyla shows notable differences at each site, with a greater proportion of the phylum Firmicutes (broadly equivalent to the traditional Gram positive bacteria) highest in the industrial effluent, and a greater proportion of the phylum Proteobacteria (broadly equivalent to the traditional Gram negative bacteria such as the Enterobacteriaceae) highest in the WwTW effluent. A clear difference in community profile was noted between the potato store and the industrial effluent, with a shift away from Proteobacteria towards Firmicutes

(Figure 4). The diversity at each site was lowest in the industrial effluent and greatest in South Bay (data not shown).

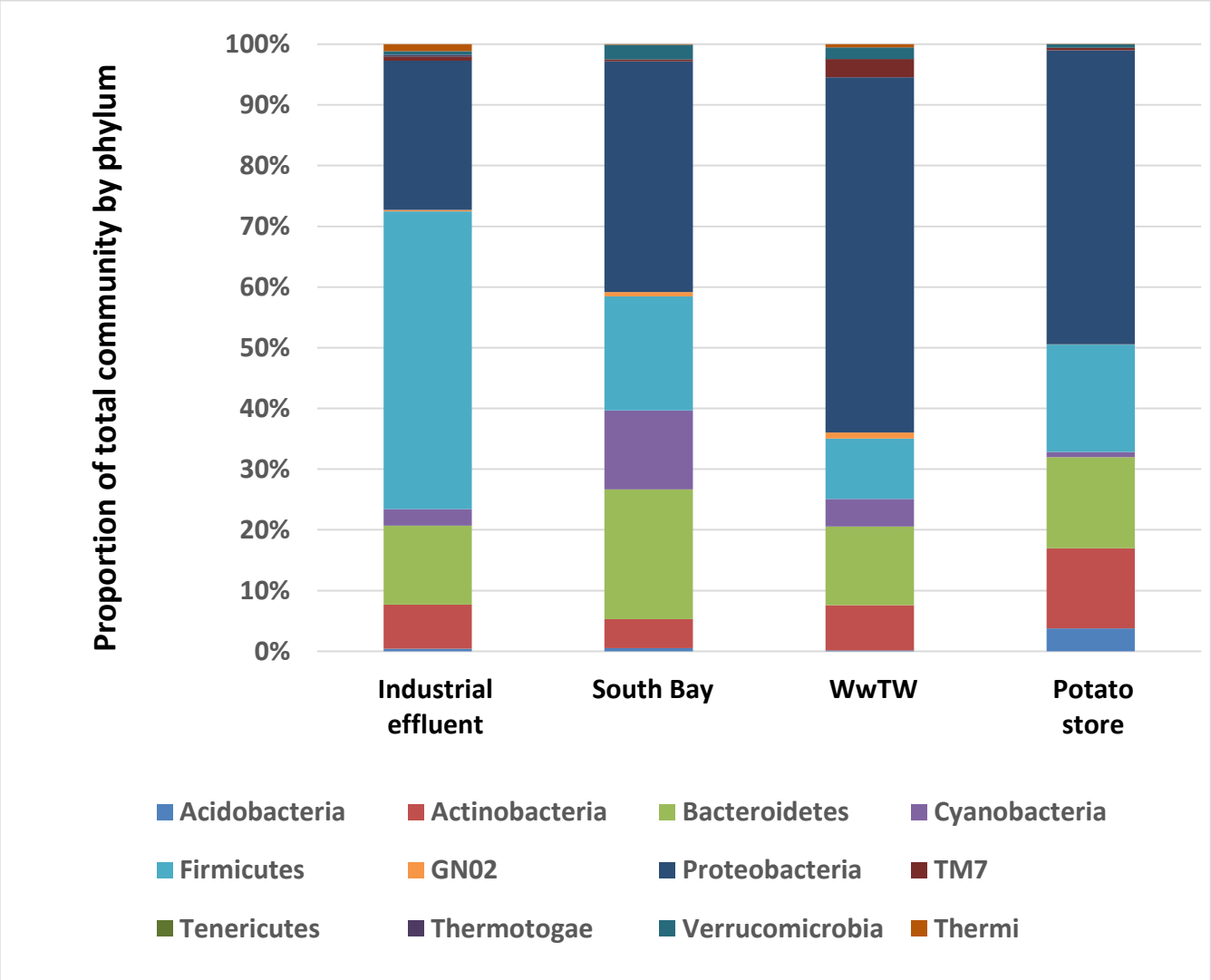

Figure 4. Proportion of total bacterial community diversity from different bacterial phyla at 4 of the target sites.

Similarities within the dataset were explored more fully using non-metric multidimensional scaling. Multidimensional scaling attempts to plot the data in such a way that the distance between the points shows (in some way) the experimental dissimilarity of all the samples within the dataset. In this particular case (Figure 5), each sample in which bacterial community diversity was assessed is represented by a coloured circle. The colours reflect the site at which the sample was taken, e.g. all samples taken from the WwTW are coloured purple. The NMDS plot represents the data so that the distance between the data points (sample sites) reflects how different the bacterial communities were from each other. The further the data points are from each other, the more different they are; however, the scale is not necessarily linear or exact (that is twice the distance may not mean twice as dissimilar). The fact that all the coloured circles are close together is because they are less dissimilar to each other than they are to all the other sites. (When the data were coloured by times, no apparent patterns or clustering were present; data not shown.) When plotted by site, there was a clear clustering of several sites - the industrial effluent, the potato store and the WwTW effluent (Figure 5). Exceptions to this were noted, with two samples from the industrial effluent being separate from the other industrial effluent samples (Figure 5).

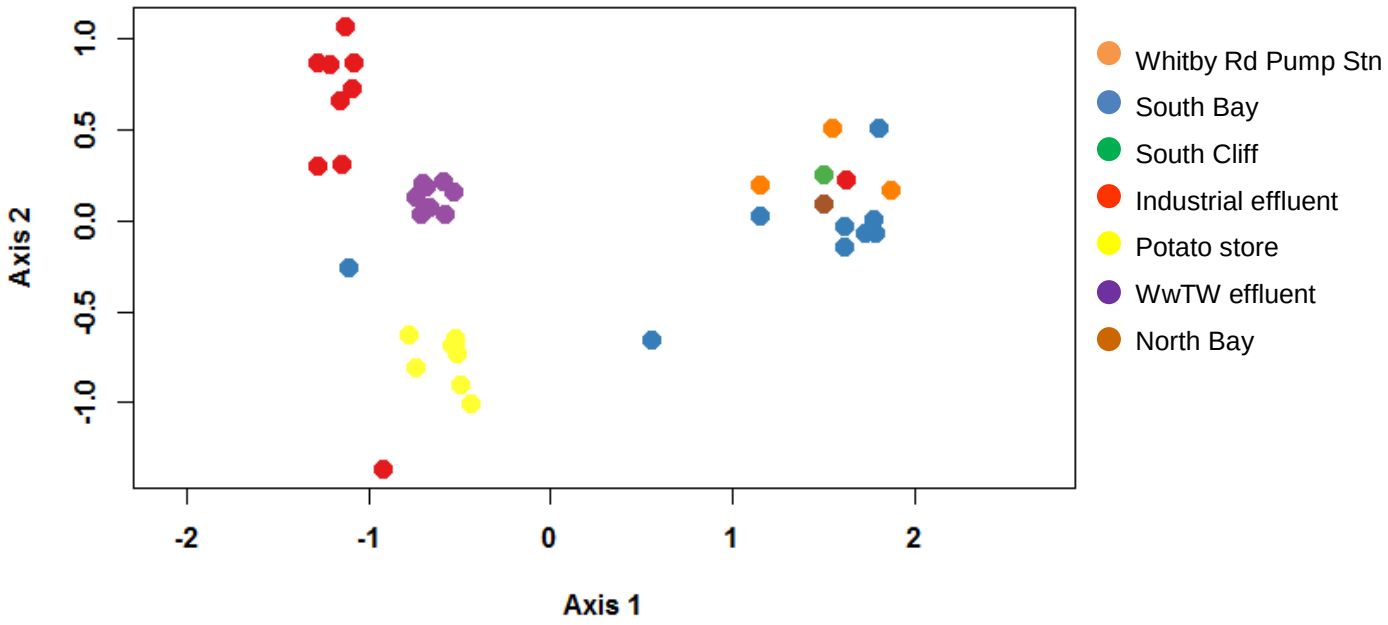

Figure 5. Non-metric multidimensional scaling plots of NGS data from bacterial community profiling at different sites.

Overall, the NGS data suggest that the bacterial community diversity found at the South Bay site was the highest of the samples analysed, with less diversity within the effluent samples. The NMDS plots suggest that for the majority of the time, sequences similar to those from the industrial effluent were not found at the South Bay sampling point, but that on occasion (2 samples out of 10; 20% of the time) sequences from the industrial effluent were detected. Interestingly, the date of the industrial effluent sample that grouped with the South Bay sample (11th August, 2016) is the date of the South Bay sample that is also dissimilar to the other South Bay samples (approximate coordinates Axis 1: -1.1; Axis 2 -0.3; Figure 5).

Isolate profiling of intestinal enterococci using rep-PCR.

Individual IE isolates from sites over time were profiled using rep-PCR. Following quality control of the data, analyses were undertaken using molecular weight sizes rounded to 20 bp and to 50 bp. The data from the 20 bp data only are presented and discussed here; similar results were found from with 50 bp data.
Rep-PCR produces a profile of a bacterial isolate based upon the frequency and position of very short repeated sequences within the host bacterium genome. A wide variety of small mobile genetic elements have been found in bacteria; simplistically, they can be thought of as short sections of infectious DNA. As such, the position, frequency and number of the repeated sequences can be thought of as representing the infection history of the isolate being studied.
Having produced a rep-PCR profile of each isolate, the initial question was whether identical profiles were present within the full dataset. Few exact duplicates were detected; those that were detected were from the South Bay site (either water:water or water:sediment); one pair matching between South Bay sediment and North Bay sediment; and one pair within the bird faecal sample. More duplicates (and triplicates) profiles were detected in the 50bp data set, as would be expected from a loss of resolution; however, nothing in either data set from exact profile matches suggested an easily-measured impact from the industrial effluent IE on the South Bay samples, whether water or sediment.
Given that exact matches were not found frequently in the isolates rep-PCR profiles, the profiles were then compared for closest matches. Comparisons were performed in multiple ways as no definitive procedure exists to cluster data together. The rep-PCR data are binary profiles (present/absent of repeat sequences at

given distances between the insertion events) and as such are not amenable to cluster analysis using Euclidean distances.

Isolate profiles were initially compared using the Sorensen-Dice similarity coefficient. This compared every isolate with every other isolate within the dataset. The average coefficients within and between sites were then calculated and are presented below (Table 5). Several sites contained isolates more similar to others from the same site than from different sites (e.g. bird faeces, raw process water from the plant; Table 5) while other sites contained isolates with profiles more similar to those from different sites (e.g. South Bay sediments). Overall, the similarity indices were not notably high (that is, the isolate profiles were dissimilar) with the overwhelming majority (>95%) of all the Sorensen-Dice coefficients being less than 0.499. This strongly suggests a high level of diversity amongst the IE bacteria in these samples.

	BF	JB	NBS	NBU	NBW	Out	PS	Raw	SBD	SBS	SBU	SBW	SCS	WW
BF	**0.20**	0.09	0.11	0.08	0.07	0.11	0.07	0.06	0.08	0.12	0.10	**0.12**	**0.15**	0.08
JB		**0.13**	**0.14**	0.15	0.15	0.12	0.10	0.14	0.13	0.12	0.11	0.11	0.10	0.12
NBS			0.11	0.13	**0.17**	0.14	0.08	0.13	0.11	**0.14**	0.12	**0.12**	0.13	0.12
NBU				**0.27**	0.08	**0.17**	0.11	0.15	**0.14**	**0.14**	**0.13**	0.11	0.12	0.10
NBW					0.30	0.13	0.10	0.16	**0.14**	0.13	0.12	0.09	0.14	**0.15**
Out						0.16	0.09	0.12	0.11	**0.14**	0.12	0.11	0.12	0.10
PS							**0.13**	0.12	0.11	0.10	0.09	0.09	0.08	0.10
Raw								**0.17**	**0.14**	0.12	0.11	0.10	0.08	0.13
SBD									**0.14**	0.12	0.11	0.11	0.11	0.12
SBS										0.13	0.11	0.11	0.12	0.11
SBU											0.10	0.10	0.13	0.10
SBW												0.11	0.11	0.10
SCS													0.07	0.08
WW														0.12

Table 5. Average Sorensen-Dice similarity coefficients within and between sites. Values in bold represent the highest average similarity by column. BF – bird faeces; JB – South Cliff; NBS – North Bay sediment; NBU – North Bay sediment undisturbed at time of sampling; NBW – North Bay water; Out – industrial effluent; PS – potato store; Raw – raw process water; SBD – South Bay sediment, disturbed before sampling; SBS – South Bay sediment; SBU – South Bay sediment, undisturbed at time of sampling; SBW – South Bay water; SCS – South Cliff; WW – Wastewater Treatment Works.

The binary profiles were further compared using a variety of approaches; parsimony, distance and Bayesian. A typical set of data from parsimony and distance methods are shown in Figure 6, with that from Bayesian methods in Figure 7. All approaches suggested a similar pattern in that some isolates from each site were more similar to each other than to isolates from other sites. This was most apparent from the Bayesian analyses (Figure 7). However, all approaches suggested at least some isolates were more variable, and therefore grouped with isolates from other sites (Figures 6 and 7). Given the number of isolates and sites involved, presenting the data with different colours and labels is difficult (Figures 6 and 7) and therefore the cluster data are also summarised for a sub-set of sites (Table 6).

The known sources of IE within the data set are bird faeces, the industrial effluent and the WwTW. IE from other sources can be assumed to be from a mixture of sources. Data from Table 6 show that for each clustering/tree method, when the clusters are split at a similarity level that produces the same number of clusters, with approximately the same number of isolates in each cluster, isolate profiles representing IE from bird faeces and from the industrial effluent are less widely distributed across the clusters than isolates from the WwTW effluent. Nearly half the clusters contain at least one IE profile from the WwTW effluent, less than a quarter contain an isolate profile from bird faeces (Table 6).

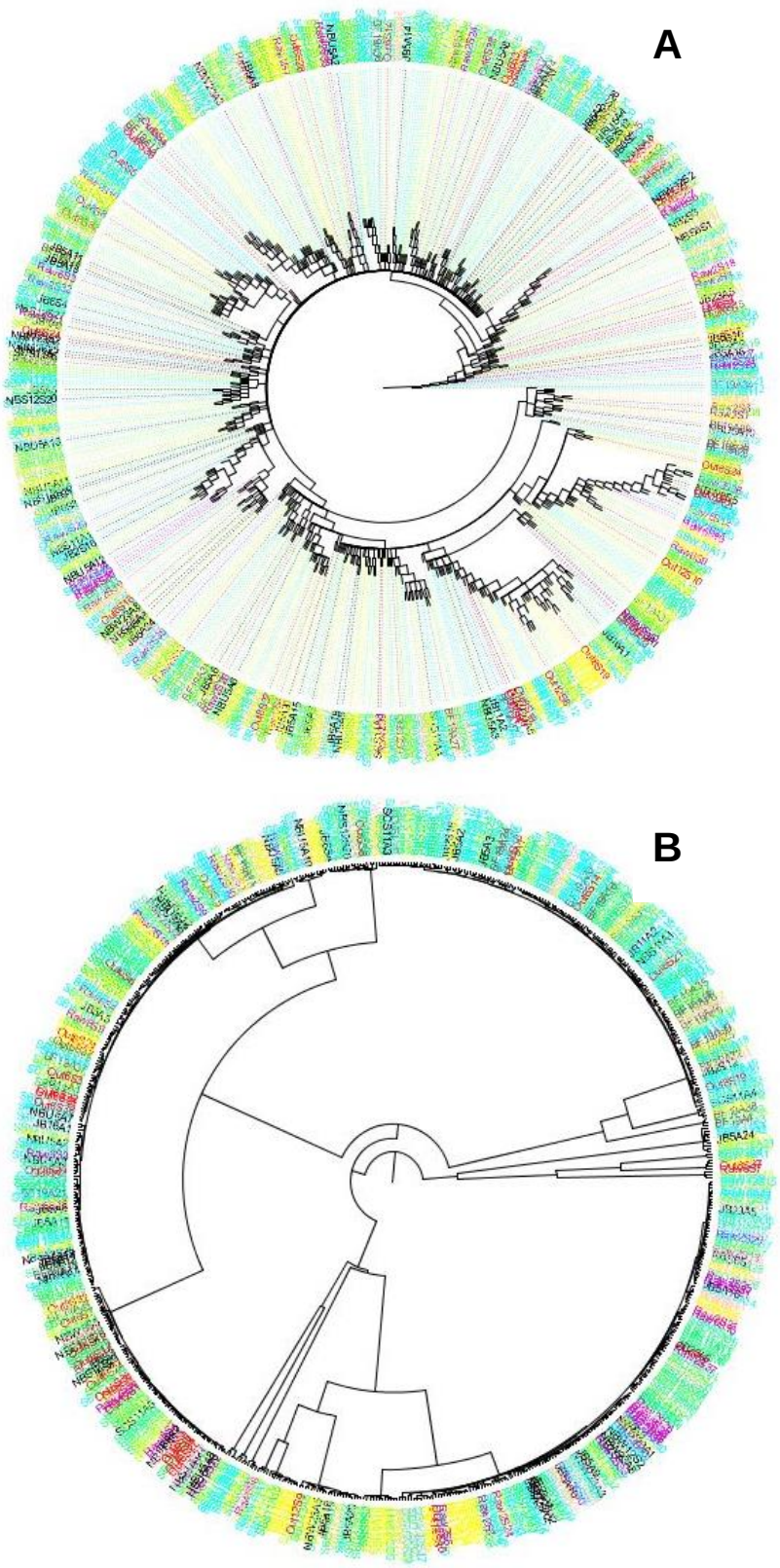

Figure 6. Example output from 20 bp rep-PCR isolate profiles. A –consensus parsimony tree produced from bootstrap sub-sampling (1000 replicates). B –distance tree produced from bootstrap sub-sampling (1000 replicates) with jumbled input order (x100). Light blue = isolates from South Bay water; light green = isolates from South Bay sediment; red = isolates from the industrial effluent; purple = isolates from the process water; pink = isolates from the wastewater treatment works; yellow = isolates from the potato store; brown/orange = isolates from bird faeces.

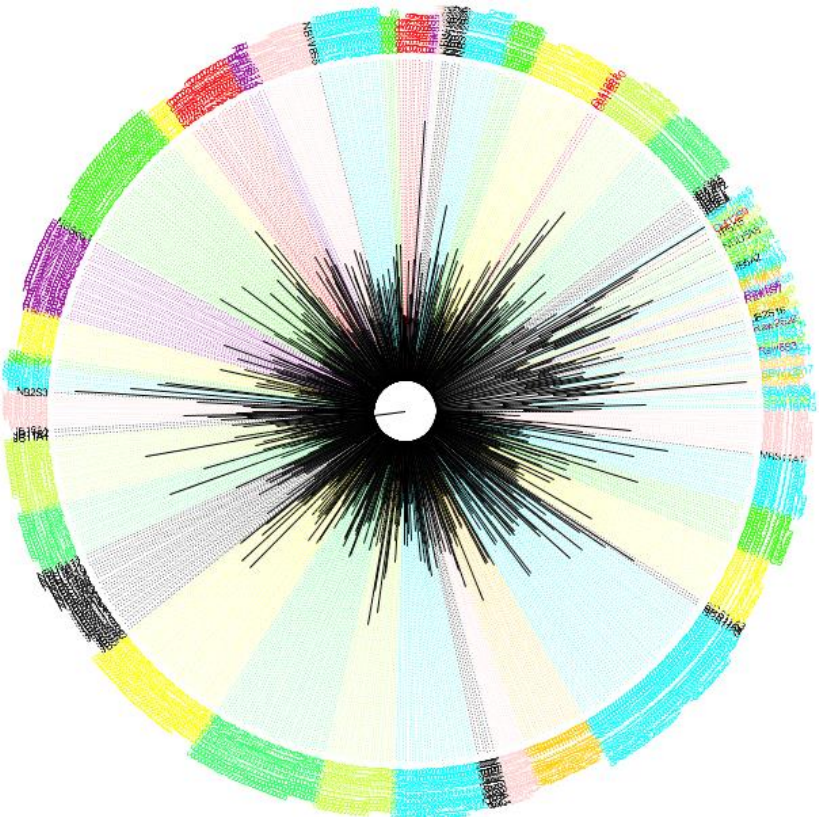

Figure 7. Example output from 20 bp rep-PCR isolate profiles analysed for similarity using Bayesian inference. Light blue = isolates from South Bay water; light green = isolates from South Bay sediment; red = isolates from the industrial effluent; purple = isolates from the process water; pink = isolates from the wastewater treatment works; yellow = isolates from the potato store; brown/orange = isolates from bird faeces.

	Parsimony	Distance	Method of analysis					
			Parsimony			Distance		
			BF	Out	WW	BF	Out	WW
Number of clusters	49	49						
Frequency (%) source-specific isolates present in a cluster			22.9	27.1	47.9	12.2	20.4	51.0
Average number total isolates/cluster	14.3	13.9						
Average number source-specific isolates/cluster			0.5	0.7	1.2	0.5	0.7	1.1

Table 6. Summary statistics of cluster analysis data presented in Figure 6. BF - bird faeces; Out – industrial effluent; WW – waste water treatment works effluent.

Chemical analysis.

Chlorpropham was only detected consistently in the industrial effluent, and nearly always at concentrations <1 µg/l (only 2 data points exceeded this value; 1.17 µg/l on 2nd September, when an unusually high pH value was also noted and 1.2 µg/l on 8th August). Chlorpropham was not detected in South Bay water (limit of detection 0.005 µg/l). None of the chemistry data supported any link between the industrial effluent and the bathing water.

Interpretation.

The body of water comprising Scarborough South Bay is clearly a large, complex system; attempting to fully describe and explain the relative sources, concentrations and movement of the different bacterial groups is always going to be difficult. Several samples collected from the industrial effluent had very low measured concentrations of bacteria compared to the majority of the samples from that effluent and one effluent sample (2nd September) had a high measured pH; apart from this, there is nothing to suggest that the sampling period covered atypical circumstances for the area. However, clearly only a limited number of samples could be processed, and there is very little replication of any variable in the system. All conclusions must be interpreted within this context. It must also be recognised that other sources of FIO, not yet considered, could impact the water quality.

The situation outlined at the start of this report was that there are multiple possible sources for the faecal bacteria found in Scarborough South Bay bathing water at any particular time. Although one clear candidate (point) source was the industrial effluent, other possible sources also required investigation. The data presented above represent an effort to determine bacteria pollution sources using several different approaches. Several definite conclusions can be drawn from the overall dataset; however, the data should not be used as the sole evidence base for management activities and expenditure. It is considered worthwhile, therefore, to use the data from this study, in conjunction with other published datasets, to interpret and generalise further for Scarborough South Bay. It must be highlighted that doing so is highly speculative, and should not be taken in any way as definitive findings.

The overall dataset suggest that the industrial effluent is not a constant source of bacteria to the South Bay bathing water; although “spikes” of bacteria from this effluent may impact water quality. There is evidence to suggest that pollution from humans, seabirds and dogs is also present in South Bay. The highest concentrations of human and seabird MST markers in South Bay seem to have at least some association with spills from the sewerage infrastructure, with associated higher concentrations of FIO, with the spills presumed to be linked to rainfall in some way.

Pollution from human sources.

The evidence suggesting human faecal pollution is present comes from DNA that was detected in the South Bay samples. If this DNA has passed through a sewage treatment works with fully operational U.V. disinfection, there is a strong likelihood that the culturable FIO in the WwTW effluent will have been greatly reduced (invisible to the methods used for regulatory monitoring, but not to the molecular biological methods). If the DNA is from waste not subject to U.V. irradiation (e.g. a CSO spill) then associated higher concentrations of live FIO will be present. It is recognised that any treated effluent is variable in nature and will be dispersed within the receiving water following discharge.

Pollution from sea birds.

The evidence suggesting seabird faecal pollution is present comes from DNA indicative of a bacterial species demonstrated to be present in high numbers in the seabird gut. The same bacterium has been found in other bird species (that is non-seabird birds) but in far lower numbers (approximately 3 or 4 orders of magnitude difference). The bacterium was originally isolated from carcasses of a grey seal (*Halochoerius grypus*) and a harbour porpoise (*Phocena phocena*) as described by Lawson *et al.*, (2006). Subsequent studies targeting

faeces of these, and related species, failed to detect the marker bacterium, leading to suggestions that the two animal carcasses were contaminated with gull faeces or that the bacterium is an opportunistic pathogen (Lu *et al.*, 2008; Green *et al.*, 2012). If, as seems highly likely, the bacterium is an effective indicator of microbiological pollution from seabirds, the faeces will be untreated, and either deposited directly into the water, or washed from nearby surfaces into the water.

Pollution from dogs.

The evidence suggesting faecal pollution from dogs comes from the detection of mitochondrial DNA from dogs in the South Bay samples. This reflects DNA from dogs, rather than DNA from bacteria found in the dog gut. Dog DNA is present in dog faeces (and associated microbiological pollution) but could also come from dog skin cells from a dog swimming in the water, without the associated microbiological pollution. Experience within the microbiology and molecular biology section of the National Laboratory Service (Starcross) is that mitochondrial DNA exists in far lower concentrations in bathing water samples than the DNA from the bacteria used to assess water quality. The evidence is therefore that dog DNA arrived in the water directly from skin, or indirectly from faeces; whichever route, dogs were present at, or near to, the sampling point before samples were collected.

No data were available on the number of dogs or seabirds/gulls per day at different locations on the beach for the duration of the sampling for this work. Anecdotally, dogs have been observed on the beach at South Bay (e.g. 10 dogs were observed by an Environment Agency employee close to the sampling point in July 2017). Additionally, some data on numbers of birds are available from informal surveys carried out in 2015. Of the bird species recorded in the surveys, herring gulls and kittiwakes were numerically the most abundant at South Bay, in the range 30 to >130 individuals of both species present at different times. Anecdotally again, opportunistic feeding by gulls may be observed around the beaches in Scarborough, with a small sub-population of birds attracted to, and actively scavenging, from people. The following scenario thus attempts to compare the relative importance of the different sources (human, seabird, dog), and to estimate whether any of these sources as sole contributor could account for the measured FIO concentrations at Scarborough South Bay for the period of this study.

Incorporating other published data.

A review of the literature was undertaken to survey the numbers/concentrations of FIO found in different animal faeces. The relative abundances of the different MST markers and associated FIO in known samples was also recorded where available. Some data were available from this report for the seabird marker and FIO from a sample collected from Scarborough South Bay. These data were used as the average values for this section of the overall dataset. Data available from the MST testing service carried out by the NLS was included for the human-specific MST marker. The majority of published work provided concentrations of FIO per gram wet or dry weight animal faeces; relatively few studies provided data on the actual mass of material egested by individual faecal events, or on direct measurements of the daily output of FIO per animal. The data were then summarised in two ways; using direct measurements of FIO (per animal per day or per faecal event) and using calculated values from the literature to derive measurements of FIO (per animal per day or per faecal event). The two approaches gave comparable estimates. The full data and sources are provided in Appendix 3, with the final illustrative data only being presented here. The lowest and highest reported values of FIO were included with average values, to provide an estimate of the range of FIO that may be encountered in different situations.

The profile of the beach at Scarborough South Bay was based on the report of Hick (2017). This report forms part of the series of reports monitoring changes in the coastline of the NE of England undertaken by Royal HaskoningDHV. The beach at Scarborough South Bay has a very gentle slope; at least within the survey area.

It was then necessary to use the data together with the beach profile to provide some context for Scarborough South Bay. Given the shallow profile, a 1000 m^3 volume of water can be used to describe part of the bathing water, 100 m long by 10 m out by 1m depth. Bacterial concentrations for bathing beach monitoring and reporting are in bacteria per 100 ml; 1000 m^3 contains ten million 100 ml volumes.

Although data were collected and/or derived to produce estimates of FIO per faecal event and per animal per day, further limits were placed on the data to help illustrate the relative importance of the different sources. Although one faecal event per dog per visit to the beach was considered as one assumption, it was recognised that this could be an overestimate; also that at least some dog owners would clean up after their dog. Overall, therefore, the assumption was made of 0.1 faecal event per dog per day.

Gulls are known to spend time both on and over the water at Scarborough South Bay, making a direct link between gull faecal events and water pollution. However, gulls are clearly highly mobile, therefore an assumption of half the faecal load per gull per day entering the water (either directly or via surface run-off) was made.

The speculative nature of this section of the report must be reiterated again at this point; several assumptions have been listed above, any or all of which may be considered unrealistic. Numerous other criticisms may also be justified. The highly variable nature of the microbiological composition of all faecal material must be acknowledged, along with the many and varied routes by which bacteria from the faeces may end up in the bathing water. The effect of environmental variables on the survival of those bacteria, and therefore their subsequent detection using the standard methods, also severely limits this illustration.

Within these constraints, estimates of the numbers of FIO that could come from dogs or gulls were produced (Table 7). The variability within these estimates is immediately apparent covering up to 6 orders of magnitude in bacterial numbers (Table 7). Regardless of the limitations of the data and the assumptions which have been made, the potential for either source to have a significant impact on the microbiological quality of the bathing water is very clear (Table 7). Additionally, these estimates are based on a uniform distribution of the bacteria throughout the 1000 m^3 volumes; a more clustered distribution, which may be argued to be more realistic, could readily exceed these values (or, conversely, mean the pollution is missed entirely when sampling).

	FC/*E. coli*			FS/IE		
	Low	Mean	High	Low	Mean	High
Dog	2	46	233	0.015	73	485
Gull	0.06	1850	25000	0.011	80	600

Table 7. Estimated (derived) values of the number of bacteria/100 ml that would be found in a 100m length x 10m width x 1m depth volume of water at Scarborough South Bay from dogs and gulls under the assumptions of 0.1 faecal event per dog per day and 50% of the faecal output of one gull per day.
FC/*E. coli* –different published studies used different methods of enumerating bacteria; for this report, faecal coliforms have been assumed to be an accurate estimator of *E. coli*.
FS/IE - different published studies used different methods of enumerating bacteria; for this report, faecal streptococci have been assumed to be the equivalent of intestinal enterococci.
Low – estimated using the lowest reported numbers from the literature survey
Mean – estimated using the average of the reported numbers from the literature survey
High – estimated using the highest reported numbers from the literature survey

The estimates presented in Table 7 are for dogs and gulls only; despite extensive data on the concentrations of FIO in both raw and treated sewage (including UK-specific measurements; Kay *et al.*, 2008 a and b) it was considered difficult to incorporate human inputs into these estimates using similar assumptions (proportion of a faecal event, or proportion total faecal events per day). However, data were available on concentrations of the human-specific MST marker and FIO in bathing beaches around England, compiled as part of the original validation data for the use of MST by the Environment Agency. These data were numerically ranked according the concentrations of the different MST markers, and a dataset produced including only samples for which the ruminant-specific marker was absent (that is, no detectable faecal pollution from ruminant animals as was the case for Scarborough South Bay in this study) and for which the human-specific marker was present (that is, at least some human faecal pollution was present). The geometric mean values for the human MST marker and for both FIO concentrations were used to estimate average values; with the lower and upper quartile values used to estimate low and high values.

Estimates of the numbers of the seabird marker to numbers of FIO were compiled from the measured concentrations taken at Scarborough South Bay to use as the average values; other data taken from published studies were used to calculate low and high range values for seabird marker concentrations and FIO (full data set and references in Appendix 3).

These values were used together with the concentrations of FIO measured in the Scarborough South Bay samples (Table 3) to assess whether the measured concentrations of FIO could have come from a single source only; or whether at least one other source of pollution would be required. Results are presented in Table 8.

Estimate of source-specific FIO numbers	Sample date[a]	*E. coli*			IE		
		Human	Seabird	Dog	Human	Seabird	Dog
Average							
	05/08	Sole source	Other	Other	Other	Other	Other
	08/08	Other	Other	Sole source	Other	Other	Sole source
	11/08	Other	Other	Sole source	Other	Other	Sole source
	16/08	Other	Other	Sole source	Other	Other	Other
	19/08[b]	Other	Other	Other	Other	Other	Other
	19/08[c]	Other	Other	Other	Other	Other	Other
	23/08	Other	Other	Other	Other	Other	Other
	08/09	Other	Other	Sole source	Other	Other	Sole source
% other sources		87.5	100	50	100	100	62.5
Low							
	05/08	Other	Other	Other	Other	Other	Other
	08/08	Other	Other	Other	Other	Other	Other
	11/08	Other	Other	Other	Other	Other	Other
	16/08	Other	Other	Other	Other	Other	Other
	19/08[a]	Other	Other	Other	Other	Other	Other
	19/08[b]	Other	Other	Other	Other	Other	Other
	23/08	Other	Other	Other	Other	Other	Other
	08/09	Other	Other	Other	Other	Other	Other
% other sources		100	100	100	100	100	100
High							
	05/08	Sole source	Sole source	Sole source	Sole source	Sole source	Sole source
	08/08	Other	Sole source	Sole source	Other	Sole source	Sole source
	11/08	Other	Sole source	Sole source	Other	Sole source	Sole source
	16/08	Sole source	Sole source	Sole source	Other	Sole source	Sole source
	19/08[a]	Other	Sole source	Sole source	Other	Sole source	Sole source
	19/08[b]	Other	Sole source	Sole source	Other	Sole source	Sole source
	23/08	Other	Sole source	Other	Other	Sole source	Other
	08/09	Sole source	Sole source	Sole source	Sole source	Sole source	Sole source
% other sources		62.5	0	12.5	75	0	12.5

Table 8. Estimation of the frequency of multiple sources of faecal pollution contributing to the measured concentrations of FIO at Scarborough South Bay.

[a]All samples were collected in 2016

[b]undisturbed sample, taken according to standard bathing water sampling practice.

[c]disturbed sample; after the standard sample was collected, the sand and sediment was deliberately disturbed before another sample was taken. Data from both undisturbed and disturbed samples are included in the % other sources.

% other sources; the frequency at which at least one other source of pollution would be required to produce the concentrations of FIO measured in the 8 samples from Scarborough South Bay.

Data presented in Table 8 are only able to illustrate the frequency at which single sources of pollution (if dispersed throughout the 1000 m^3 volume of water) could have accounted for the measured bacterial concentrations. Estimations based on average FIO numbers, which must be assumed to represent the typical situation, would strongly suggest that for both *E. coli* and IE, it may be rare for faecal pollution from either human or seabird sources to contribute sufficient numbers of bacteria to be considered the sole source of the FIO (Table 8). Faecal pollution from dogs may be sufficient to be a sole source more frequently (Table 8) but even so, at least one other source seems likely to be required over 50% of the time.

Estimations using the lowest estimate of FIO numbers shows that more than one source would be required for all sampling times, for both *E. coli* and IE, as may be expected (Table 8). Estimations using high FIO numbers may be of more interest, however, and suggest that even when high FIO from human sources are present, other sources are still likely to be required to account for the measured FIO. Using high estimates for seabird FIO numbers could account for the total FIO concentrations measured on every sampling occasion; with high estimates for dog FIO numbers also able to account for the measured concentrations in the majority of samples (Table 8).

Culturable FIO counts (Figure 2) suggest that concentrations of IE exceeded those of *E. coli* in Scarborough South Bay for the first 5 sets of samples, but not for the remainder. Concentrations of IE consistently exceeded those of *E. coli* in the industrial effluent (Figure 2). Several authors have reported that dog and bird faeces (pigeon and gull) are comparatively high in enterococci (Wright *et al.*, 2009; Kelty *et al.*, 2012; Ervin *et al.*, 2013), although this may vary with location (Fogarty *et al.*, 2003). Caution may therefore be needed when considering the higher IE (over *E. coli*) concentrations of the industrial effluent to explain higher IE concentrations in the bathing water.

Gulls have been cited in other reports as significant contributors of microbiological pollution in recreational waters. Removing or controlling gull numbers (using trained dogs) has been demonstrated to reduce IE concentrations (Converse *et al.*, 2012). Use of trained dogs at a site on Lake Michigan (freshwater) reduced average gull numbers from 665 to 17, with significant improvements in water quality. Greater reductions in concentrations of IE than *E. coli* were noted (Converse *et al.*, 2012). It may also be relevant to note that numbers of enterococci may vary greatly throughout a day; generally highest in the morning and declining through the day (Converse *et al*., 2012, freshwater; Boehm, 2007, marine water).

For the three sources identified as causing faecal pollution at Scarborough South Bay (human, seabird, dog), when more than a single source is likely to be involved in the total FIO pollution, it is not possible from this data to say whether just one, or both of the other sources would be required to account for the measured concentrations.. It must also be remembered that bacteria from the industrial effluent which may be present at Scarborough South Bay cannot be readily assessed using MST, thus further complicating the picture.

Conclusions.

The aims of this study were to determine:
i. Is an industrial effluent a factor in the intestinal enterococci (IE) and *Escherichia coli* counts at the designated sampling point in South Bay?
ii. If so, how much of an impact does this industrial effluent have?
iii. Are there other factors, and how much do they contribute?

The following conclusions have been drawn from this study:

1. The MST and FIO data give a strong indication that pollution from bird faeces should be considered as a factor contributing to the microbiological pollution at South Bay.

2. The MST and FIO data give a strong indication that pollution from humans should be considered as a factor contributing to the microbiological pollution at South Bay.

3. The MST data showed that pollution from ruminant animals was not detected in these samples. The MST data suggest that pollution from donkeys was not a significant source of FIO in these samples.

4. A marker specific to dogs was detected during the sampling period, during the period of the dog ban.

5. The MST and FIO data show a very large number of MST gene copies and of IE may come from bird faeces. At the concentrations estimated in this study, 1 g of bird faeces could be polluting approximately 80 litres of bathing water.

6. High numbers of IE and *E. coli* are present in the industrial effluent on site.

7. Associations and patterns in the FIO, MST and NGS data were only present when data were examined by site; not by time.

8. The impact, if any, of the IE from the industrial effluent cannot be easily measured using MST and qPCR, as no specific marker for this source is available.

9. The NGS data, assessing the numerically abundant, non-culturable bacterial communities from the sites sampled in this study, show a large proportion of Firmicute bacteria within the industrial effluent community. These bacteria may persist differently than the Proteobacteria that dominate the wastewater treatment and the potato store communities.

10. The NGS data only showed associations within the data when analysed by site; not by time.

11. The NGS data suggest that the bacterial communities in the industrial effluent, the potato store and the wastewater treatment works are generally discrete from the seawater community. However, there is evidence that the bacterial community from the industrial effluent can impact on the bacterial community at South Bay. This estimated to happen 10 – 20% of the time from this (limited) data set.

12. The rep-PCR data, which profile individual IE isolates from the different sites, showed very few exact duplicates; the few pairs of duplicate profiles that were obtained did not suggest a direct link between IE from either effluent source being present in the South Bay.

13. The rep-PCR data suggest that it is common for the IE diversity to be more similar within a site than between sites (that is, the IE from bird faeces are in general more similar to other IE from bird faeces than to IE from other sources). However, the total IE diversity was large, and some isolates were more similar to others at different sites.

14. The chlorpropham data do not support a direct link between the industrial effluent and South Bay water, though this may be due to dilution of the compound.

15. Incorporating possible bacterial contributions derived from a literature survey into several assumptions stated in the report allowed speculation about the relative contribution of FIO from different sources as determined by MST. Any bacterial contribution from the industrial effluent is missed by the MST approach.

Within the limits of the assumptions, estimations based on average FIO numbers suggest that it is unlikely either human or seabird faecal pollution alone contribute sufficient numbers of FIO to be considered the sole source of the measured FIO at Scarborough South Bay. Under some circumstances, it is possible that faecal pollution from dogs could account for the bacterial concentrations measured at the South Bay sampling point. However, overall, it seems most likely that multiple sources of pollution all contribute to the issue.

References.

Alderisio, K.A. and N. DeLuca. 1999. Seasonal enumeration of fecal coliform bacteria from the feces of ring billed gulls (*Larus delawarensis*) and Canada geese (*Branta canadensis*). Appl. Environ. Microbiol. 65:5628-5630.

Anderson, SA, S.J. Turner and G.D. Lewis. 1997. Enterococci in the New Zealand environment: Implications for water quality monitoring. Wat. Sci. Technol. 35:325-311.

Bernhard, A. E., Field, K. G., 2000. A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroidetes-Prevotella* genes encoding 16S rRNA. Appl. Environ. Microbiol. 66:4571-4574.

Boehm, A.B. 2007. Enterococci concentrations in diverse coastal environments exhibit extreme variability. Env. Sci. Technol. 41:8227-8223.

Caporaso, J.G., C.L. Lauber, W.A. Walters, D. Berg-Lyons, C.A. Lozupone, P.J. Turnbaugh, N. Fierer, and R. Knight. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. USA 108:4516-4522.

Converse, R.R., J.L. Kinzelman, E.A. Sams, E. Hudgens, A.P. Dufour, H. Ryu, J.W. Santo-Domingo, C.A. Kelty, O.C. Shanks, S.D. Siefring, R.A. Haugland and T.J. Wade. 2012. Dramatic improvements in beach water quality following gull removal. Env. Sci. Technol. 46:10206-10213.

Cox,P., M. Griffith, M. Angles, D. Deere and C. Ferguson. 2005. Concentrations of pathogens and indicators in animal feces in the Sydney watershed. Appl. Environ. Microbiol. 71:5929–5934.

Devane. M. and B. Gilpin. 2015. Human health risks of faecal pollution from different sources: A review of the literature. Institute of Environmental Science and Research Limited (New Zealand). 35 pages.

Ervin, J.S., T.L. Russell, B.A. Layton, K.M. Yamahara, D. Wang, L.M. Sassoubre, Y. Cao, C.A. Kelty, M. Sivaganesan, A.B. Boehm, P.A. Holden, S.B. Weisberg and O.C. Shanks. 2005. Characterization of fecal concentrations in human and other animal sources by physical, culture-based, and quantitative real-time PCR methods. Wat. Res. 47:6873-6882.

Felsenstein, J. 2013. PHYLIP. Phylogeny Inference Package version 3.695. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle. http://evolution.genetics.washington.edu/phylip/

Fogarty, L.R., Haack, S.K., Wolcott, M.J. and R.L. Whitman. 2003. Abundance and characteristics of the recreational water quality indicator bacteria *Escherichia coli* and enterococci in gull faeces. J. Appl. Microbiol. 94:865-878.

Geldreich, E.E. 1966. Sanitary significance of fecal coliforms in the environment. Robert A. Taft Sanitary Engineering Center, Cincinnati, Ohio.

Gould, D.J. and M.R. Fletcher. 1978. Gull droppings and their effects on water quality. Wat. Res. 12:665-672.

Green, H.C., L.K. Dick, B. Gilpin, M. Samadpour and K.G. Field. 2012. Genetic markers for rapid PCR-based identification of gull, Canada goose, duck, and chicken fecal contamination in water. Appl. Environ. Microbiol. 78:503-510.

Haack, S., L.R. Fogarty and C. Wright. 2003. Escherichia coli and enterococci at beaches in the Grand Traverse Bay, Lake Michigan: Sources, characteristics, and environmental pathways. Environ. Sci. Technol. 37:3275-3282.

Harwood, V.J., C. Staley, B.D. Badgley, K. Borges and A. Korajkic. 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.

Hick, E. 2017. Cell 1 Regional Coastal Monitoring Programme Update Report 9 Partial Measures Survey for Scarborough Borough Council Royal HaskoningDHV Last accessed (16th August 2017) at: http://www.northeastcoastalobservatory.org.uk/data/Reports/index.php

Hill, R.C., C.F. Burrows, G.W. Ellison, M.D. Finke, J.L. Huntington and J.E. Bauer. 2011. Water content of faeces is higher in the afternoon than in the morning in morning-fed dogs fed diets containing texturised vegetable protein from soya. British Journal of Nutrition 106:S202–S205.

Jones, K. 2002. UK bathing waters: a success story, but… there may be trouble ahead. Microbiology Today 29:186-188.

Kay, D., J. Crowther, C.M. Stapleton, M.D. Wyer, L. Fewtrell, S. Anthony, M. Bradford, A. Edwards, C.A. Francis, M. Hopkins, C. Kay, A.T. McDonald, J. Watkins and J. Wilkinson. 2008a. Faecal indicator organism concentrations and catchment export coefficients in the UK. Wat. Res. 42:2649-2661.

Kay, D., J. Crowther, C.M. Stapleton, M.D. Wyer, L. Fewtrell, A. Edwards, C.A. Francis, A.T. McDonald, J. Watkins and J. Wilkinson. 2008b. Faecal indicator organism concentrations in sewage and treated effluents. Wat. Res. 42:442-454.

Kelty, C.A., M. Varma, M. Sivaganesan, R.A. Haugland and O.C. Shanks. 2012. Distribution of genetic marker concentrations for fecal indicator in sewage and animal feces. Appl. Environ. Microbiol. 78:4225-4232.

Kinzelman, J., S.L. McLellan, A.D. Daniels, S. Cashin, A. Singh, S. Gradus and R. Bagley. 2004. Non-point source pollution: Determination of replication versus persistence of *Escherichia coli* in surface water and sediments with correlation of levels to readily measurable environmental parameters. Journal Water and Health 2:103-114.

Kitts, C.L., M.W. Black, M.M. Moline, M. Yeung, A.K. Hamrick, I.C. Robbins, A.A. Schaffner and N.I. Boutet. 2010. Pismo beach fecal contamination source identification study. Final Report. 116 pages.

Koskey, A.M., J.C. Fisher, M.F. Traudt, R.J. Newton and S.L. McLellan. 2014. Analysis of the gull fecal microbial community reveals the dominance of *Catellicoccus marimammalium* in relation to culturable enterococci. Appl. Environ. Microbiol. 80:757-765.

Lawson, P.A., M.D. Collins, E. Falsen and G. Foster. 2006. *Catellicoccus marimammalium* gen. nov., sp. nov., a novel Gram-positive, catalase-negative, coccus-shaped bacterium from porpoise and grey seal. Int. J. Syst. Evol. Microbiol. 56:429-432.

Lee, C., J.W. Marion, J. Lee. 2013. Development and application of a quantitative PCR assay targeting *Catellicoccus marimammalium* for assessing gull-associated fecal contamination at Lake Erie beaches. Sci. Total. Environ. 454–455:1–8.

Lu, J., J.W. Santo Domingo, R. Lamendella, T. Edge and S. Hill. 2008. Phylogenetic diversity and molecular detection of bacteria in gull feces. Appl. Environ. Microbiol. 74:3969-3976.

Muyzer, G., E.C. de Waal and A.G. Uitterlinden. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. 59:695-700.

Pommepuy, M., D. Hervio-Heath, M.P. Caprais, M. Gourmelon, J.C. Le Saux, and F. Le Guyader. 2006. Fecal contamination in coastal areas:An engineering approach. In:Oceans and health:Pathogens in the marine environment. Belkin and Colwell (eds.) Springer, New York.

Rambaut, A. 2015. FigTree version 1.4.2. http://tree.bio.ed.ac.uk/software/figtree/

Reischer, G. H., Kasper, D. C., Steinborn, R., Mach, R. L., Farnleitner, A. H., 2006. Quantitative PCR method for sensitive detection of ruminant fecal pollution in freshwater and evaluation of this method in alpine karstic regions. Applied and Environmental Microbiology, 72:5610-5614.

Ronquist, F., M. Teslenko, P. van der Mark, D. Ayres, A. Darling, S. Hohna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61:539-542.

Ryu, H., J.F. Griffith, I.U.H. Khan, S. Hill, T.A. Edge, C. Toledo-Hernandez, J. Gonzalez-Nieves and J. Santo Domingo. 2012. Targeting the 16S rRNA genes of *Catellicoccus marimammalium* and *Streptococcus* spp. Appl. Environ. Microbiol. 78:1909-1916.

Stapleton, C.M., D. Kay,, M.D. Wyer, C. Davies, J. Watkins, C. Kay, A.T. McDonald, J. Porter and A. Gawler. 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. Water Research 43:4888-4899.

Stewart, J.R., A.B. Boehm, E.A. Dubinsky, T. Fong, K.D. Goodwin, J.F. Griffith, R.T. Noble, O.C. Shanks, K. Vijayavel and S.B. Weisberg. 2013. Recommendations following a multi-laboratory comparison of microbial source tracking methods. Wat. Res. 47:6829-6838.

Versalovic, J., M. Schneider, F.J. de Bruin and J.R. Lupski. 1994. Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. Methods in Molecular and Cellular Biology 5:25-40.

Whitman, R.L., D.A. Shively, H. Pawlik, M.B. Nevers and M.N. Byappanahalli. 2003. Occurrence of *Escherichia coli* and enterococci in *Cladophora* (Chlorophyta) in nearshore water and beach sand of Lake Michigan. Appl. Environ. Microbiol. 69:4714–4719.

Wright, M.E., H.M. Solo-Gabriele, S. Elmir and L.E. Fleming. 2009. Microbial load from animal feces at a recreational beach. Marine Pollution Bulletin 58:1649–1656.

Zhou,J., Z. He, Y. Yang, Y. Deng, S.G. Tringe and L. Alvarez-Cohen. 2015. High-throughput metagenomic technologies for complex microbial community analysis: Open and closed formats. mBio 6:e02288-14. doi:10.1128/mBio.02288-14.

Appendix 1.

Details of the analytical methods and data analysis used in this study.

FIO enumeration.

Water samples were processed for *E. coli* and IE using the standard approach used for all regulatory bathing water samples in England. Briefly, sample volumes ranging from 0.1 to 10 ml were captured on 0.45 µm cellulose nitrate/cellulose acetate filters. For *E. coli*, the filters were placed on TBX agar plates (Merck) and incubated at 30 ºC for 4 h, followed by 44 ºC for a further 14 h. After incubation, the numbers of blue colonies (confirmed *E. coli*) were recorded. For IE, the filters were placed on Slanetz and Bartley agar (Oxoid), and incubated at 30 ºC for 4 h, followed by 44 ºC for a further 40 h. After incubation, the numbers of red colonies (presumptive IE) were recorded. Where appropriate, a sub-set of up to ten colonies was then further sub-cultured and tested for aesculin hydrolysis using kanamycin aesculin agar (with incubation at 44 ºC for 21h). The proportion testing positive was used to produce the final confirmed IE count. For solid samples, 10 g wet weight material was made up to 100 ml with sterile quarter-strength Ringer's solution, and shaken vigorously with 5g sterile Ballotini beads. A range of volumes of this suspension was captured on the filters before incubation.

Sample processing for microbial source tracking.

After filtration of water samples for FIO enumeration, the remainder of the sample (up to a maximum of one litre volume; exact volumes were determined by weighing the sample bottle before and after) was passed through a filter prior to storage at -20 ºC prior to DNA extraction.
DNA was extracted from those samples showing the highest concentrations of the FIO. Homogenisation and physical sample lysis was performed using the Fast Prep 24 instrument (MP Biomedicals). DNA was extracted using the Fast DNA Spin kit for Soil (MP Biomedicals Europe) according to the manufacturer's instructions. Purified DNA was eluted in 0.1 ml buffer before qPCR amplification of target genes.

A total of five different qPCR assays were used in this study. All amplifications were performed singly (i.e. no multiplexing) for each assay. Two markers were used for MST analysis of total DNA using *Bacteroides*. The human-specific HF8 sequence cluster (HuBac) identified by Bernhard and Field (2000) was assayed as described in Stapleton *et al.,* (2009) and the BacR assay of Reischer *et al.,* (2006) was used to enumerate ruminant-specific pollution (RuBac). Pollution from dogs was assessed by targeting mitochondrial DNA using in-house qPCR assays. Seabird pollution was assayed broadly as described (Green *et al.,* 2012; Ryu *et al.*, 2012; Lee *et al.,* 2013).

All qPCR assays were performed using TaqMan Universal PCR master mix (Life Technologies) with the UNG carry-over control system. Cycling conditions were 900 nM each primer, 50 nM probe and one to five µl template in a final reaction volume of 25 µl. Cycling conditions were 10 minutes at 50 ºC (UNG activity), 10 minutes at 95 ºC (initial denaturation) followed by 40 cycles of 15 seconds at 95 ºC/1 minute hybridisation-extension at assay-specific temperatures. Reporter dye fluorescence was read at the end of each extension step. Amplifications were performed using a Stratagene MxPro 3005 thermal cycler. Appropriate controls (laboratory blanks, reagent blanks, negative blanks, positive controls and quantification controls) were included. Quantification controls were produced from consensus sequences of the genes from the nucleic acid sequence databases (accessed through EMBL) and synthesised as Ultramers™ (Integrated DNA Technologies, Iowa). Data were assessed and exported using the MxPro software version 4.10.

Sample processing for next-generation sequencing.

Aliquots of DNA extracted for the qPCR work were also used for bacterial community profiling using a next-generation sequencing approach following PCR amplification of the 16S rRNA gene from the bacterial communities using primers 341F (5'- CCTACGGGAGGCAGCAG -3'; Muyzer *et al.*, 1993) and 806R (5'-

GGACTACHVGGGTWTCTAAT -3'; Caporaso *et al.*, 2011) using a MiSeq (Illumina) NGS instrument. After sequencing, data were quality filtered, aligned against the QIIME (Green Genes) reference database, subsampled, denoised, and chimeras removed. Operational taxonomic units (OTU) were picked by clustering of aligned sequences at a range of identity levels, with subsequent alpha and beta diversity analyses carried out using non-metric multidimensional scaling.

Sample processing for repetitive element PCR of IE isolates.

Up to 40 isolates per sample were from individually sub-cultured from the original Slanetz and Bartley plates and incubated at 44 ºC for a further 24 to 48 h. Following sub-culture, isolates were picked into individual wells of 96 well plates containing 200 µl volumes of buffered peptone water. Isolates were stored in this way at 4ºC. For processing, the wells were aspirated using a pipette and 150 µl volumes were added to 5 ml volumes of nutrient broth in sterile 15 ml centrifugation tubes. Following incubation at 37 ºC for 24-48h, the tubes were centrifuged and the resulting pellet were resuspended in 1 ml TE buffer before 10 µg lysozyme (Sigma Aldrich) was added and the tubes incubated at 37 ºC for 10 min, followed by placement in a boiling bath for 10 minutes. The tubes were centrifuged one further time before the supernatant was used for PCR.
Isolates were profiled using repetitive element polymerase chain reaction (rep-PCR) and the BOXA1R primer (Versalovic *et al.,* 1994). Amplification used VWR Chemicals *Taq* polymerase, standard Key buffer and dNTPs (VWR) and oligonucleotides from Integrated DNA Technologies (IDT).

Following rep-PCR, amplification products were electrophoresed at 1.5V per cm. through 1.5% wt/vol agarose gels, stained with SYBR Safe (Life Technologies) and an Amresco 100 bp molecular size ladder covering 3000 to 100 bp (VWR). Gel images were captured under U.V. illumination and saved as .TIFF files. Images were read into the MYImageAnalysis software (2014 release, ThermoFisher). Gels were analysed automatically as far as possible, using the functions within the software, with each band size estimated from the molecular size standards (one standard was run at each side of every gel). Data were exported into Microsoft Excel. Replicate positive controls using DNA from *Enterococcus faecalis* were included in every amplification and gel (alongside replicated negative controls).

A profile (presence/absence of a rep-PCR product across all sizes found throughout the work) was produced for each isolate using Excel and rounded to the nearest 20 bp before producing a binary profile. PCR product smaller than 100 bp were removed from the dataset. Very few isolates produced PCR products greater than 3500 bp and these data points were also removed. Finally, isolates that had produced less than 4 bands across the range 100 to 3500 bp (in 20 bp increments) were removed from the dataset.

Binary profiles (1=present; 0=absent) were then available for each remaining isolate. Isolates were assigned a unique code, reflecting site, date and isolate number for that sample, with the resulting profiles analysed for diversity estimates.

Data processing and analysis.

Bacterial numbers and gene copy numbers (HuBac and RuBac assays) were expressed per 100 ml of original sample volume. *E. coli* and IE colony counts that were less than or greater than values were used at that value (e.g. <10cfu/100 ml was treated as 10 cfu/100 ml). Gene copy number (qPCR) estimates below the limits of detection were treated as zero or were excluded as appropriate when manipulating data. Standard correlation analysis (both parametric and ranked) were performed in Microsoft Excel. Bacterial community comparisons were carried out on NGS data using the GreenGenes database, through the Illumina BaseSpace software and within the R environment. IE colony diversity was carried out by comparison of binary profiles. Sorensen-Dice similarity coefficients were calculated using Excel. Profiles were compared directly for exact matches; and for relatedness using parsimony and distance methods within the Phylip 3.695 set of packages (Felsenstein, 2013). Data were sub-sampled and consensus trees produced as appropriate. Relatedness among isolates was also assessed using the same binary profiles through a Bayesian (Markov Chain Monte Carlo) approach using MrBayes 3.2.6 (Ronquist *et al*., 2012). Trees were produced using FigTree 1.4.2 (Rambaut, 2012).

Chemical analysis.

Samples were also collected for chemical analysis according to standard operating procedures. Parameters analysed included dry solids @ 30°C, Chlorpropham (µg/l), GCMS Screen: Semi-Volatile (unitless).

Appendix 2.

Table A2.1. Sample details and associated data.

Date	Time	Matrix[a]	Notes[b]	Sample point[c]	*E. coli*	IE	MST[d]				NGS[e]
							Human	Seabird	Ruminant	Dog	
05/08	11:16	Water		Industrial effluent	6.10	7.78	3.45	2.08	ND	ND[f]	Y
05/08	11:20	Water		Raw process water	<1	<1					
05/08	11:40	Water		Whitby Rd PS	2.61	2.23	4.07	3.65	ND	ND	Y
05/08	12:25	Water	Undist	North Bay	1.73	<1					
05/08	12:27	Water	Dist	North Bay	1.26	<1					
05/08	12:41	Water		WwTW effluent	3.61	2.66	5.43	2.14	ND	ND	Y
05/08	12:57	Water	Undist	South Bay	1.80	2.53	3.78	5.19	ND	ND	Y
05/08	13:05	Water	Dist	South Bay	2.03	2.28					
05/08	13:40	Water		South Cliff	<1	<1	3.36	3.75	ND		Y
08/08	09:50	Water		Industrial effluent	6.34	9.48	5.32	ND	ND	ND	Y
08/08	09:55	Water		Raw process water	<1	<1					
08/08	10:00	Soil		Potato Store	4.21	5.65	ND	2.70	ND		Y
08/08	10:47	Water		South Cliff	<1	1.43	2.58	4.17	ND		Y
08/08	11:57	Water		Whitby Rd PS	<1	1.65	2.13	3.44	ND		Y
08/08	12:55	Sed		North Bay	<1	<1					
08/08	12:55	Water	Undist	North Bay	<1	<1					
08/08	12:55	Water	Dist	North Bay	<1	<1					
08/08	13:15	Sed		South Bay	<1	1.08					
08/08	13:15	Water	Dist	South Bay	<1	<1					
08/08	13:15	Water	Undist	South Bay	<1	1.43	ND	4.45	ND	ND	Y
11/08	12:20	Sed		South Cliff	<1	<1					
11/08	12:21	Water		South Cliff	<1	<1					
11/08	13:00	Sed		South Bay	<1	1.91					
11/08	13:00	Water		Industrial effluent	6.16	9.30	4.56	ND	ND	ND	Y
11/08	13:00	Water		South Bay	1.26	1.56	ND	5.03	ND	1.83	Y
11/08	13:05	Water		Raw process water	1.73	3.40	ND	ND	ND		Y
11/08	13:30	Water		North Bay	<1	<1					
11/08	13:31	Sed		North Bay	<1	<1					
11/08	13:50	Soil		Potato Store	3.28	5.08	ND	2.91	ND		Y
11/08	14:05	Water		WwTW effluent	4.19	3.76	6.25	3.40	ND		Y
11/08	14:10	Water		Whitby Rd PS	<1	<1					
16/08	11:00	Water		Whitby Rd PS	<1	<1					
16/08	11:05	Water		Industrial effluent	1.95	1.95					
16/08	11:08	Water		Raw process water	<1	<1					
16/08	11:10	Soil		Potato Store	5.00	5.49					
16/08	11:40	Sed		North Bay	<1	<1					
16/08	11:40	Water		North Bay	<1	<1					
16/08	11:55	Sed		South Bay	<1	1.83					
16/08	11:55	Water		South Bay	1.56	1.91	2.66	5.11	ND		Y
16/08	12:15	Water		WwTW effluent	3.88	3.03	5.44	ND	ND		Y
16/08	12:40	Water		South Cliff	<1	<1					
19/08	10:52	Water		Raw process water	<1	<1					
19/08	11:00	Soil		Potato Store	2.95	4.03	ND	ND	ND		Y
19/08	11:05	Water		Industrial effluent	6.46	9.30					
19/08	11:45	Sed		North Bay	<1	<1					
19/08	11:45	Water		North Bay	<1	<1					
19/08	11:45	Water		Whitby Rd PS	1.43	<1					
19/08	12:30	Sed		South Bay	<1	<1					
19/08	12:30	Water	Undist	South Bay	2.03	2.19	2.77	5.00	ND		Y
19/08	12:30	Water	Dist	South Bay	2.03	2.35	2.74	4.85	ND		Y
19/08	12:45	Faeces		Bird faeces	7.17	6.88	ND	11.02	ND		Y
19/08	13:25	Water		North Bay	1.26	<1					

Table A2.1. (cont.) Sample details and associated data.

Date	Time	Matrix[a]	Notes[b]	Sample point[c]	*E. coli*	IE	MST[d]				NGS[e]
							Human	Seabird	Ruminant	Dog	
23/08	10:40	Water		Whitby Rd PS	<1	1.26	3.46	3.38	ND		Y
23/08	11:03	Water		Industrial effluent	6.41	9.23	4.23	ND	ND	ND	Y
23/08	11:05	Water		Raw process water	<1	<1					
23/08	11:07	Soil		Potato Store	5.65	6.23	ND	ND	ND		Y
23/08	11:10	Sed		North Bay	<1	<1					
23/08	11:10	Water		North Bay	<1	<1					
23/08	11:25	Sed		South Bay	<1	<1					
23/08	11:25	Sed		South Bay	<1	<1					
23/08	11:25	Water		South Bay	2.81	2.72	3.20	5.95	ND	ND	Y
23/08	11:55	Water		WwTW effluent	2.70	2.10	5.26	3.24	ND	ND	Y
02/09	11:15	Water		Industrial effluent	1.95	2.26	4.22	2.76	ND		Y
02/09	11:20	Water		Raw process water	<1	1.41	ND	ND	ND		Y
02/09	11:25	Soil		Potato Store	4.34	5.95	ND	ND	ND		Y
02/09	13:00	Water		Whitby Rd PS	<1	<1					
02/09	13:30	Water		WwTW effluent	3.91	3.38	5.86	2.56	ND		Y
02/09	14:15	Water		North Bay	<1	<1					
02/09	14:16	Sed		North Bay	<1	<1					
02/09	14:30	Water		South Bay	1.26	<1					
02/09	14:33	Sed		South Bay	<1	<1					
06/09	10:54	Water		Industrial effluent	1.95	3.07	2.41	ND	ND		Y
06/09	11:01	Water		Raw process water	<1	<1					
06/09	11:04	Water		Whitby Rd PS	1.26	<1					
06/09	11:07	Soil		Potato Store	3.16	4.62	ND	ND	ND		Y
06/09	11:42	Water		North Bay	<1	<1	2.58	3.87	ND		Y
06/09	11:45	Sed		North Bay	<1	<1					
06/09	12:12	Water		South Bay	1.26	<1					
06/09	12:15	Sed		South Bay	<1	<1					
06/09	12:20	Sed		South Bay	<1	<1					
06/09	12:50	Water		WwTW effluent	3.61	2.61	5.11	ND	ND		Y
08/09	10:50	Water		Raw process water	<1	<1					
08/09	10:52	Soil		Potato Store	3.96	5.49	ND	ND	ND		Y
08/09	10:55	Water		Industrial effluent	<1	3.16	2.92	ND	ND	ND	Y
08/09	10:57	Water		Whitby Rd PS	<1	<1					
08/09	11:16	Water		North Bay	<1	<1					
08/09	11:17	Sed		North Bay	<1	<1					
08/09	11:30	Water		WwTW effluent	3.51	2.93	5.30	ND	ND	ND	Y
08/09	11:40	Sed		South Bay	<1	<1					
08/09	11:41	Sed		South Bay	<1	<1					
08/09	11:42	Water		South Bay	1.45	1.04	2.61	4.54	ND	2.28	Y
12/09	09:45	Water		WwTW effluent	4.81	4.28	6.40	2.62	ND		Y
12/09	09:46	Water		Whitby Rd PS	1.26	<1					
12/09	10:19	Water		North Bay	<1	<1					
12/09	10:20	Sed		North Bay	<1	<1					
12/09	10:35	Water		Raw process water	<1	<1					
12/09	10:36	Sed		South Bay	<1	<1					
12/09	10:37	Water		South Bay	1.65	<1					
12/09	10:38	Sed		South Bay	<1	<1					
12/09	10:50	Soil		Potato Store	3.96	4.86	ND	ND	ND		Y
12/09	10:55	Water		Industrial effluent	<1	1.63	2.83	ND	ND		Y

[a] sample matrix, water, sediment or soil. For water matrices, data are $\log_{10}$ count /100 ml volume; for solid matrices, data are $\log_{10}$ count /g wet weight

[b] Undist - undisturbed, sample collected according to standard procedures. Dist – after initial sampling, the sampler deliberately disturbed the sand/sediment and collected a second sample.

[c] further details of sample points is given in main report, Table 1 and Figure 1.
[d] MST – microbial source tracking, for targets specific to an animal host. Data are given as $\log_{10}$ gene copies /100 ml. Human, seabird, ruminant – bacterial MST targets; dog – mitochondrial DNA
[e] NGS – samples having bacterial community diversity assessed using next generation sequencing
[f] ND – not detected

Appendix 3.

Literature survey data from which low, average/intermediate and high estimates of FIO and MST numbers from human, gull and dog faecal wastes were derived.

A1. Concentrations of faecal indicator organisms in dog and seabird (gull) faeces.

Reported concentrations/gram wet weight

Animal	Faecal coliform/*E. coli*			Faecal streptococci/IE			Reference
	Low	Mean	High	Low	Mean	High	
Dog	8.40E+06	3.10E+07	1.20E+08				Cox *et al*., (2005)
Medium-large dog		6.40E+07					Wright *et al*., (2009)
Small dog		5.90E+06					Wright *et al*., (2009)
Dog		2.30E+04					Anderson *et al*., (1997)
Dog		2.40E+06			2.50E+06		Kitts *et al*., (2010)
Dog		2×107			2.50E+08		Pommepuy *et al*., (2006)
Gull (Chicago)	1.60E+06	4.87E+08	1.90E+09	2.00E+04	1.73E+07	6.50E+07	Fogarty *et al*., (2003)
Gull (Traverse City)	1.00E+05	1.38E+07	6.90E+07	1.00E+06	5.73E+07	2.40E+08	Fogarty *et al*., (2003)
Gull (beach)		1.80E+07					Jones (2002)
Gull (landfill)		1.00E+10					Jones (2002)
Gull		1.40E+07			5.00E+07		Haack *et al*., (2003)
Gull		4.00E+06			2.00E+06		Kitts *et al*., (2010)

A2. Wet weight faeces produced by dog and seabird (gull) per day.

Animal	Excretion (g/day)	References
Dog	413	Geldreich (1966) cited in Pommepuy *et al.,* (2006)
Dog (24 kg)	227	Hill *et al.*, (2011)
Gull	11.2	Gould and Fletcher (1978)
Gull	24.9	Gould and Fletcher (1978)
Ring-billed gull		
Autumn	29.16	Alderiso and DeLuca (1999)
Winter	26.04	Alderiso and DeLuca (1999)
Spring	34.5	Alderiso and DeLuca (1999)
Summer	30.08	Alderiso and DeLuca (1999)
Overall average	29.945	Alderiso and DeLuca (1999)

Gull: low, mean, high 0.43, 0.48, 0.53 g per faecal event for gull (Alderiso and DeLuca, 1999)
Dogs: low, mean, high 62, 133.5, 194 g per faecal event for medium dog (Hill *et al.*, 2011)

A3. Ranges/day from literature (reported directly; relatively few studies)

	Faecal coliform/*E. coli*			Faecal streptococci/IE		
	Low	Mean	High	Low	Mean	High
Dog	2.88E+09	3.00E+09	3.30E+09	no data	no data	no data
Gull	3.02E+08	1.29E+09	1.03E+10	2.10E+05	1.38E+08	4.44E+08

Results from Wright et al., (2009) include data from a small dog (3.2 kg body weight) which has been considered here to be an outlier; including this data in these estimates lead to the suggestion of the derived numbers greatly exceeding a directly-measured number (per faecal event). However, it is recognised that dog body sizes vary greatly, and that these estimates, using average weights and counts taken from a number of studies will have the effect of reducing the actual data range.

A4. Typical ranges/day from literature (taken in part from data derived by Devane and Gilpin [2015] using data from a greater number of studies).

	Faecal coliform/*E. coli*			Faecal streptococci/IE		
	Low	Mean	High	Low	Mean	High
Dog	5.45E+08	1.30E+10	4.96E+10	5.22E+06	2.06E+10	1.03E+11
Gull	1.12E+06	3.69E+10	5.00E+11	2.24E+05	1.60E+09	1.20E+10

A5. Assumptions and derivations using above data.

Dogs, may be reasonable to assume one faecal event per visit to the beach, of which 0.1 ends up in the bathing wat.
Gulls, may be reasonable to assume 50% of the faecal events end up in the bathing wat?
(May all change after rainfall.)

Combining datasets under these assumptions gives:

	Faecal coliform/*E. coli*			Faecal streptococci/IE		
	Low	Mean	High	Low	Mean	High
Dog	1.54E+07	4.60E+08	2.33E+09	1.47E+05	7.32E+08	4.85E+09
Gull	5.60E+05	1.85E+10	2.50E+11	1.12E+05	8.00E+08	6.00E+09

Pollution of 1000 cubic metres of wat with the above estimates, and using the assumptions of one-tenth of a faecal event from a dog with 50% of the daily faecal output from a gull dispersing in 1000 m^3 wat, results in the concentrations per 100 ml presented in Table 7 (main report).

A6. MST marker/FIO ratios.

	Faecal coliform/*E. coli*			Faecal streptococci/IE		
	Low	Medium	High	Low	Medium	High
Human	9	50	349	16	76	135
Gull	59	7079	64371	121	13804	64376

Human-specific Bacteroides/FIO from Porter, J. unpublished MST validation data
Gull-specific bacteria/FIO, from Koskey *et al.*, (2014) (Medium values derived from bird faecal sample, this study.)