Molecular Source Tracking Report for Skagit County, Washington

To: **Rick Haley**

From: **Kate Field,** Ph.D.
Department of Microbiology
Oregon State University
Corvallis, OR 97331
kate.field@oregonstate.edu

Attached:

1) Results table: TABLE 1.xls
2) Description of collection sites: TABLE 2: OSU MST site descriptions.xls

Executive Summary

This report summarizes and interprets the results of host directed quantitative polymerase chain reaction (qPCR) analyses of water samples from the area of the Samish River watershed in Skagit County, Washington, and presents the methods used for sample processing. The purpose of the study was to identify the sources of fecal contamination in the water samples. We analyzed DNA extracted from the water samples with assays that target molecular markers in human (2 assays), ruminant (2 assays), dog, avian, gull, horse and dog feces. The assays target gene sequences in fecal bacteria specific to host species or groups; targeted bacteria are not culture-based indicator bacteria (FIB; *E. coli*, fecal and total coliforms, and enterococci), because FIB do not provide enough information to distinguish among hosts.

The entire data set is given in the attached Excel table (Table 1). We found evidence of contamination from human, ruminant, avian, and dog feces. We did not find evidence of gull or horse contamination. The most frequent source of contamination detected was ruminant: 58% of the samples tested positive with at least one ruminant marker. In addition, quantities of ruminant markers were significantly negatively correlated with dissolved oxygen, reflecting the high biological oxygen demand from ruminant feces. The second most frequent source of contamination was avian: 54% of the samples tested positive with the avian assay. Third was human: a total of 17% of the samples tested positive with at least one human assay. We detected dog feces in only one sample.

It is necessary to understand the limitations of the technique in order to avoid over-interpretation or under-interpretation of the results. Sensitivity and specificity of the markers are important; composition of fecal bacteria in hosts changes in relation to diet, temporally, and geographically, so some hosts in a given area may not carry the particular fecal bacteria targeted by the assays. The analysis is also limited by the lack of markers for additional species or groups that may affect water quality, such as rodents and raccoons. Additionally, because naturally occurring compounds, such as humic acids, inhibit qPCR, we diluted inhibited samples to relieve inhibition. In some samples this may have lowered the concentration of the markers below the threshold concentration at which they can be detected.

Although the analysis is quantitative, it can only roughly quantify the relative or proportional contribution from each source of fecal contamination, for several reasons. The different bacterial targets of the assays have different numbers of target genes per bacterial cell. Individual hosts and populations vary in amounts of targeted fecal bacteria and FIB they contain, in relation to diet, temporally, and geographically. This means, for example, that a certain number of copies of a ruminant marker may not equate to the same number of copies of a human marker in terms of contamination. In addition, each bacterial group has a unique survival curve once released from the host into the environment, influencing how long it can be detected. Thus the analyses are best interpreted qualitatively, as either absence, or presence of some, a middling amount, or a lot of contamination from each source.

This study produced strong evidence that ruminants and birds are responsible for a significant part of the microbial contamination found in the water samples analyzed. In addition, as both domestic ruminant and human contamination are likely to be amenable to remediation, the analysis allows specific areas where remediation would be most useful to be identified.

Methods

Overview. DNA was extracted from water samples and added to quantitative polymerase chain reactions (qPCR) using "host-specific" primers or markers. These primers were all developed in my laboratory at Oregon State University (OSU). The primers target ribosomal genes in groups of fecal bacteria that have been shown to be specific to a particular animal host species or group, such as humans or ruminants. If fecal contamination from the host group is present in the original water sample, qPCR will detect the host-specific marker in the DNA that was extracted from the water sample, and will measure the number of copies of the marker in the sample. Controls, quality assurance measures, and knowledge about specificity and sensitivity of the markers allow the results to be interpreted, in combination with data from sanitary surveys and fecal indicator bacteria (FIB).

Sample Processing. We were provided with water samples and sample environmental data, including fecal coliforms, turbidity, and dissolved oxygen. Sites where water samples were collected are described in Table 2 (attached). Five hundred ml environmental water samples were collected from a variety of freshwater sites in and around the Samish River in Washington State, and mailed on ice by overnight express mail to our laboratory at OSU. Within 1hr of receipt by our lab, 5 µl AF504 cryostock (recovery control) was added to 100 ml of each sample and samples were filtered through 0.22 micron pore size filters. Filters were soaked in 700 µL GITC lysis buffer, and stored at -80ºC. All filters and blanks were given unique laboratory sample numbers ("SN", column B in Table 1). DNA was extracted using the PowerWater DNA Isolation kit (MoBio, Carlsbad, CA) according to the manufacturer's directions. After elution, PAO DNA (inhibition control) was added to DNA extracts and they were tested for PCR inhibitors using a separate qPCR assay that targeted PAO. DNA extracts testing positive for inhibition were diluted and retested. Samples were analyzed by qPCR using the markers shown in Table 3. DNA extraction recovery was separately estimated for each sample by qPCR targeting AF504, the recovery control, and results of each host-specific assay were adjusted for recovery.

Table 3. Fecal qPCR assays and their host targets

PCR Assay	Target	Reference
HF134 and HF183	Human	Bernhard and Field, 2000 Walters and Field, 2009
CF128 and CF193	Ruminant	Bernhard and Field, 2000 Walters and Field, 2009
GFD	Avian	Green *et al*., in review
GFC	Gull	Green *et al*., in review
DF155	Dog	Dick and Field, 2005
Ho597	Horse	Dick *et al.,* 2005
GenBac	*Bacteroides* (fecal bacteria)	Bernhard and Field, 2000 Dick and Field, 2004

Controls. *Extraction efficiency:* A full-process control was added to each sample prior to filtration, consisting of a known number of cells of *E. coli* AF504, a strain with a unique genetic insert. We then amplified the AF504 genetic marker from each extracted DNA sample and used the results of this amplification to estimate recovery (as gene copies recovered/gene copies added). Results for each sample were adjusted for recovery. In Table 1, sample recovery is shown in column C ("rec"). Numbers shown for each marker have been adjusted for recovery.

Filtration blanks: for each set of samples, we included 4 filtration blanks (pure water) and these were filtered, extracted and analyzed as above.

Extraction blanks: DNA was extracted from filters in sets of 24, of which 1 was an extraction blank (pure water), which was analyzed as above.

Sample inhibition: Sample inhibition was estimated based on qPCR amplification of the inhibition control, *P. aeruginosa* strain PAO-T7 (PAO). We added a known number of genomes of PAO to each extracted sample DNA and then amplified a PAO genetic marker. If inhibition was observed in a sample (less PAO amplified than expected), samples were diluted. Dilution has been shown by many investigators to relieve inhibition. The dilution factors are shown in Column N in Table 1, and results shown for each marker assay have been adjusted to reflect this dilution factor.

PCR negative controls: qPCR was performed in 96-well plates. For each plate, a minimum of 4 no-template controls was included.

PCR positive controls: a minimum of 2 positive controls consisting of known fecal DNA was included for each qPCR assay for each plate.

Results

The entire data set is shown in Table 1 (attached). Numbers for each assay are marker copies per 100 ml of each marker present in the given sample, measured by the qPCR assay; these numbers have been adjusted for dilution and for recovery. "NA" indicates that the marker was below the limits of detection.

Fecal contamination in samples, by qPCR: Markers for horse and gull feces were not found in any samples. However, markers that indicate the presence of human, ruminant, bird, and dog fecal contamination were found. We detected no markers in 15 samples (22%). Table 5 lists the samples testing positive for each of the markers, ranked as containing relatively low (+; < 20,000 marker copies/100 ml for human and ruminant markers or <7,000 marker copies per 100 ml for the avian marker), medium (++; < 200,000 marker copies/100 ml for human and ruminant markers or <70,000 marker copies per 100 ml for the avian marker), or high (+++) amounts of the markers. The dosage rankings are for convenience, to make it easier to interpret the data. The numbers of marker gene copies per 100 ml in each sample are shown in Table 1.

The most frequent source of contamination detected was ruminant: 58% of the samples tested positive with at least one ruminant fecal marker. The second most frequent source of contamination was avian: 54% of the samples tested positive with the avian feces assay. Third was human: 17% of the samples tested positive with at least one human fecal marker. Only one sample was positive for dog feces.

Table 4. Marker occurrence among all samples that tested positive, ordered from least to most contaminated for each marker. "Total ruminant" and "total human" indicate the total number and percentage of samples testing positive with at least one ruminant or at least one human marker, respectively.

Marker	Source (target)	Number testing positive (n=69)	Percentage testing positive	Positive Sample ID and approximate dosage	
GFD	**Avian**	**37**	**53.6**	061411-23	++
				060811-7	++
				060811-2	++
				060811-3	++
				060811-4	++
				061411-1	++
				061411-10	++
				061411-4	++
				060811-10	++
				060811-10	++

Marker	Source (target)	Number testing positive (n=69)	Percentage testing positive	Positive Sample ID and approximate dosage	
				060811-17	++
				061411-18	++
				061411-17	++
				060811-9	++
				052611-3	++
				052611-2	+++
				061411-6	+++
				061411-8	+++
				060811-11	+++
				060811-16	+++
				060811-24	+++
				052611-4	+++
				061411-19	+++
				060811-8	+++
				052611-10	+++
				052611-7	+++
				052611-5	+++
				052611-19	+++
				052611-16	+++
				031611-3	+++
				052611-18	+++
				052611-13	+++
				052611-23	+++
				052611-11	+++
				052611-21	+++
				052611-15	+++
				052611-8	+++
CF193	Ruminant	38	55	061411-1	+
				061411-8	+
				061411-4	+
				061411-23	+
				061411-17	++
				060811-17	++
				060811-10	++
				060811-7	++
				060811-4	++
				060811-9	++
				061411-15	++
				060811-8	++
				060811-3	++
				052611-3	++
				061411-6	++
				060811-10	++
				052611-19	+++
				061411-19	+++

Marker	Source (target)	Number testing positive (n=69)	Percentage testing positive	Positive Sample ID and approximate dosage	
				052611-4	+++
				061411-3	+++
				060811-23	+++
				060811-11	+++
				052611-10	+++
				052611-13	+++
				052611-15	+++
				052611-8	+++
				052611-7	+++
				052611-5	+++
				052611-18	+++
				052611-23	+++
				031611-3	+++
				060811-13	+++
				060811-16	+++
				060811-21	+++
				052611-16	+++
				052611-21	+++
				031611-1	+++
				061411-21	+++
CF128	Ruminant	8	11.6	060811-3	+
				061411-17	+
				052611-4	+
				061411-3	+
				052611-10	+
				052611-8	+
				052611-16	++
				052611-21	+++
Total ruminant		**40**	**57.8**		
HF134	Human	8	11.6	061411-15	+
				061411-23	+
				060811-10	+
				061411-6	+
				060811-8	++
				052611-8	++
				052611-23	++
				031611-1	+++
HF183	Human	6	8.7	061411-8	+
				061411-6	+
				052611-19	+
				060811-8	+
				052611-15	++
				052611-8	++
Total human		**12**	**17.4**		
DF155	dog	1	1.4		

Marker	Source (target)	Number testing positive (n=69)	Percentage testing positive	Positive Sample ID and approximate dosage	
				061411-3	+++
Ho597	Horse	0	0.0		
GFC	Gull	0	0.0		
GenBac	Fecal *Bacteroides*	50	72.5	(refer to Table 1)	

We detected markers from two species or groups (e.g., ruminant and avian) in 23 samples (33%) and from three species or groups (ruminant, human, avian) in 10 samples (14%) (Table 5).

Table 5. Samples with more than one source of contamination

Source	Sample
Ruminant and Avian	052611-13, 052611-21, 052611-16, 052611-3, 052611-18, 052611-4, 052611-5, 052611-7, 031611-3, 060811-16, 060811-11, 060811-17, 060811-7, 060811-3, 060811-4, 060811-9, 061411-19, 061411-4, 061411-1, 061411-17
Ruminant and Dog	061411-3
Ruminant and Human	031611-1, 061411-15
Avian and Human	
Ruminant, Avian and Human	052611-15, 052611-23, 060811-10, 060811-10, 060811-8, 061411-8, 061411-6, 061411-23, 052611-19, 052611-8

Fifteen samples (22%) were negative for all tested molecular markers for fecal contamination (Table 6). A high fecal coliform count in a negative sample suggested that another source that was not tested could be important. A high dilution factor suggested that dilution may have lowered markers below the minimum threshold concentration necessary for detection.

Table 6. Samples testing negative for all markers

Sample	Coliforms	Sample Dilution Factor
052611-14	1600	1
052611-22	540	1
052611-12	9200	1
031611-2	2	100
052611-17	350	100
052611-6	5400	100
052611-9	3500	10
052611-20	2200	100
060811-19	130	100
060811-1	46	10
031611-3	110	10
060811-22	33	100

Sample	Coliforms	Sample Dilution Factor
031611-4	350	100
031611-5	49	1
061411-20	130	10

Correlations. There were significant or highly significant positive correlations between quantities of the GenBac, human, ruminant and avian molecular markers and number of fecal coliforms present in each sample (Figure 1). Although not all samples with one human marker or one ruminant marker had the other human or ruminant marker, quantities of the two human markers were positively correlated, as were quantities of the two ruminant markers. The quantity of ruminant markers was strongly negatively correlated with sample dissolved oxygen, but this correlation was not seen with human and avian markers.

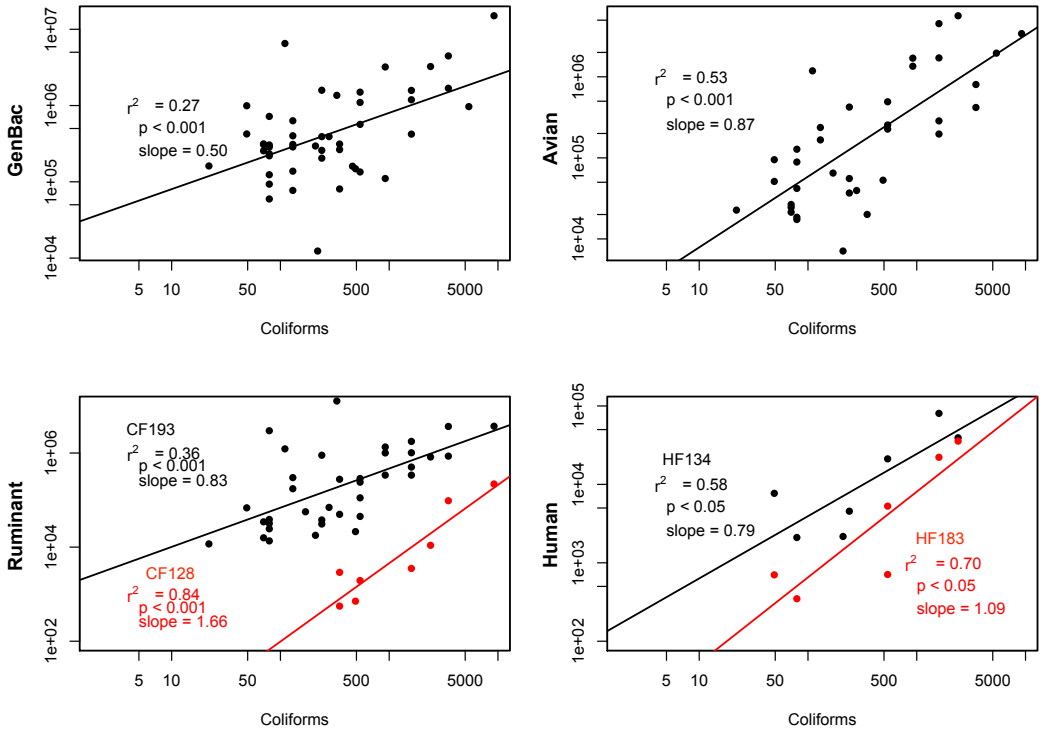

Figure 1. Significant positive correlations between numbers of copies of fecal molecular markers recovered in a sample and fecal coliforms in the same sample. X axes: MPN fecal coliforms/100ml. Y axes: copies of the specific marker recovered by qPCR. r^2: correlation coefficient. p: probability of this pattern occurring by chance alone.

Controls. All of the filtration blanks were negative for all markers. Slight contamination occurred in one out of four extraction blanks; it was positive for GenBac, but the number of copies measured (374) was two orders of magnitude smaller than the lowest number of copies found in any water sample (12,373). The lowest number of GenBac marker copies in any sample testing positive with one of the specific markers was 80,549. Thus the amount of GenBac contamination seen in the extraction blank would not change the results, even if present in many samples.

All of the negative controls were negative for all markers.

Marker Specificity and Sensitivity. Marker specificity is the percentage of samples amplified by a particular marker that is actually in the targeted host group, or true positives (it measures how often the

marker or assay gives false positive results, or cross-reactions). Sensitivity measures what percentage of the targeted host group tests positive for the assay (it measures the likelihood of finding feces from the targeted host group if it is present; some members of the targeted group may not carry the marker). Marker specificity and sensitivity for each assay were established previously by testing them with DNA extracts from human, cow, gull, goose, duck, chicken, pig, horse, sheep, cat, dog, seal, and sea lion feces collected from a variety of locales in the US, with the majority from the Northwest. However, as we were not provided with local fecal samples from the specific region of this study, we did not establish specificity and sensitivity for the markers in the specific region.

The GenBac marker, which is often treated as though it was a general fecal indicator marker, has been found in varying amounts in feces from every mammal we have tested. However, our laboratory and others have found that most avian species have low numbers of the *Bacteroides* bacteria targeted by GenBac, so this assay targets general mammalian fecal contamination, and to a lesser extent, avian fecal contamination.

Limits of Detection and Limits of Quantification. Limits of detection (LOD) and limits of quantification (LOQ) have been previously established for each marker, calculated as the lowest number of gene copies that a marker can detect, and the lowest number of copies that it can detect quantitatively, respectively. We previously tested serial dilutions of fresh feces in order to establish the relationship between the limits of quantification and most probable number (MPN) of fecal indicator bacteria (FIB). We found that with the exception of the dog marker, LOQ of all the markers were below the EPA mandated FIB cut-offs for recreational waters. For example, the LOQ of the avian marker, GFD, of 200 copies corresponded to a chicken fecal dilution producing 87 MPN total coliforms or 12.6 MPN *E.coli*/100 ml (Green et al., 2011). The LOQ of the dog assay was 10 to 100 times higher.

Discussion

This study provided evidence for fecal contamination from ruminants and birds, and to a lesser degree, humans.

Ruminant. The ruminant markers theoretically can measure contamination from a range of ruminant species, such as cattle, sheep, goats, deer and elk. In a study in the Tillamook Bay watershed, Oregon (Shanks et al., 2006) we used an elk-specific assay on samples positive for the ruminant markers, as elk were suspected to be a potential contributor to contamination. However, in that study we found no elk-positive samples, and we concluded that the most likely source of ruminant contamination in the Tillamook watershed was cows. That conclusion was supported by land-use data in that watershed, which included heavy dairy farming. Without more data for the Samish watershed, we cannot distinguish the contributions of cattle from deer and elk.

Avian. The avian marker, GFD, is 100% avian specific, and amplifies fecal DNA from gulls, geese, ducks, and chickens, as well as from a variety of other seabirds (Green et al., 2011). GFD detects a relatively high percentage of fecal samples from the west coast of the US and Canada. Because in this study, no samples were positive for the gull marker, GFC, we concluded that gulls did not contribute to the high number of samples exhibiting avian contamination. Other than excluding gulls, we currently do not have the technology to further identify the bird species responsible, and suggest that this will require field surveys. However, in general, exposure to bird feces is considered less harmful to humans than exposure to other sources of fecal contaminants, especially that of humans (Schoen and Ashbolt, 2010; Soller et al., 2010). Wild birds can spread pathogens to domestic poultry; an example is avian influenza.

Human. Both human markers, HF134 and HF183, detect contamination from fresh human feces, sewage and septage, and both are in use around the world. However, it has frequently been observed that the distributions of these markers differ widely among different human populations. In general, we saw much higher levels of contamination from ruminants and birds than from human sources.

Dog. Although only one sample (061411-3) was positive for DF155, the dog assay, it had a relatively high level of contamination (627,615 copies/100 ml), causing us to suspect that it had been spiked. However, this sample was also positive for both ruminant markers.

Horse. Because we did not detect Ho597, the horse fecal marker, in any samples, it would be a good idea to analyze horse fecal samples from the study area to make sure that they carry Ho597; if not, there are other published markers for horse fecal contamination that could be tried.

Marker distribution. Without a survey of fecal samples from the local host populations, we cannot estimate how many populations or individuals carry the marker bacteria and at what concentrations, and thus how likely it is that some contamination could have been missed. The results for the human and ruminant markers reveal that the markers have very different distributions within their targeted host groups. Although copy numbers of the two ruminant markers were positively correlated, only 8 water samples tested positive for CF128, while 38 samples tested positive for CF193. There were no samples testing positive for CF128 that were not also positive for CF193, indicating that using CF128 in this sample set did not provide any information that was not also provided by CF193, and future studies in this locale could consider dropping CF128. Although copy numbers measured by the two human markers were also positively correlated, of the 12 samples positive for a human marker, 8 were positive for only one of the two. Thus using both human markers gave more information than just one would have done. However, of the two, more samples were positive for HF134 but negative for HF183 (6) than were positive for HF183 but negative for HF134 (2). Not using HF183 in this study would have resulted in 16% fewer samples identified as containing human fecal contamination.

Correlations. The good correlations among quantities of human, ruminant and avian markers and fecal coliforms were expected, because both FIB and molecular source tracking assays are methods of enumerating fecal bacterial contamination, although they are based on different groups of fecal bacteria.

The negative correlations between quantities of ruminant markers and dissolved oxygen (DO) is likely to be caused by the high biological oxygen demand (BOD) needed by microbes to break down an influx of fresh ruminant feces. However, notably, the same thing was not seen with the human markers. This could indicate that the human markers were coming from sewage, or some similar well-aged source, as secondary sewage treatment removes a significant portion of BOD.

It would be informative to add rainfall data to this analysis.

Limitation of the Study. It is necessary to understand the limitations of the technique in order to avoid over- or under-interpretation of the results. The most important limitation relates to sensitivity and specificity of the markers. Not all population members of each host species in a given area carry the particular fecal bacteria targeted by the molecular assays, because an individual's composition of fecal bacteria changes in relation to diet, temporally, and geographically.

The analysis is also limited by the lack of markers for other species or groups that may be important contaminators of water, such as rodents and raccoons.

The frequent presence of inhibition, measured with the inhibition control PAO, required us to dilute inhibited samples in an effort to extract data from these samples. Although this was often successful, in some samples it may have diluted the markers below their limit of detection (Table 6).

Although the *data* are quantitative, they cannot be used to directly measure the relative or proportional contribution from each potential source of fecal contamination. All the assays used in this study targeted ribosomal genes in host-specific fecal bacteria. The *Bacteroides* group, targeted by the human, ruminant, horse and dog assays, has an average of 6 copies of the ribosomal gene per bacterial cell; thus measuring 6 gene copies in a sample equates to detecting a single *Bacteroides* cell (Lee et al., 2009). The bacterial target of the avian assay, *Helicobacter,* has only two copies of the ribosomal gene per cell; measuring 2 gene copies in a sample equates to detecting a single *Helicobacter* cell (Lee et al., 2009). Thus when we ranked samples as "+", "++", or "+++", we took this difference into account.

Water quality managers typically measure water quality in terms of FIB, and would prefer both to know the source of the FIB, and to be able to estimate how much FIB each source contributed. However, individual hosts and species vary in amounts of targeted fecal bacteria, not only in response to diet but also temporally and geographically. This is also true for FIB; for example, bird feces are high in enterococci, while human feces are much higher in *E. coli.* In addition, each bacterial group has a unique survival curve once released into the environment, influencing how long it can be detected. Thus although the qPCR analyses produce quantitative data, we interpret it qualitatively, as either absence, or presence

of some, a middling amount, or a lot of contamination from each source. More research is needed on how these amounts correlate to pathogens.

A final limitation is inherent in all water sampling: water is a dynamically moving system, and sampling may miss some things. Careful sampling design, a good sanitary survey and a detailed understanding of the watershed can help to offset this risk, and is particularly important considering the much higher cost of molecular assays compared to culture-based assays for FIB.

Conclusions

This study produced strong evidence that ruminants and birds are responsible for a significant part of the microbial contamination found in the water samples analyzed. In addition, as both domestic ruminant and human contamination are likely to be amenable to remediation, the analysis allows specific areas where remediation would be most useful to be identified.

References

Bernhard, A. E., and K. G. Field. 2000. A PCR assay to discriminate human and ruminant feces based on host differences in *Bacteroides-Prevotella* 16S ribosomal DNA. *Appl Environ Microbiol* 66: 4571-4574

Dick, L. K. and K. G. Field 2004. Rapid estimation of numbers of fecal *Bacteroidetes* by use of a quantitative PCR assay for 16S rRNA genes. *Appl Environ Microbiol* 70: 5695-5697

Dick, L. K., A. E. Bernhard, T. J. Brodeur, J. W. Santo Domingo, J. M. Simpson, S. P. Walters, and K. G. Field 2005. Host Distributions of Uncultivated Fecal *Bacteroidales* Reveal Genetic Markers for Fecal Source Identification. *Appl Environ Microbiol* 71: 3184-3191

Dick, L. K. and K. G. Field 2005. Microplate subtractive hybridization to enrich for source-specific *Bacteroidales* fecal pollution indicators. *Appl Environ Microbiol* 71: 3179-3183.

Green, H. C., Dick, L. K., Gilpin, B., Samadpour, M., and K. G. Field. 2011 (in review) Genetic Markers for Rapid PCR-based Identification of Gull, Canada Goose, Duck, and Chicken Fecal Contamination in Water. *Appl Environ Microbiol*

Green, H. C., Shanks, O. C., Sivaganesan, M., Haugland, R. A., and Field, K. G. 2011 (in press). Extended Survival of Human Fecal *Bacteroides* in Marine Water. *Environ Microbiol*

Lee Z.M., Bussema C. 3rd and Schmidt, T.M. 2009. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Res* pp1-5.

Schoen, M.E., and Ashbolt, N.J. 2010. Assessing Pathogen Risk to Swimmers at Non-Sewage Impacted Recreational Beaches. *Environ Sci Technol* 44: 2286-2291.

Shanks, O. C., C. Nietch, M. T. Simonich, M. Younger, D. Reynolds, and K. G. Field. 2006. Basin-wide analysis of the dynamics of fecal contamination and fecal source identification in Tillamook Bay, Oregon. *Appl Environ Microbiol* 72: 5537-5546

Soller, J.A., Schoen, M.E., Bartrand, T., Ravenscroft, J.E., and Ashbolt, N.J. 2010. Estimated human health risks from exposure to recreational waters impacted by human and non-human sources of faecal contamination. *Water Research* 44: 4674-4691.

Walters, S. P. and K. G. Field. 2009. Survival and persistence of human and ruminant-specific fecal *Bacteroidales* in freshwater mesocosms. *Environ Microbiol* 11:1410-21.