Enterococci Concentrations in a Coastal Ecosystem are a Function of Fecal Source Input, Environmental Conditions, and Environmental Sources

Derek Rothenheber[a] and Stephen Jones[ab#]

University of New Hampshire, Molecular and Cellular Biomedical Sciences, Durham, New Hampshire, USA[a], University of New Hampshire, Department of Natural Resources, Durham, New Hampshire, USA[b]

Running Head: Fecal Source Influence on Enterococci Concentrations

Key Words: fecal pollution, coastal ecosystem, enterococci, microbial source tracking

#Address correspondence to Steve Jones, Stephen.jones@unh.edu

ABSTRACT

Fecal pollution at coastal beaches requires management efforts to address public health and economic concerns. Fecal-borne bacterial concentrations are influenced by different fecal sources, environmental conditions, and ecosystem reservoirs, making their public health significance convoluted. In this study, we sought to delineate the influences of these factors on enterococci concentrations in southern Maine coastal recreational waters. Weekly water samples and water quality measurements were conducted at freshwater, estuarine, and marine beach sites from June through September 2016. Samples were analyzed for total and particle-associated enterococci concentrations, total suspended solids, and microbial source tracking markers (PCR: Bac32, HF183, CF128, DF475, Gull2; qPCR: AllBac, HF183, GFD). Water, soil, sediment, and marine sediment samples were also subjected to 16S rRNA sequencing and SourceTracker analysis to determine the influence from these environmental reservoirs on water sample microbial communities. Enterococci and particle-associated enterococci concentrations were elevated in freshwater, but suspended solids concentrations were relatively similar. Mammal fecal contamination was significantly elevated in the estuary, with human and bird fecal contaminant levels similar between sites. A partial least squares regression model indicated particle-associated enterococci and mammal marker concentrations had the most significant positive relationships with enterococci concentrations across marine, estuary, and freshwater environments. Freshwater microbial communities were significantly influenced by underlying sediment while estuarine/marine beach communities were influenced by freshwater, high tide height, and estuarine sediment. Elevated enterococci levels were reflective of a combination of increased fecal source input, environmental sources, and environmental conditions, highlighting the need for encompassing MST approaches for managing water quality issues.

IMPORTANCE

Enterococci have long been the federal standard in determining water quality at estuarine and marine environments. Although enterococci are highly abundant in the intestines of many animals they are not exclusive to that environment and can persist and grow outside of fecal tracts. This presents a management problem for areas that are largely impaired by non-point source contamination, as fecal sources might not be the root cause of contamination. This study employed different microbial source tracking methods to delineate influences from fecal source input, environmental sources, and environmental conditions to determine which combination of variables are influencing enterococci concentrations in recreational waters at a historically impaired coastal town. Results showed that fecal source input, environmental sources and conditions all play a role in influencing enterococci concentrations. This highlights the need to include an encompassing microbial source tracking approach to assess the effects of all important variables on enterococci concentrations.

INTRODUCTION

Fecal contamination of coastal recreational waters is a significant public health concern, as fecal material, often from nonpoint sources, can harbor an array of different pathogens. The US EPA has established regulations based on enterococci bacteria as the indicator of fecal-borne pollution to help manage water quality at estuarine and marine beaches (1). These organisms correlated well with predicted public health outcomes in several epidemiological studies that served as the basis for their adoption as the regulatory water quality indicator (2–5). The presence of human feces can present an elevated public health risk in recreational waters compared to non-human

65 sources due to the lack of an "inter-species barrier" for diseases and the higher density of human
66 pathogens that humans can carry (6–8). Although human pollution represents the greatest public
67 health risk, other fecal sources that contain enterococci and possibly human pathogens can be
68 chronic or intermittent sources of both, making beach water quality management and
69 remediation efforts more complex.

70

71 The need to differentiate fecal sources in recreational waters led to the emergence of microbial
72 source tracking (MST) methods in the early 2000s, most notably the PCR-based assays that
73 target the 16S rRNA gene in *Bacteroides* spp. (9, 10). There are a wide range of species-specific
74 genetic markers designed to identify human fecal sources and various domestic and wildlife fecal
75 sources. These assays have been in use for well over a decade and are supported by numerous
76 and rigorous laboratory evaluations and field applications (11–17). Initial field studies
77 investigated the relationship between MST markers and FIB concentrations in recreational
78 waters to better elucidate potential sources of fecal pollution. Some studies have found strong
79 relationships between the MST markers and enterococci (12, 18) while other studies have found
80 either weak or no relationships (19–21), many of which are discussed in a review by Harwood et
81 al. (22). One main factor affecting the relationship between enterococci and the relative strength
82 of different sources of fecal contamination is that enterococci can persist and grow in the
83 environment, which can significantly influence their concentrations in recreational water (23).

84

85 Due to the pervasiveness of enterococci in natural ecosystems, recent studies have been
86 conducted to not only elucidate environmental parameters controlling their growth, but also to

identify naturalized niches that can act as reservoirs for enterococci and the associated influence on water quality measurements. Specifically, enterococci have been shown to persist in fresh water sediments (24–26) and marine sediments (24, 27), and in some cases their relative concentrations in sediments are several orders of magnitude higher than the overlying water (24, 28–30). In addition, enterococci persist in soils affected by anthropogenic activities (31) as well as more natural soil environments (32–34). Thus, soil can act as a significant reservoir of enterococci that can, if eroded, confound concentrations observed in recreational waters. Evaluating the influence of sediment and or soil on water quality has, in some studies, been conducted by measuring total suspended solids as a surrogate for sediment-associated enterococci (27, 35, 36), however this non-specific approach does not indicate the specific type of source(s) of the suspended solids. With the advent of next generation sequencing, sources of sediment or soil bacteria can be fingerprinted via 16S rRNA sequencing, and programs like SourceTracker can then determine relative fractions of source-specific 16S fingerprints within a water sample (37).

This study examined the coastal and estuarine beaches of Wells, ME where there has been historically elevated enterococci levels, as reported by the Maine Healthy Beaches Program (38). Prior to this study, only a ribotyping-based MST study (39) that also involved other indicator tracking work had been conducted in this area. In that study, the two major freshwater inputs, the Webhannet River and Depot Brook were found to be the major influences on water quality related to an array of fecal contamination sources. To investigate potential sources of enterococci we measured three major categories of variables (fecal source input, environmental conditions,

and environmental sources) and then used a partial least squares regression model approach to determine the most significant influences on the enterococci concentrations in water samples.

RESULTS

Total and particle-associated enterococci concentrations and total suspended solids in water. During this study, total enterococci concentrations were highest in freshwater sites, with concentrations significantly decreasing from there to the estuary and then the marine beach areas (Figure 2). The geometric mean enterococci concentrations were 197 and 40 CFU/100 ml at the Depot and Webhannet sites, respectively, with 71% of samples exceeding 104 CFU/100 ml at the Depot site compared to 21% at the Webhannet site. In contrast, the geometric mean enterococci concentrations at the other sites were all <15 CFU/100 ml and samples exceeded 104 CFU/100 ml 0% (at Wells Beach) to 25% of the time. In addition to measuring enterococci concentrations in water samples, particle-associated enterococci and suspended solid concentrations were measured to better understand the potential mode of transport of these bacteria within this coastal watershed. Throughout the study period (June-September 2016), levels of total and particle-associated enterococci varied by site. Concentrations were lowest at the marine beach (Wells Beach) compared to other sites, with levels significantly higher in all estuary sites (W11-W15) and freshwater sites (Depot & Webhannet; Figure 2).

Both total and particle-associated enterococci geometric mean concentrations were statistically similar at the estuary beach (W11, W12, W13) and estuary (W14, W15) sites. Freshwater sites (Webhannet and Depot) however, had statistically higher enterococci concentrations than other

131 sites (Figure 2; $p < 0.05$). The ratio of total to particle-associated enterococci varied throughout
132 the season, with an average of 36.3% (SD ± 30) across all sites. Sites within the estuary beach
133 showed the highest ratio (41%, SD ± 32), however there were no significant differences observed
134 between sites or types of sites. Average TSS concentrations were relatively low and similar for
135 most sites, with an overall average of 2.9 mg TSS/L (SD± 1.2). The Webhannet freshwater site,
136 however, had a significantly lower average TSS concentration (1.2 mg/L ± 1.0SD, $p < 0.05$)
137 (Figure 2), despite, as previously mentioned, having higher enterococci concentrations. The
138 relationship between particle-associated enterococci and TSS was not significant ($r^2 = 0.0011$),
139 and significant rainfall events were seldom and sparse with only one greater than 1 in 48 h prior
140 to sampling. Overall, this study showed enterococci concentrations were significantly different
141 by site and were ubiquitously associated with particles, which was independent of suspended
142 solids concentrations.

143

144 **Presence of fecal sources in fresh, estuarine, and marine waters.** The concentration of fecal
145 pollution in this study area was determined using both PCR and quantitative PCR MST assays to
146 identify and quantify predominant sources of fecal contamination present in the water. The
147 mammal fecal marker (Bac32) was detected via PCR at all sites 100% of the time throughout the
148 study period. (Supplementary Material 1E). The human fecal marker (HF183) was detected in
149 51% of all water samples, with the highest detection rate in fresh water (56%) and the lowest
150 detection rate in marine beach water (46%). Differences in the percent detection of the gull fecal
151 marker (Gull2) were most pronounced between freshwater (10%) and all other sites (>77%). The
152 dog fecal marker (DF475) detection rate was highest in the estuary beach water (10/44 = 23%),
153 however 8 of the 10 positive samples were detected in July (8/13 = 61%). For all other sites, an

154 increase in the detection of dog fecal marker also occurred during July, with 44% (16/36)
155 detection, compared to 0% for August and September and <1% for June. Thus, most of the dog
156 contamination at all sites was associated with unknown dog-related conditions during July.

157

158 **Concentrations of mammal, human, and bird fecal sources.** We used qPCR to provide
159 relative quantitative measures of mammal, human and bird fecal contamination levels. Water at
160 estuary and estuary beach sites contained significantly higher levels of mammal (AllBac) fecal
161 marker copies, with an average of 1.54×10^7 compared to 2.62×10^6 in freshwater and 3.9×10^6
162 copies/100 ml in marine beach ($p < 0.05$). Average concentrations of human (HF183) and bird
163 (GFD) fecal markers were not statistically different between sites, however, concentrations of the
164 human marker in individual samples varied from <LOD - 2.04×10^4 copies/100 ml (Figure 3),
165 while bird fecal marker concentrations were relatively stable across all sites. No significant
166 temporal trends were observed for any of the quantitative fecal marker levels. Compared with
167 presence/absence detection of fecal sources, quantitative measurements also did not show strong
168 spatial patterns, except mammal marker levels showed significant increases at estuary and
169 estuary beach sites compared to marine and freshwater sites.

170

171 **Differences between water, soil, and sediment bacterial community compositions**. 16S
172 amplicon sequencing was used to characterize the microbial community present in water and
173 other sample matrices (soil, sediment, and marine sediment), which was the nexus for ensuing
174 SourceTracker analysis. A total of 3,276,196 reads and 7,706 unique OTUs were obtained from
175 the 177 samples of fresh, estuary, estuary beach and marine beach water and soil, sediment, and

marine sediment. The number of OTUs assigned and the Shannon diversity index were significantly higher for soil, sediment, and marine sediment when compared to water samples (Figure 4, $p < 0.05$). Most taxa in the estuary and marine beach water samples were identified as Flavobacteriia, Alphaproteobacteria, and Gammaproteobacteria classes, which together accounted for 84% of the total assigned taxa. Cyanobacteria accounted for 34% of the taxa in marine sediment, and Betaproteobacteria was one of the top three most abundant taxa in fresh water, soil and sediment (Figure 4). A Non-Metric Multi-Dimensional Scaling **(**NMDS) ordination was used to determine if the bacterial communities from water and other matrices (soil and sediments) differed based on their taxonomic composition. Bacterial communities from the marine beach and estuary (All Estuary) waters were similar, but were statistically different from fresh water (Figure 5, $p < 0.05$). The bacterial communities associated with soil, sediment and marine sediment were all distinct when compared to each other and water samples, indicating unique groups of OTUs (Figure 5, $p < 0.05$). Samples taken from different areas within the watershed (soil, estuarine water, freshwater, etc.) contained unique bacterial compositions, allowing for downstream analysis with the SourceTracker software to discern relative contributions of these different communities to the make-up of microbial communities in the different types of water samples.

Environmental source contribution to water samples. The fraction of freshwater, sediment, soil, estuarine sediment, and marine beach water source bacterial communities within estuary and estuary beaches water samples were calculated using the Bayesian mixing model SourceTracker. Freshwater sample analysis showed a high probability of taxa originating from underlying sediment (74%) and much lower probability of taxa originating from soil (2.6%).

199 Initial results for the estuary and estuary beach indicated that marine beach water was the
200 dominant source of bacteria (Table 1). However, given that likely fecal sources are coming from
201 the watershed, we excluded marine beach water as a potential source and included it as a sink
202 then re-analyzed the data. These second results showed that freshwater taxa had a high
203 probability of being a significant fraction of estuary (73%), estuary beach (66%) and marine
204 beach (35%) water communities, with a significantly higher percentage for the estuary locations
205 compared to the marine beach (Table 1, $p < 0.05$), which is more influenced by ocean microbial
206 taxa. Despite the significant percentage of freshwater taxa assignments in the estuary, estuary
207 beach, and marine beach waters there were no freshwater sediment or soil taxa assignments for
208 these sites. The data for the percent of unidentifiable taxa showed the opposite trend compared
209 to percent of assigned freshwater taxa. Unidentifiable taxa in the marine beach were significantly
210 higher (46%; $p < 0.05$), which is not surprising given that marine beach water community would
211 likely be most influenced by non-terrestrial sources. Estuarine sediment was the highest likely
212 identified source in the water from the marine beach site (19%), and it was significantly higher
213 than percentages calculated for all estuary sites ($p < 0.05$). Overall results showed that freshwater
214 source-related taxa were a pervasive source throughout the estuary and marine beach, and while
215 sediment source-related taxa were highly abundant in the freshwater they were not observed
216 within the estuary or marine beach.

217

218 **Relationships between environmental conditions, fecal source concentrations,**
219 **environmental sources and enterococci concentrations**. Two PLSR models were created to
220 determine relationships between enterococci and fecal source concentrations, environmental
221 sources, and environmental conditions (outlined in the Methods). The first 'freshwater' PLSR

model indicated particle-associated enterococci concentration, concentration of mammal fecal marker, TSS concentration, percent of sediment source, percent of unknown source, and salinity were important variables (VIP > 0.8) in resolving variation in enterococci concentrations (Table 1). A one-factor (single PLSR regression) model was deemed optimal (root mean PRESS = 0.735), and showed that all variables (except salinity) had positive associations with enterococci concentrations. Values for model performance ($R^2Y = 0.6$, $R^2X = 0.5$, and $Q^2 = 0.4$) indicated that the model fit the data moderately well ($R^2X \geq 0.5$) but had poor predictive capability of enterococci concentrations ($Q^2 < 0.5$; Supplementary Material 3). Out of all the important variables, particle-associated enterococci (Particle ENT) concentrations showed the strongest relationship to total enterococci concentrations (Table 2). The second PLSR model, a two-factor/two PLSR regressions model, was the best fit (root mean PRESS = 0.744) from the PLSR constructed for the estuary, estuary beach, and marine beach sites. The analysis identified particle-associated enterococci concentration, mammal fecal source concentration, percent of freshwater, unidentified and estuarine sediment sources, water temperature, and high tide height as significantly related to enterococci concentrations. Factor one showed that all variables were positively associated, except for the percent unidentified and marine sediment sources. The second factor showed mammal fecal sources, freshwater sources, and water temperatures were negatively related to enterococci concentrations, which was the opposite of their associations for factor one. The high tide height and marine sediment were positively related to enterococci concentrations for factor 2 of the PLSR (Table 2). Together both factors explained 61.8% in the variation observed in enterococci concentrations, and model performance ($R^2Y = 0.6$, $R^2X = 0.5$, and $Q^2 = 0.6$) indicated better predictive ability with a similar fit to the data compared to the freshwater model (Supplementary Table 3). Out of all the potential variables measured (19 total)

245 across three categories (fecal source input, environmental source contribution, and environmental
246 conditions), particle-associated enterococci and mammal fecal marker concentrations had the
247 most significant relationships to enterococci concentrations. The relationships between other
248 variables and enterococci concentrations were specific to freshwater and estuary/marine beach
249 models, indicating ecosystem specific relationships. However, the joint relationship of particle-
250 associated and mammal fecal marker across freshwater and estuary/marine environments
251 indicate their overarching importance in determining enterococci concentrations.

252

253 **4 Discussion:**

254 Geometric mean enterococci concentrations at the marine beach, estuary, and estuary beach
255 sampling sites were all less than the State of Maine water quality standard of 35 CFU/100 ml and
256 the majority of single sample concentrations were less than the State 104 CFU/100 ml single
257 sample maximum standard, indicating the water quality was typically considered acceptable for
258 recreational use. Previous monitoring by the Maine Healthy Beaches Program in 2014 had
259 shown the Wells Beach area was one of 7 beaches in Maine that had a greater than 20%
260 exceedance rate, with suspicion that freshwater inputs are a significant source of contamination
261 (38). Our findings confirmed that enterococci concentrations were statistically higher at both
262 major freshwater tributaries to the estuary, especially at the Depot Brook site where levels were
263 regularly above the 104 CFU/100 ml single sample standard. The Depot Brook site is located in
264 a watershed with a higher fraction of developed land (0.27-0.50) and more people per km^2 (325-
265 2,650 people) compared to the Webhannet site watershed that has a lower developed fraction
266 (0.13-0.25) and 150-325 people per km^2; 40). This could help explain the difference in
267 enterococci concentrations between freshwater sites as a more urbanized watershed can increase

transport of more pollution from the watershed to the freshwater tributary. However, the summer of 2016 was especially dry in this region (41) with just one event with >1 inch of rain (1.73 in., 6/28/16) 48 h prior to the sampling time. This overall dry condition likely contributed to less fecal contamination transport (via freshwater discharge) from the watershed to the estuary and marine beach. This suggests that more typical rainfall conditions would probably have resulted in more freshwater discharge and higher enterococci concentrations than what we observed.

Enterococci were significantly associated with suspended particles of >3.0 µm diameter (R^2 = 0.96, $p < 0.05$). On average, 36% (SD ± 30) of the total enterococci concentrations were associated with particles, which suggests particles as a potentially important transport mechanism. Other studies conducted in estuary and storm waters have found similar fractions of particle associated enterococci, but they noted enterococci demonstrated a preference for a larger particle size of >30 µm (42–44). The large standard deviation for particle-associated enterococci could be attributed to the complex nature of particle interactions (sedimentation rate, electrostatic, hydrophobic, and other surface-surface interactions) and hydrogeological dynamics (salinity-driven turbidity maximum) (45). The mechanisms underlying enterococci-particle interactions may also be related to ionic strength in surface waters, as *Enterococcus faecalis* is negatively charged over a broad pH range (2-8 pH units) and in the presence of different ion concentrations (46). Results for this study indicate that TSS and particle-associated enterococci had no linear relationship, indicating particle-associated enterococci were not dependent on the total amount of suspended material and thus the association is likely due to other factors influencing cell-particle interactions.

291 Quantitative PCR assessment of several fecal sources is a potentially useful strategy to determine
292 the relative significance of the different sources in a single sample and over time at sites of
293 interest. PCR detection showed a chronic presence of mammalian fecal source(s) (100% of
294 samples) with human fecal source(s) detected in approximately half of all samples, so qPCR
295 analysis is useful for bringing context to the significance of these findings. For example, Mayer
296 et al. (47) showed that wastewater effluent contains about 10^8 copies/100 ml of the AllBac
297 mammal fecal marker, Sowah et al. (48) found that streams impacted by septic systems could
298 contain $10^5 - 10^7$ copies/100 ml depending on the season, and Bushon et al. (49) determined that
299 under storm flow conditions in an urban watershed mammal marker copy numbers could exceed
300 10^8 copies/100 ml. Results for this study ranged from 10^5 to 8.6 x 10^7 copies/100 ml, values that
301 are within previously reported ranges and likely a concentration reflective of a predominantly
302 non-urbanized watershed and intermediate mammal source loading. The estuary and estuary
303 beach area showed a statistically higher concentration of the mammal marker, however, there
304 was no responsive increase in the concentrations of the human associated fecal marker (HF183),
305 which may indicate that humans are not the primary mammalian source for the increased fecal
306 contamination.

307

308 The average concentration of the human marker was 1,500 copies/100 ml across all sites
309 (geometric mean 167 copies/100 ml), with the highest concentration being 20,364 copies/100ml
310 (Webhannet 6/22/16). Boehm et al. (50) showed that 4,200 copies/100 ml of HF183 in
311 recreational waters contaminated by raw sewage is the cutoff for where GI illnesses exceed the
312 EPA acceptable risk level of approximately 30/1000 for swimmers (1) Another recent study also
313 established a copy number cutoff for the HF183 assay using a threshold of 36 predicted GI

illness per 1000 swimmers as a benchmark. Their findings show at 3,220 copies/100 ml in untreated sewage and at 3,660 copies/100 ml of HF183 in secondary treated sewage predicted GI illness exceeds the 36/1000 swimmers threshold (51). It is important to note that both studies simulated risk based on either raw sewage or secondary treated sewage contamination and do not account for differential decay between pathogens and molecular markers. In our study, the primary source of contamination seems to not be sewage, thus direct application of these thresholds to our study might not convey the same risk. However, we chose to compare our HF183 copy numbers to the threshold published by Boehm et al. (50) due to their use of the EPA acceptable risk level of 30/1000 threshold and to give a general context for potential health risk for this study. On average, sites in this study did not exceed this benchmark level, however, there were 10 occasions when sites were above the 4,200/100ml threshold (7 different sites across 4 sampling dates), indicating that sporadic events or conditions can cause elevated human fecal contamination and potential public health concerns (Supplementary 4). Boehm et al. also showed that at the LOQ for most assays, 500 copies/100ml or 1000 copies/100ml, there is still a predicted GI illness of 4 or 8 cases per 1000 swimmers, suggesting positive detection at the LOQ is indicative of low level health risk (50). For this study, the LOQ was 250 copies/100ml for the HF183 assay and 67 of 117 samples (57%) tested positive at or above this limit, suggesting that over half of collected water samples indicated the presence of a low-level health risk. Although there were no statistical differences between sites for human fecal contamination, W11 did contain the highest geometric mean (493 copies/100 ml; Supplementary 4). This could be reflective of the location of the site as it's where drainage from the Webhannet and Depot watershed meets and is also directly downstream from a boat marina with the harbor sewage pump station, which could be a possible point source of contamination. Nonetheless, even

337 though sites on average were below published thresholds, detection of human contamination
338 even at low concentrations is a concern.

339

340 Although human fecal sources are the greatest public health concern (6, 7, 22, 52) we did not
341 observe any relationship between human fecal contamination and enterococci concentrations.
342 We did however observe a positive relationship between mammal fecal contamination and
343 enterococci concentrations through PLSR analysis, thus suggesting other mammalian fecal
344 source(s) are more influential in explaining the variation observed in this study. It is important to
345 note that the mammal marker is at a higher copy number in sewage/feces compared to the human
346 marker making it by default easier to detect. Also, the mammal and human molecular markers
347 could decay at different rates in the environment thus masking any potential relationship between
348 human fecal contamination and enterococci concentrations. Interestingly gull fecal sources were
349 detected in 77% or more of the samples in the estuary and marine beach area, however only 10%
350 of the samples were positive within the fresh water (Supplementary Material 1), despite there
351 being no decrease in the bird fecal marker concentration, suggesting the presence of different
352 bird sources in these areas. Anecdotally, Canada geese were observed upstream of both the
353 Webhannet and Depot freshwater sites periodically throughout the season, which could be a
354 significant source of bird fecal contamination in the fresh water locations (53).

355

356 One of the unique findings of this study was the relative contribution of different sources to the
357 bacterial community in the estuarine water. The bacterial community in estuarine water primarily
358 originated (>90%) from marine beach water, which is not surprising for a well-flushed estuary

like the study site. Because the study period was minimally influenced by rainfall and associated runoff of freshwater, we expected that the influence of freshwater sources would be low. In ensuing analyses, we chose not to include marine beach water as a potential source for a variety of reasons. First, the samples were always collected during low tide before the ebb when the estuary water was draining and water was moving from the watershed towards the marine beach. Secondly, we had already shown that the OTU compositions for the marine beach and estuary samples were very similar, increasing the possibility of a type I error (false positive) for identifying marine beach as the likely source of enterococci. Lastly, fecal pollution sources most likely come from the watersheds and not from marine water, so excluding marine beach water helps to enhance the determination of watershed influences. Our second analysis (marine beach source excluded) showed that freshwater was a significant source of bacteria to the estuary (>65% assignment) compared to soil, sediment, and estuarine sediment. This implicates freshwater as a major conduit for bacterial transport, as well as the major source of enterococci to the estuary. Overall this finding highlights the importance of freshwater discharge as a controlling factor in transporting contamination from the watershed to the coast. The specific percent assignment of freshwater source could be an over-estimate, however the trend observed is a likely scenario given the rational discussed.

Analysis of environmental reservoirs of enterococci (soil, sediment, etc.) and their presence within water samples using SourceTracker revealed a variety of source contributions to freshwater, estuary and marine waters. To date there have been limited studies using SourceTracker to identify soil and sediment-associated taxa within water samples, and none of these studies have focused on a coastal watershed with the potential for freshwater, estuarine and

marine sources. One study conducted in the upper Mississippi River identified up to 14% of sediment and 1.4% of soil sources of the taxa within the river water (54). This study, however showed that the sediment source was much more abundant in freshwater (74%), indicating a greater degree of mixing between the freshwater and underlying sediment communities. The amount of sediment and soil sources within water samples may be related to site specific characteristics such as relief or soil texture, which has been shown with TSS fluxes on a global scale (55). Thus, the degree to which the underlying sediment community mixes with the overlaying water is likely site specific. Interestingly, even though freshwater contained a significant amount of sediment source taxa, no sediment source was observed at the estuary and marine beach sites through the SourceTracker analysis. This difference could indicate that rapid sedimentation happens during transit to and within the estuary and at the estuarine turbidity maximum zone (56). TSS concentrations and the ratio of particle-associated to total enterococci concentrations, however, showed no differences between freshwater and estuary/marine sites. This could be related to the separate and quite different hydrodynamics within these different water systems. The percent of sediment source in the freshwater samples observed here might also be an over-estimate/over fit from SourceTracker given the limited number of potential sources used, but results consistently showed an elevated presence of sediment in all freshwater samples in this study. SourceTracker analysis also revealed that the freshwater source was significant (35% or more) in estuary and marine beach water samples, suggesting that fresh water is a significant conduit for microbial, and fecal contamination, transport from the watershed to the estuary and marine beach.

Applied and Environmental Microbiology

404 The use of predictive models for water quality has been a focus in the field in parallel with the
405 adoption of bacterial indicator organisms as the gold standard for water quality determination.
406 The goal of this research was to identify significant influences on enterococci concentrations by
407 measuring a wide variety of variables. To distill this information, we used a PSLR model, which
408 has been shown to out-perform similar multiple linear regression and principle components
409 regression analyses (57) and has gained popularity in the water quality field (58, 59). Results
410 from the PLSR analysis in this study showed that particle-associated enterococci and
411 concentrations of mammal fecal sources were the driving force behind variation in enterococci
412 concentrations, as described by both PLSR models constructed. Other factors were found to
413 influence enterococci concentrations, however, these differed between the freshwater and
414 estuary/marine beach models. For example, TSS concentration as well as the percent of both
415 freshwater sediment and unknown sources positively influenced enterococci concentrations at
416 freshwater sites. This indicates that sediment is a likely source of enterococci that influences
417 concentrations measured in the water. Positive influences from the unidentified source taxa
418 suggests that there is either an alternative source (not measured in this study) within the
419 watershed that also influences enterococci concentrations or that SourceTracker could simply not
420 resolve all the potential sources we used. This finding is not surprising given the vast number of
421 potential sources of fecal pollution within a watershed and that fecal sources were not a part of
422 the SourceTracker analysis. Results from the estuary and marine beach model returned a two-
423 factor regression, with each factor essentially being the inverse of each other. Specifically, it
424 highlighted freshwater being a major conduit for microbial transport to and through the estuary.
425 Negative influences from the unknown source reaffirms this finding, along with positive
426 influences from the previous high tide height. The second factor explained approximately 15% of

the variation in enterococci concentration, therefore its importance must be weighed proportionately to factor one, which explained almost 50% of the variation. However, positive loadings from previous high tide height and percent of estuarine sediment indicate estuarine sediment could be a source of enterococci whose influence is dependent on tide height. The negative loadings from mammal fecal source(s) may indicate that enterococci originating from the estuarine sediment are not from mammal fecal sources.

Overall, the results from this study demonstrated that concentrations of enterococci in the coastal estuarine/marine beach study area were largely controlled by particle-associated enterococci and mammal fecal source input. The influence of these factors is likely universal across freshwater and estuarine environments, however other ecosystem factors likely play a role as well. For freshwater portions of the coastal watershed, sediment may act as a significant enterococci reservoir that is frequently re-suspended within the water column. Freshwater itself could act as a major conduit for bacterial transport to an estuary and marine beach area where other environmental factors (water temperature and high tide height) can influence enterococci concentrations as well. These findings highlight the dynamic nature of enterococci in natural aquatic ecosystems outside of the mammalian fecal tract, and that concentrations within fresh water and estuary/marine beach water are influenced by a variety of factors.

Materials and Methods:

Site description. This study was conducted in Wells, Maine, USA (Figure 1). Eight different sites were used to monitor water quality (n = 2 freshwater, n = 2 estuary, n = 3 estuary beaches, n

449 = 1 marine beach) as well as twelve soil, twelve fresh-water sediment and four estuarine
450 sediment sampling sites. Data for air temperature and rainfall amount for the 48 h prior to
451 sampling were obtained from Weather Underground
452 (https://www.wunderground.com/cgibin/findweather/getForecast?query=Wells,%20ME) and
453 characteristics of tides during sampling were obtained from US Harbors
454 (www.meusharbors.com).

455

456 **Water sampling**. Surface water samples were collected weekly from June to September 2016 (n
457 = 117). Sampling started two hours before low tide to maximize the potential impacts of
458 freshwater pollution sources, and samples from all estuary and marine beach sites were collected
459 before the slack tide. Water samples were collected in autoclaved 1L Nalgene™ Wide-Mouth
460 Lab Quality PPCO bottles (Thermo Fisher Scientific, Waltham, MA, USA), and environmental
461 parameters were measured with a YSI Pro2030® dissolved oxygen, conductivity, and salinity
462 Instrument (YSI Incorporated, Yellow Springs, Ohio, USA). A field replicate was collected at a
463 different site for each sampling event.

464

465 **Soil, sediment, and marine sediment collection.** Environmental sources were collected twice
466 throughout the sampling season to build source libraries that were "finger-printed" with 16S
467 sequencing and SourceTracker analysis. Six soil and sediment samples were collected upstream
468 of both freshwater sites (Webhannet and Depot; Figure 1). Soil samples were collected at the
469 crest of the stream embankment, where a 10 x 10 cm a plastic square template was placed down
470 and all soil (O-horizon) within the template at a 2 cm depth was collected. Samples were sieved

(USA Standard No. 5) to remove any loose-leaf litter and roots to only sample smaller soil particles and their microbes. Underlying stream sediments were collected using a Van Veen sediment sampler from depositional sites chosen based on the presence of fine grain sediments. One grab sample was collected for each site and then the top 2 cm of sediment was subsampled for analysis. Sediments were sieved (USA Standard No. 45) to remove coarse grain and gravel size particles. Four estuarine sediments were collected during low tide when intertidal sediments were exposed using the Van Veen sampler, and the top 2 cm were again collected for analysis.

Enterococci and total suspended solids quantification. Total and particle-associated enterococci were enumerated using the EPA Method 1600 membrane filtration protocol (60) and particle-associated enterococci were determined via filtration through a 0.47 mm diameter 3.0 µm pore size polycarbonate filter (Millipore™, Darmstadt, Germany) as first reported by Crump et al. (61). The filters were rolled onto plates containing mEI agar and incubated at 41°C ± 0.5°C; representative colonies were counted in 24 ± 2 hours. Total suspended solids (TSS) were measured using EPA method 160-2, where 500 ml of the water sample was used to determine TSS concentrations (62).

DNA extractions. DNA extraction from all matrices was performed with the PowerSoils® DNA Extraction Kits (MO BIO Laboratories, Carlsbad, CA, USA), with modifications to the manufacture's protocol needed to optimize the extraction from water sample filters. For water samples, 500 ml collected water sample was filtered through 0.47 mm diameter 0.45 µm pore size polycarbonate filter (Millipore™, Darmstadt, Germany), which was stored in a sterile 2 ml cryotube at -80°C for at least 24 h. Prior to DNA extraction, frozen filters were crushed into small pieces with an ethanol sterilized razor blade, a practice commonly used to maximize DNA

494 recovery (63-65). To minimize additional DNA loss during the extraction process solutions C2
495 and C3 (from manufacturer's protocol) were halved in volume and combined into a single step
496 (as per communication with manufacture). DNA extraction from soil, freshwater sediment, and
497 marine sediment were conducted per the manufacture's protocol.

498

499 **Microbial source tracking (MST) PCR and qPCR assays**. MST PCR assays that target
500 Mammals (Bac32; 65), Humans (HF183; 9), Gulls (Gull2; 66), Dogs (DF475; 10) and
501 Ruminants (CF128; 9) were used to determine the presence of fecal sources in water samples
502 (Table 3). Positive control plasmids were created for each PCR assay from fresh fecal samples
503 that came from each target organism (Human, Gull, Dog, and Cow). The TOPO™ TA™ Cloning
504 Kit was used (Invitrogen, Carlsbad, CA, USA), with a blue/white screen of *E. coli* transformants
505 on kanamycin (50 µg/mL) selective TSA plates. Positive *E. coli* colonies were screened with
506 their respective PCR assay, and PCR positive colonies were then grown in TSB and extracted
507 with the PureLink® Quick Plasmid Miniprep Kit (Invitrogen, Carlsbad, CA, USA). PCR assays
508 were run on a T100™ Thermal Cycler (BioRad, Hercules, CA, USA) with the GoTaq® Green
509 MasterMix (Promega, Madison, WI, USA). Cycling conditions and amplification protocols for
510 each assay targeted the different source specific markers and followed protocols delineated by
511 different studies: Bac32 (67) and HF183 (67), CF128 (68), DF475 (69), and Gull2 (66).
512 Quantitative PCR assays were also run to determine fecal source strength for Mammals (AllBac;
513 70), Humans (HF183; 71), and Birds (GFD; 72), primer and probe sequences can be found in
514 Table 3. All qPCR assays were run on a Mx3000P cycler (Agilent Technologies, Santa Clara,
515 CA, USA), TaqMan assays used the PerfecCTa® FastMix® II (QuantaBio, Beverly, MA, USA)
516 master mix and the SYBR green assay used the FastSYBR™ Green Master Mix (Applied

517 Biosystems, Foster City, CA, USA). A standard curve ranging from 10^6-10^2 copies (Mammal
518 assay) or 10^5-10^1 copies (Human & Bird assay) was also run for each experimental run with the
519 limit of quantification (LOQ) being 100 copies (Mammal) or 10 copies (Human & Bird) per
520 PCR. The Ct values, amplification efficiency, slope, and R^2 values for each standard curve were
521 compared to previously run standard curves, to ensure satisfactory performance before being
522 used to calculate copy numbers for that run. Each environmental sample was diluted 1:10 and
523 run in triplicate and the reaction volume (25 µl) contained a final concentration of 0.2 mg/ml
524 BSA. Amplification/cycling conditions were preformed per published protocols for AllBac (73),
525 HF183 (73), and GFD (16). TaqMan assays were run with an internal amplification control (74)
526 with a down-shift of 1 cycle considered inhibition. Samples spiked with a plasmid containing 10^4
527 copies of GFD amplicon were used as inhibition controls for the SYBR assay, with a recovery of
528 less than 10^4 copies (100%) considered inhibition. For a list of primers, probes, and standard
529 curve performance, see Supplementary Material 1.

530

531 **16S library preparation**. The V4 region of the 16S rRNA gene, using the 515F-806R primer-
532 barcode pairs, was used for amplicon sequencing (75). The Earth Microbiome Project protocol
533 was used for amplification and pooling of samples, with minor modifications (76). The Qubit®
534 dsDNA HS assay was used to quantify sample concentrations, and 500 ng of DNA was pooled
535 per sample. The pool was then run on a 1.2 % low-melt agarose gel to separate primer-dimers
536 from acceptable product, and bands between 300-350 bps were cut and extracted as described
537 above. The final DNA sample was then run on the Agilent Technologies 2200 TapeStation
538 system (Santa Clara, CA, USA) to determine final size, quality, and purity of sample. Each

library was sent to the Hubbard Center for Genome Studies at the University of New Hampshire to be sequenced (2 x 250 bp) on the Illumina HiSeq 2500 (San Diego, CA, USA).

Quality filtering and Operational Taxonomic Unit (OTU) picking. QIIME 1.9.1 was used to perform all major quality filtering, and OTU picking (77). Forward and reversed reads were quality trimmed (μ P25) and removed of Illumina adapters via Trimmomatic (78). Any reads that were less than 200 bps were discarded, and reads were merged with the QIIME joined_paired_ends.py, using a minimum overlap of 10 bps and a maximum percent difference of 10%. Paired-end data were analyzed using the QIIME open-reference OTU picking strategy with UCLUST for *de novo* picking and the Greengenes 13_8 database (79) for taxonomic assignment. Alternative OTU picking strategies were also tested to determine best workflow, for performance of difference strategies refer to Supplementary Material 2. Data for all sequenced samples are publicly available through NCBI BioProject (http://www.ncbi.nlm.nih.gov/bioproject/431501).

SourceTracker analysis. Samples from 4 source types (fresh water, soil, sediment, and marine sediment) and 4 sink types (fresh water, estuary water, estuary beach water, and marine beach water) were analyzed by the open-source software SourceTracker v1.0 (37). Default parameters were used (rarefaction depth 1000, burn-in 100, restart 10, alpha (0.001) and beta (0.01) dirichlet hyperparameters) in accordance with previously published literature (53, 80). A 'leave one out' cross validation was performed to assess the general performance of the model and source samples were iteratively assigned as sinks to assess how well a known sink would be assigned

(i.e. source = soil and sink = soil). The percent assignments from SourceTracker are the result of the Gibbs Sampler assigning OTUs from an unknown sample to sources in a random and iterative fashion, and then calculating likelihood of that OTU originating from said source. The final output can be interpreted as the percent (or likelihood) of OTUs present in an unknown sample originating from the sources used in the analysis

Partial least squares regression model. A partial least squares regression (PLSR) model was used to determine the most important and significant variables affecting enterococci concentrations (81). Two models were created, one for the estuary, estuary beach, and marine beach sites, and one for the freshwater sites. Particle-associated enterococci, environment variables (water temperature, air temperature, dissolved oxygen, salinity, height of previous high tide, rainfall in previous 48 h), fecal source strength (mammal, human, and bird), and percent of environmental source (fresh water, soil, sediment, and marine sediment) were used as explanatory variables for the non-freshwater model. The same parameters, except height of previous high tide and percent of freshwater source, were used for the freshwater model. All data except the percent assignments from SourceTracker were log (x+1) transformed before performing the analysis. A KFold cross validation (K=7) with the NIPALS method was used to determine optimal factors and variable importance (VIP > 0.8) for each model. Models were then re-run with only explanatory variables that were determined to be significant. To see model validation and diagnostic plots, refer to Supplementary Material 3.

Routine statistical analysis and data visualizations. All routine statistical analyses were performed in R v3.4.0, Python 3.6.1, or JMP Pro13, while multivariate analyses were performed with PC-ORD v6. Graphing was performed in IPython notebook with matplotlib, seaborn,

pandas, and numpy packages. All pairwise comparisons were done using the Kruskal-Wallis nonparametric method, with Dunn's nonparametric multiple comparisons run *post hoc* using a Bonferroni correction.

ACKNOWLEDGEMENTS

We would like to thank Meagan Sims and Keri Kaczor at the Maine Healthy Beaches program and Sean Smith PhD at the University of Maine for their guidance and help with general knowledge of the Wells, ME area and planning of field sampling. Field work, sample processing, and molecular work were assisted by Christine Bunyon, Alexandra Bunda, Jackie Lemaire, Audrey Beresnson, and Joseph Sevigny assisted in optimizing bioinformatic workflows. This work was funded by the National Science Foundation New Hampshire EPSCoR IIA-1330641 grant.

REFERENCES

1. US Environmental Protection Agency. 2012. Recreational water quality criteria. US Environ Prot Agency 1–69. https://www.epa.gov/wqc/2012-recreational-water-quality-criteria

2. Cabelli VJ. 1989. Swimming-associated illness and recreational water quality criteria. Water Sci Technol 2:13–21.

3. Cabelli VJ, Dufour AP, McCabe LJ, Levin MA. 1983. A marine recreational water quality criterion consistent with indicator concepts and risk analysis. J Water Pollut Control Fed 55:1306–1314.

4. Dufour AP. 1984. Bacterial indicators of recreational water quality. Can J Public Health 75:49–56.

5. Colford JM, Wade TJ, Schiff KC, Wright CC, Griffith JF, Sandhu SK, Burns S, Sobsey M, Lovelace G, Weisberg SB. 2007. Water quality indicators and the risk of illness at beaches with nonpoint sources of fecal contamination. Epidemiology 18:27–35. 10.1097/01.ede.0000249425.32990.b9

6. Soller JA, Schoen ME, Bartrand T, Ravenscroft JE, Ashbolt NJ. 2010. Estimated human health risks from exposure to recreational waters impacted by human and non-human sources of faecal contamination. Water Res 44:4674–4691. 10.1016/j.watres.2010.06.049.

7. Soller JA, Schoen ME, Varghese A, Ichida AM, Boehm AB, Eftim S, Ashbolt NJ, Ravenscroft JE. 2014. Human health risk implications of multiple sources of faecal indicator bacteria in a recreational waterbody. Water Res 66:254–264. 10.1016/j.watres.2014.08.026 0043-1354/

8. World Health Organization. 1999. Health-based monitoring of recreational waters: the feasibility of a new approach (the "Annapolis Protocol"). World Health Organization - Geneva 1:1-50. http://www.who.int/iris/handle/10665/66477.

9. Bernhard AE, Field KG. 2000. A PCR assay to discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA. Appl Environ Microbiol 66:4571-4574. 10.1128/AEM.66.10.4571-4574.2000.

10. Dick LK, Simonich MT, Field KG. 2005. Microplate subtractive hybridization to enrich for Bacteroidales genetic markers for fecal source identification. Appl Environ Microbiol 71:3179–3183. 10.1128/AEM.71.6.3179-3183.2005.

11. Schriewer A, Goodwin KD, Sinigalliano CD, Cox AM, Wanless D, Bartkowiak J, Ebentier DL, Hanley KT, Ervin J, Deering LA, Shanks OC, Peed LA, Meijer WG, Griffith JF, SantoDomingo J, Jay JA, Holden PA, Wuertz S. 2013. Performance evaluation of canine-associated Bacteroidales assays in a multi-laboratory comparison study. Water Res 47:6909–6920. 10.1016/j.watres.2013.03.062.

12. Reischer GH, Kavka GG, Kasper DC, Winter C, Mach RL, Farnleitner AH. 2008. Applicability of DNA based quantitative microbial tracking (QMST) evaluated on a large scale in the Danube River and its important tributaries. Large Rivers 18:117–125. 10.1127/lr/18/2008/117.

13. Ebentier DL, Hanley KT, Cao Y, Badgley BD, Boehm AB, Ervin JS, Goodwin KD, Gourmelon M, Griffith JF, Holden PA, Kelty CA, Lozach S, McGee C, Peed LA, Raith M, Ryu H, Sadowsky MJ, Scott EA, Domingo JS, Schriewer A, Sinigalliano CD, Shanks OC, Van De Werfhorst LC, Wang D, Wuertz S, Jay JA. 2013. Evaluation of the repeatability and reproducibility of a suite of qPCR-based microbial source tracking methods. Water Res 47:6839–6848. 10.1016/j.watres.2013.01.060.

644 14. Fremaux B, Gritzfeld J, Boa T, Yost CK. 2009. Evaluation of host-specific Bacteroidales
645 16S rRNA gene markers as a complementary tool for detecting fecal pollution in a prairie
646 watershed. Water Res 43:4838–4849. 10.1016/j.watres.2009.06.045.

647 15. Sinigalliano CD, Ervin JS, Van De Werfhorst LC, Badgley BD, Ballesté E, Bartkowiak J,
648 Boehm AB, Byappanahalli M, Goodwin KD, Gourmelon M, Griffith J, Holden PA, Jay J,
649 Layton B, Lee C, Lee J, Meijer WG, Noble R, Raith M, Ryu H, Sadowsky MJ, Schriewer
650 A, Wang D, Wanless D, Whitman R, Wuertz S, Santo Domingo JW. 2013. Multi-
651 laboratory evaluations of the performance of Catellicoccus marimammalium PCR assays
652 developed to target gull fecal sources. Water Res 47:6883–6896.
653 10.1016/j.watres.2013.02.059.

654 16. Ahmed W, Harwood VJ, Nguyen K, Young S, Hamilton K, Toze S. 2016. Utility of
655 Helicobacter spp. associated GFD markers for detecting avian fecal pollution in natural
656 waters of two continents. Water Res 88:613–622. 10.1016/j.watres.2015.10.050.

657 17. Ahmed W, Goonetilleke A, Powell D, Chauhan K, Gardner T. 2009. Comparison of
658 molecular markers to detect fresh sewage in environmental waters. Water Res 43:4908–
659 4917. 10.1016/j.watres.2009.09.047.

660 18. Schriewer A, Miller WA, Byrne BA, Miller MA, Oates S, Conrad PA, Hardin D, Yang
661 HH, Chouicha N, Melli A, Jessup D, Dominik C, Wuertz S. 2010. Presence of
662 Bacteroidales as a predictor of pathogens in surface waters of the central California coast.
663 Appl Environ Microbiol 76:5802–5814. 10.1128/AEM.00635-10.

664 19. Flood C, Ufnar J, Wang S, Johnson J, Carr M, Ellender R. 2011. Lack of correlation
665 between enterococcal counts and the presence of human specific fecal markers in

Mississippi creek and coastal waters. Water Res 45:872–878. 10.1016/j.watres.2010.09.026.

20. Shibata T, Solo-Gabriele HM, Sinigalliano CD, Gidley ML, Plano LRW, Fleisher JM, Wang JD, Elmir SM, He G, Wright ME, Abdelzaher AM, Ortega C, Wanless D, Garza AC, Kish J, Scott T, Hollenbeck J, Backer LC, Fleming LE. 2010. Evaluation of conventional and alternative monitoring methods for a recreation marine beach with non-point source of fecal contamination. Environ Sci Technol 44:8175–8181. 10.1021/es100884w.

21. Santoro AE, Boehm AB. 2007. Frequent occurrence of the human-specific Bacteroides fecal marker at an open coast marine beach: relationship to waves, tides and traditional indicators. Environ Microbiol 9:2038–2049. 10.1111/j.1462-2920.2007.01319.x.

22. Harwood VJ, Staley C, Badgley BD, Borges K, Korajkic A. 2014. Microbial source tracking markers for detection of fecal contamination in environmental waters: Relationships between pathogens and human health outcomes. FEMS Microbiol Rev 38:1–40. 10.1111/1574-6976.12031.

23. Byappanahalli MN, Nevers MB, Korajkic A, Staley ZR, Harwood VJ. 2012. Enterococci in the environment. Microbiol Mol Biol Rev 76:685–706. 10.1128/MMBR.00023-12.

24. Anderson KL, Whitlock JE, Valerie J, Harwood VJ. 2005. Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. Appl Environ Microbiol 71:3041–3048. 10.1128/AEM.71.6.3041–3048.2005.

686 25. Haller L, Amedegnato E, Poté J, Wildi W. 2009. Influence of freshwater sediment
687 characteristics on persistence of fecal indicator bacteria. Water Air Soil Pollut 203:217-
688 227. 10.1007/s11270-009-0005-0.

689 26. Badgley BD, Thomas FIM, Harwood VJ. 2010. The effects of submerged aquatic
690 vegetation on the persistence of environmental populations of Enterococcus spp. Environ
691 Microbiol 12:1271–1281. 10.1111/j.1462-2920.2010.02169.x.

692 27. Jeng HC, Sinclair R, Daniels R, Englande AJ. 2005. Survival of Enterococci facalis in
693 estuarine sediments. Int J Environ Stud 62:283–291. 10.1080/0020723042000275132.

694 28. Craig DL, Fallowfield HJ, Cromar NJ. 2002. Enumeration of faecal coliforms from
695 recreational coastal sites: evaluation of techniques for the separation of bacteria from
696 sediments. J Appl Microbiol 93:557–565. 10.1046/j.1365-2672.2002.01730.x.

697 29. Ferguson DM, Moore DF, Getrich MA, Zhowandai MH. 2005. Enumeration and
698 speciation of enterococci found in marine and intertidal sediments and coastal water in
699 southern California. J Appl Microbiol 99:598–608. 10.1111/j.1365-2672.2005.02660.x.

700 30. Niewolak S. 1998. Total viable count and concentration of enteric bacteria in bottom
701 sediments from the Czarna Hancza River, Northeast Poland. Polish J Environ Stud
702 7:295–306.

703 31. Jamieson RC, Gordon RJ, Sharples KE, Stratton GW, Madani A. 2002. Movement and
704 persistence of fecal bacteria in agricultural soils and subsurface drainage water: a review.
705 Can Biosyst Eng 44:1-9.

32. Desmarais TR, Solo-Gabriele HM, Carol J, Palmer CJ. 2002. Influence of soil on fecal indicator organisms in a tidally influenced subtropical environment. Appl Environ Microbiol 68:1165–1172. 10.1128/AEM.68.3.1165–1172.2002.

33. Byappanahalli MN, Roll BM, Fujioka RS. 2012. Evidence for occurrence, persistence, and growth potential of Escherichia coli and Enterococci in Hawaii's soil environments. Microbes Environ 27:164–170. 10.1264/jsme2.ME11305.

34. Fujioka R, Sian-Denton C, Borja M, Castro J, Morphew K. 1998. Soil: the environmental source of Escherichia coli and Enterococci in Guam's streams. Environ Toxicol 6:185-195.

35. Jamieson RC, Joy DM, Lee H, Kostaschuk R, Gordon RJ. 2005. Resuspension of sediment-associated Escherichia coli in a natural stream. J Environ Qual 34:581–589. 10.2134/jeq2005.0581.

36. Hathaway JM, Hunt WF. 2011. Evaluation of first flush for indicator bacteria and total suspended solids in urban stormwater runoff. Water Air Soil Pollut 217:135–147. 10.1007/s11270-010-0574-y.

37. Knights D, Kucyznski J, Charlson, Zaneveld ES, Mozer J, Mozer MC, Collman RG, Bushman FD, Knight R, Kelley ST. 2013. Bayesian community-wide culture-independent microbial source tracking. Nat Methods 8:761–763. 10.1038/NMETH.1650.

38. Kaczor K, Sims M. 2014. Maine healthy beaches 2014 Report to US EPA April 1, 2015. Maine Sea Grant Publications 1-17. http://digitalcommons.library.umaine.edu/seagrant_pub/41/

39. Grant-Whiting K, Dalton C, Dillon F. 2003. Microbial source tracking in two sourthern Maine watersheds - Webhannet River watershed report. Maine Sea Grant Publications 1-11. https://www.seagrant.umaine.edu/extension/microbial-source-tracking

40. Roy S, Gerard B, Smith SM, McGreavy B. 2017. A runoff-based vulnerability analysis to examine and communicate the dynamics of bacteria pollution events in the Gulf of Maine. Maine Sustainability & Water Conference, 2017 {presentation unpublished}. https://umaine.edu/mitchellcenter/2017-conference/h-understanding-managing-land-sea-connections-along-maine-coast/

41. Piscataqua Region Estuaries Partnership. 2017. State of our estuaries report 2018. PREP Publications 1-50. http://scholars.unh.edu/prep/391

42. Characklis GW, Dilts MJ, Simmons OD, Likirdopulos CA, Krometis LAH, Sobsey MD. 2005. Microbial partitioning to settleable particles in stormwater. Water Res 39:1773–1782. 10.1016/j.watres.2005.03.004

43. Mote BL, Turner JW, Lipp EK. 2012. Persistence and growth of the fecal indicator bacteria enterococci in detritus and natural estuarine plankton communities. Appl Environ Microbiol 78:2569–2577. 10.1128/AEM.06902-11.

44. Boehm AB, Sassoubre LM. 2014. Enterococci as indicators of environmental fecal contamination. In: enterococci from commensals to lead causes of drug resist infections. Boston: Massachusetts Eye and Ear Infirmary 1–21. https://www.ncbi.nlm.nih.gov/books/NBK190421/

747 45. Talke SA, Swart HE De, Schuttelaars HM. 2009. Feedback between residual circulations
748 and sediment distribution in highly turbid estuaries: an analytical model. Cont Shelf Res
749 29:119–135. 10.1016/j.csr.2007.09.002

750 46. Schinner T, Letzner A, Liedtke S, Castro FD, Eydelnant IA, Tufenkji N. 2010. Transport
751 of selected bacterial pathogens in agricultural soil and quartz sand. Water Res 44:1182–
752 1192. 10.1016/j.watres.2008.11.038.

753 47. Mayer RE, Vierheilig J, Egle L, Reischer GH, Saracevic E, Mach RL, Kirschner AKT,
754 Zessner M, Sommer R, Farnleitner AH. 2015. Automated sampling procedures supported
755 by high persistence of bacterial fecal indicators and Bacteroidetes genetic microbial
756 source tracking markers in municipal wastewater during short-term Storage at 5°C. Appl
757 Environ Microbiol 81:5134–5143. 10.1128/AEM.00998-15

758 48. Sowah RA, Habteselassie MY, Radcliffe DE, Bauske E, Risse M. 2017. Isolating the
759 impact of septic systems on fecal pollution in streams of suburban watersheds in Georgia,
760 United States. Water Res 108:330–338. 10.1016/j.watres.2016.11.007

761 49. Bushon RN, Brady AMG, Christensen ED, Stelzer EA. 2017. Multi-year microbial
762 source tracking study characterizing fecal contamination in an urban watershed. Water
763 Environ Res 89:127-143. 10.2175/106143016X14798353399412

764 50. Boehm AB, Soller JA, Shanks OC. 2015. Human-associated fecal quantitative
765 polymerase chain reaction measurements and simulated risk of gastrointestinal illness in
766 recreational waters contaminated with raw sewage. Environ Sci Technol Lett 2:270–275.
767 10.1021/acs.estlett.5b00219

768 51. Ahmed W, Hamilton KA, Lobos A, Hughes B, Staley C, Sadowsky MJ, Harwood VJ.
769 2018. Quantitative microbial risk assessment of microbial source tracking markers in
770 recreation water contaminated with fresh untreated and secondary treated sewage.
771 Environ Int 117:243-249. 10.1016/j.envint.2018.05.012

772 52. Wong M, Kumar L, Jenkins TM, Xagoraraki I, Phanikumar MS, Rose JB. 2009.
773 Evaluation of public health risks at recreational beaches in Lake Michigan via detection
774 of enteric viruses and a human-specific bacteriological marker. Water Res 43:1137–1149.
775 10.1016/j.watres.2008.11.051

776 53. Green HC, Dick LK, Gilpin B, Samadpour M, Field KG. 2012. Genetic markers for rapid
777 PCR-based identification of gull, Canada goose, duck, and chicken fecal contamination in
778 water. Appl Environ Microbiol 78:503–510. 10.1128/AEM.05734-11

779 54. Staley C, Gould TJ, Wang P, Phillips J, Cotner JB, Sadowsky MJ. 2016. Sediments and
780 soils act as reservoirs for taxonomic and functional bacterial diversity in the upper
781 Mississippi River. Microb Ecol 71:814-824. 10.1007/s00248-016-0729-5

782 55. Meybeck M, Laroche L, Dürr HH, Syvitski JPM. 2003. Global variability of daily total
783 suspended solids and their fluxes in rivers. Glob Planet Change 39:65–93.
784 10.1016/S0921-8181(03)00018-3

785 56. Orton PM, Kineke GC. 2001. Comparing calculated and observed vertical suspended-
786 sediment distributions from a Hudson River estuary turbidity maximum. Estuar Coast
787 Shelf Sci 52:401–410. 10.1006/ecss.2000.0747

788 57. Carrascal LM, Galván I, Gordo O. 2009. Partial least squares regression as an alternative
789 to current regression methods used in ecology. Oikos 118:681–690. 10.1111/j.1600-
790 0706.2008.16881.x

791 58. Thoe W, Gold M, Griesbach A, Grimmer M, Taggart ML, Boehm AB. 2014. Predicting
792 water quality at Santa Monica Beach: evaluation of five different models for public
793 notification of unsafe swimming conditions. Water Res 67:105–117.
794 10.1016/j.watres.2014.09.001

795 59. Brooks WR, Fienen MN, Corsi SR. 2013. Partial least squares for efficient models of
796 fecal indicator bacteria on Great Lakes beaches. J Environ Manage 114:470–475.
797 10.1016/j.jenvman.2012.09.033

798 60. Messer JW, Dufour AP. 2002. Method 1600: Enterococci in water by membrane
799 filtration using membrane-Enterococcus indoxyl-β-D-glucoside agar (mEI). US Environ
800 Prot Agency 1-42.

801 61. Crump BC, Baross JA, Simenstad CA. 1998. Dominance of particle-attached bacteria in
802 the Columbia River estuary, USA. Aquat Microb Ecol 14:7–18. 10.3354/ame014007.

803 62. US Environmental Protection Agency. 1983. Methods for chemical analysis of water and
804 wastes - method 160.2 (gravimetric, dried at 103-105°C). US Environ Prot Agency 59-
805 61.

806 63. Sauer EP, Vandewalle JL, Bootsma MJ, Mclellan SL. 2011. Detection of the human
807 specific Bacteroides genetic marker provides evidence of widespread sewage
808 contamination of stormwater in the urban environment. Water Res 45:4081–4091.
809 10.1016/j.watres.2011.04.049.

64. Cloutier DD, Alm EW, McLellan SL. 2015. The influence of land-use, nutrients, and geography on microbial communities and fecal indicator abundance at Lake Michigan beaches. Appl Environ Microbiol 81:4904-4913. 10.1128/AEM.00233-15.

65. Newton RJ, Bootsma MJ, Morrison HG, Sogin ML, McLellan SL. 2013. A microbial signature approach to identify fecal pollution in the waters off an urbanized coast of Lake Michigan. Microb Ecol 65:1011–1023. 10.1007/s00248-013-0200-9.

66. Bernhard AE, Field KG. 2000. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. Appl Environ Microbiol 66:1587–1594. 10.1128/AEM.66.4.1587-1594.2000.

67. Lu J, Santo Domingo JW, Lamendella R, Edge T, Hill S. 2008. Phylogenetic diversity and molecular detection of bacteria in gull feces. Appl Environ Microbiol 74:3969-3976. 10.1128/AEM.00019-08.

68. Harwood VJ, Brownell M, Wang S, Lepo J, Ellender RD, Ajidahun A, Hellein KN, Kennedy E, Ye X, Flood C. 2009. Validation and field testing of library-independent microbial source tracking methods in the Gulf of Mexico. Water Res 43:4812–4819. 10.1016/j.watres.2009.06.029.

69. Shanks OC, Nietch C, Simonich M, Younger M, Reynolds D, Field KG. 2006. Basin-wide analysis of the dynamics of fecal contamination and fecal source identification in Tillamook Bay, Oregon. Appl Environ Microbiol 72:5537-5536. 10.1128/AEM.03059-05

70. Jokinen CC, Schreier H, Mauro W, Taboada E, Isaac-Renton JL, Topp E, Edge T, Thomas JE, Gannon VPJ. 2010. The occurrence and sources of Campylobacter spp.,

Salmonella enterica and Escherichia coli O157:H7 in the Salmon River, British Columbia, Canada. J Water Health 8:374–386. 10.2166/wh.2009.076.

71. Layton A, McKay L, Williams D, Garrett V, Gentry R, Sayler G. 2006. Development of Bacteroides 16S rRNA gene taqman-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. Appl Environ Microbiol 72:4214–4224. 10.1128/AEM.01036-05.

72. Converse RR, Blackwood AD, Kirs M, Griffith JF, Noble RT. 2009. Rapid QPCR-based assay for fecal Bacteroides spp. as a tool for assessing fecal contamination in recreational waters. Water Res 43:4828–4837. 10.1016/j.watres.2009.06.036.

73. Green HC, Dick LK, Gilpin B, Samadpour M, Field KG. 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. 10.1128/AEM.05734-11

74. Haugland RA, Varma M, Sivaganesan M, Kelty C, Peed L, Shanks OC. 2010. Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected Bacteroidales species and human fecal waste by qPCR. Syst Appl Microbiol 33:348-357. 10.1016/j.syapm.2010.06.001.

75. Nordstrom JL, Vickery MCL, Blackstone GM, Murray SL, DePaola A. 2007. Development of a multiplex real-time PCR assay with an internal amplification control for the detection of total and pathogenic Vibrio parahaemolyticus bacteria in oysters. Appl Environ Microbiol 73:5840–5847. 10.1128/AEM.00460-07.

76. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of

sequences per sample. Proc Natl Acad Sci U S A 108:4516–4522. 10.1073/pnas.1000080107.

77. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6:1621–1624. 10.1038/ismej.2012.8.

78. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Gonzalez Pena A, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turn KR. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 5:335–336. doi:10.1038/nmeth.f.303.

79. Bolger AM, Lohse M, Usadel B. 2014. Genome analysis Trimmomatic : a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. 10.1093/bioinformatics/btu170

80. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072. 10.1128/AEM.03006-05

81. Henry R, Schang C, Coutts S, Kolotelo P, Prosser T, Crosbie N, Grant T, Cottam D, O'Brien P, Deletic A. 2016. Into the deep: Evaluation of SourceTracker for assessment of faecal contamination of coastal waters. Water Res 93:242-253. 10.1016/j.watres.2016.02.029

82. Helland IS. 1990. Partial least squares regression and statistical models. Scand Stat Theory Appl 17:97–114. http://www.jstor.org/stable/4616159.

Tables

Environmental Microbial Community Source (Including Marine Beach Source)

Water Sample Type	Marine Beach	Freshwater	Estuarine Sediment	Sediment	Soil
Estuary Beach	97%	<0.01%	0.4%	<0.01%	<0.01%
Estuary	94%	2.9%	0.2%	0.02%	<0.01%
Freshwater	<0.01%	N/A	<0.01%	74%	2.6%

Environmental Microbial Community Source (Excluding Marine Beach Source)

Water Sample Type	Marine Beach	Freshwater	Estuarine Sediment	Sediment	Soil
Estuary Beach	N/A	66%	12%	<0.01%	<0.01%
Estuary	N/A	74%	7.6%	0.02%	<0.01%
Marine Beach	N/A	35%	19%	<0.01%	<0.01%
Freshwater	N/A	N/A	0%	74%	2.6%

Table 1. **The relative contribution of different sources to the microbial communities in estuarine and marine water**. SourceTracker was run with two different configurations, one where Marine Beach water was included as a potential source (top) and a second run where Marine Beach water was excluded as a potential source (bottom).

Freshwater		Estuary, Estuary Beach & Marine Beach			
PLSR 1		PLSR 1		PLSR 2	
X Variable	**Loading**	**X Variable**	**Loading**	**X Variable**	**Loading**
Particle ENT	0.501	Particle ENT	0.456	Particle ENT	0.420
qPCR Mammal	0.352	qPCR Mammal	0.438	qPCR Mammal	-0.337
TSS	0.408	% Freshwater	0.408	% Freshwater	-0.418
% Sediment	0.336	%Unknown	-0.457	% Unknown	0.389
% Unknown	0.476	Water Temp (C)	0.302	Water Temp (C)	-0.123
Salinity	-0.344	Hightide (ft)	0.170	Hightide (ft)	0.456
		% Estuarine Sediment	-0.294	% Estuarine Sediment	0.401
Total Y Variance	60.1%	Total Y Variance	47.2%	Cumulative Y Variance	61.8%

Table 2. Most significant relationships/contributions for all factors to enterococci concentrations. Shown is the output from a partial least squares regression for a freshwater and estuary/marine model. All variables shown have significant relationships for each model (VIP > 0.8), and loadings are derived from re-running models with only variables deemed significant. Model loadings are specific weights on a multivariate regression axis, positive and negative loadings refer to positive or negative relationships to enterococci concentrations. Negative loadings in the model are designated with a – before the number.

PCR Assay	Source	Taxa	Target (Size)	Primer/Probes 5' -> 3'
Bac32	Mammal	*Bacteroides-Provotella*	*16S rRNA* (696bp)	Bac32:AACGCTAGCTACAGGCTT Bac708:CAATCGGAGTTCTTCGTG
HF183	Human	Human Cluster *Bacteroides-Provotella*	*16S rRNA* (541bp)	HF183:ATCATGAGTTCACATGTCCG Bac708:CAATCGGAGTTCTTCGTG
CF128	Ruminant	Ruminant Cluster *Bacteroides-Provotella*	*16S rRNA* (595bp)	CF128:CCAACYTTCCCGWTACTC Bac708R:CAATCGGAGTTCTTCGTG
DF475	Dog	Dog Cluster *Bacteroides-Provotella*	*16S rRNA* (251bp)	DF475:CGCTTGTATGTACCGGTACG Bac708:CAATCGGAGTTCTTCGTG
Gull2	Gulls	*Catellicoccus marimammalium*	*16S rRNA* (412bp)	Gull2F:TGCATCGACCTAAAGTTTTGAG Gull2R:GTCAAAGAGCGAGCAGTTACTA
qPCR Assays	**Source**	**Taxa**	**Target (Size)**	**Primer/Probes 5'->3'**
AllBac TaqMan	Mammal	*Bacteroides-Provotella*	*16S rRNA* (108bp)	AllBac296f: GAGAGGAAGGTCCCCCAC AllBac412r: CGCTACTTGGCTGGTTCAG AllBac375Bhqr:(FAM) TGAAGGATGAAGGTTCTATGGATTGTAA ACTT (BHQ-1)
HF183 TaqMan	Human	Human Cluster *Bacteroides-Provotella*	*16S rRNA* (167bp)	HF183:ATCATGAGTTCACATGTCCG BDFRev:CGTAGGAGTTTGGACCGTGT BFDFAM:(FAM) CTGAGAGGAAGGTCCCCCACATTGGA (BHQ-1)
GFD SYBR	Avian	Unclassified *Helicobacter spp.*	*16S rRNA* (123bp)	F:TCGGCTGAGCACTCTAGGG R:GCGTCTCTTTGTACATCCCA

Table 3. Primers and probes for MST PCR and qPCR assays.

904 **Figure Legends**

905 **Figure 1: Wells Maine Study Area and Sampling Sites.** All water collection sites are marked
906 with a dark grey circle. Sites that correspond to fresh water are indicated with a (1), estuary (2),
907 estuary beach (3), and marine beach (4). Soil and sediment sites are represented with a star and
908 estuarine sediment sites are shown with a triangle. Map created with ArcGIS Online (using the
909 Light Gray Canvas Map; sources: Esri, DeLorme, HERE, MapmyIndia).

910

911 **Figure 2: Geometric Mean Concentrations of Total and Particle Associated Enterococci**
912 **and Average Total Suspended Solids Concentrations at the Eight Study Sites.** (A) Total
913 enterococci concentrations are represented with the blue bar, and particle associated enterococci
914 concentrations correspond to the green bar. Error bars are derived from variation from each site
915 across the entire study. (B) Violin plots were used to represent TSS concentrations, and the color
916 corresponds to the type of site including marine beach (red), estuary beach (purple), estuary
917 (green), or fresh water (blue). Horizontal lines go through the median of each violin plot.

918

919 **Figure 3: Relative Levels of Mammal, Human, and Bird Fecal Source at the Different**
920 **Types of Study Sites.** Box plots represent levels of microbial source tracking markers at marine
921 beach (Wells Beach), estuary beach (W11, W12, W13), estuary (W14 & W15), and fresh water
922 (Webhannet & Depot). Outlier data are represented with a black diamond.

923

924 **Figure 4: 16S Taxa Profiles and the Top Three Most Abundant Bacterial Classes in All**
925 **Source and Sink Samples.** Stacked bar plots represent percentages of the class level
926 composition of the microbial communities. Source corresponds to environmental sources that
927 were finger-printed with the SourceTracker program, and then used to determine their presence
928 within water (sink) samples. The table represents the top three classes for each group of samples
929 and * corresponds to phylum level. For a complete list of all taxa assignments refer to
930 Supplementary material 4.

931

932 **Figure 5. Differences Between Microbial Communities from Different Source Materials.**
933 Samples are color-coded based on sample matrix (i.e. soil, fresh water, etc.). Percent of variation
934 explained are displayed on the x and y axis and the minimum stress of the ordination is shown in
935 the top left corner.

936