Effects of Fecal Source Input, Environmental Conditions, and Environmental Sources on Enterococci Concentrations in a Coastal Ecosystem

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 in the Northeast, USA requires management efforts to address public health and economic concerns. Concentrations of fecal-borne bacteria 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 for multiple sources. 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. We found elevated enterococci levels are 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 fecal tracts 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 sources due to the lack of an "inter-species barrier" for diseases and the higher density of human pathogens that humans can carry (6–8). Although human pollution represents the greatest public health risk, other fecal sources that contain enterococci and possibly human pathogens can be chronic or intermittent sources of both, making beach water quality management and remediation efforts more complex.

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

Due to the pervasiveness of enterococci in natural ecosystems, recent studies have been 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

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

144

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

however 8 of the 10 positive samples were detected in July (8/13 = 61%). For all other sites, an increase in the detection of dog fecal marker also occurred during July, with 44% (16/36) detection, compared to 0% for August and September and <1% for June. Thus, most of the dog contamination at all sites was associated with unknown dog-related conditions during July.

Concentrations of mammal, human, and bird fecal sources. We used qPCR to provide relative quantitative measures of mammal, human and bird fecal contamination levels. Water at estuary and estuary beach sites contained significantly higher levels of mammal (AllBac) fecal marker copies, with an average of 1.54×10^7 compared to 2.62×10^6 in freshwater and 3.9×10^6 copies/100 ml in marine beach ($p < 0.05$). Average concentrations of human (HF183) and bird (GFD) fecal markers were not statistically different between sites, however, concentrations of the human marker in individual samples varied from 0 - 2.04×10^4 copies/100 ml (Figure 3), while bird fecal marker concentrations were relatively stable across all sites. No significant temporal trends were observed for any of the quantitative fecal marker levels. Compared with presence/absence detection of fecal sources, quantitative measurements also did not show strong spatial patterns, except mammal marker levels showed significant increases at estuary and estuary beach sites compared to marine and freshwater sites.

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

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

199 underlying sediment (74%) and much lower probability of taxa originating from soil (2.6%).
200 Initial results for the estuary and estuary beach indicated that marine beach water was the
201 dominant source of bacteria (Table 1). However, given that likely fecal sources are coming from
202 the watershed, we excluded marine beach water as a potential source and included it as a sink
203 then re-analyzed the data. These second results showed that freshwater taxa had a high
204 probability of being a significant fraction of estuary (73%), estuary beach (66%) and marine
205 beach (35%) water communities, with a significantly higher percentage for the estuary locations
206 compared to the marine beach (Table 1, $p < 0.05$), which is more influenced by ocean microbial
207 taxa. Despite the significant percentage of freshwater taxa assignments in the estuary, estuary
208 beach, and marine beach waters there were no freshwater sediment or soil taxa assignments for
209 these sites. The data for the percent of unidentifiable taxa showed the opposite trend compared
210 to percent of assigned freshwater taxa. Unidentifiable taxa in the marine beach were significantly
211 higher (46%; $p < 0.05$), which is not surprising given that marine beach water community would
212 likely be most influenced by non-terrestrial sources. Estuarine sediment was the highest likely
213 identified source in the water from the marine beach site (19%), and it was significantly higher
214 than percentages calculated for all estuary sites ($p < 0.05$). Overall results showed that freshwater
215 source-related taxa were a pervasive source throughout the estuary and marine beach, and while
216 sediment source-related taxa were highly abundant in the freshwater they were not observed
217 within the estuary or marine beach.

218

219 **Relationships between environmental conditions, fecal source concentrations,**
220 **environmental sources and enterococci concentrations**. Two PLSR models were created to
221 determine relationships between enterococci and fecal source concentrations, environmental

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

245 freshwater model (Supplementary Table 3). Out of all the potential variables measured (19 total)
246 across three categories (fecal source input, environmental source contribution, and environmental
247 conditions), particle-associated enterococci and mammal fecal marker concentrations had the
248 most significant relationships to enterococci concentrations. The relationships between other
249 variables and enterococci concentrations were specific to freshwater and estuary/marine beach
250 models, indicating ecosystem specific relationships. However, the joint relationship of particle-
251 associated and mammal fecal marker across freshwater and estuary/marine environments
252 indicate their overarching importance in determining enterococci concentrations.

253

254 **4 Discussion:**

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

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

275

276 Enterococci were significantly associated with suspended particles of >3.0 µm diameter (R^2 =
277 0.96, $p < 0.05$). On average, 36% (SD ± 30) of the total enterococci concentrations were
278 associated with particles, which suggests particles as a potentially important transport
279 mechanism. Other studies conducted in estuary and storm waters have found similar fractions of
280 particle associated enterococci, but they noted enterococci demonstrated a preference for a larger
281 particle size of >30 µm (42–44). The large standard deviation for particle-associated enterococci
282 could be attributed to the complex nature of particle interactions (sedimentation rate,
283 electrostatic, hydrophobic, and other surface-surface interactions) and hydrogeological dynamics
284 (salinity-driven turbidity maximum) (45). The mechanisms underlying enterococci-particle
285 interactions may also be related to ionic strength in surface waters, as *Enterococcus faecalis* is
286 negatively charged over a broad pH range (2-8 pH units) and in the presence of different ion
287 concentrations (46). Results for this study indicate that TSS and particle-associated enterococci
288 had no linear relationship, indicating particle-associated enterococci were not dependent on the
289 total amount of suspended material and thus the association is likely due to other factors
290 influencing cell-particle interactions.

291

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

308

309 The average concentration of the human marker was 1,500 copies/100 ml across all sites
310 (geometric mean 167 copies/100 ml), with the highest concentration being 20,364 copies/100ml
311 (Webhannet 6/22/16). Boehm et al. (50) showed that 4,200 copies/100 ml of HF183 is the cutoff
312 for where GI illnesses exceed the EPA acceptable risk level of approximately 30/1000 for
313 swimmers (1). 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 though sites on average were below published thresholds, detection of human contamination even at low concentrations is a concern.

Although human fecal sources are the greatest public health concern (6, 7, 22, 51) we did not observe any relationship between human fecal contamination and enterococci concentrations, suggesting other mammalian fecal sources are more influential in explaining the variation observed in this study. Interestingly gull fecal sources were detected in 77% or more of the samples in the estuary and marine beach area, however only 10% of the samples were positive within the fresh water (Supplementary Material 1), despite there being no decrease in the bird fecal marker concentration, suggesting the presence of different bird sources in these areas.

337 Anecdotally, Canada geese were observed upstream of both the Webhannet and Depot
338 freshwater sites periodically throughout the season, which could be a significant source of bird
339 fecal contamination in the fresh water locations (52).

340

341 One of the unique findings of this study was the relative contribution of different sources to the
342 bacterial community in the estuarine water. The bacterial community in estuarine water primarily
343 originated (>90%) from marine beach water, which is not surprising for a well-flushed estuary
344 like the study site. Because the study period was minimally influenced by rainfall and associated
345 runoff of freshwater, we expected that the influence of freshwater sources would be low. In
346 ensuing analyses, we chose not to include marine beach water as a potential source for a variety
347 of reasons. First, the samples were always collected during low tide before the ebb when the
348 estuary water was draining and water was moving from the watershed towards the marine beach.
349 Secondly, we had already shown that the OTU compositions for the marine beach and estuary
350 samples were very similar, increasing the possibility of a type I error (false positive) for
351 identifying marine beach as the likely source of enterococci. Lastly, fecal pollution sources most
352 likely come from the watersheds and not from marine water, so excluding marine beach water
353 helps to enhance the determination of watershed influences. Our second analysis (marine beach
354 source excluded) showed that freshwater was a significant source of bacteria to the estuary
355 (>65% assignment) compared to soil, sediment, and estuarine sediment. This implicates
356 freshwater as a major conduit for bacterial transport, as well as the major source of enterococci
357 to the estuary. Overall this finding highlights the importance of freshwater discharge as a
358 controlling factor in transporting contamination from the watershed to the coast. The specific

359 percent assignment of freshwater source could be an over-estimate, however the trend observed
360 is a likely scenario given the rational discussed.

361

362 Analysis of environmental reservoirs of enterococci (soil, sediment, etc.) and their presence
363 within water samples using SourceTracker revealed a variety of source contributions to
364 freshwater, estuary and marine waters. To date there have been limited studies using
365 SourceTracker to identify soil and sediment-associated taxa within water samples, and none of
366 these studies have focused on a coastal watershed with the potential for freshwater, estuarine and
367 marine sources. One study conducted in the upper Mississippi River identified up to 14% of
368 sediment and 1.4% of soil sources of the taxa within the river water (53). This study, however
369 showed that the sediment source was much more abundant in freshwater (74%), indicating a
370 greater degree of mixing between the freshwater and underlying sediment communities. The
371 amount of sediment and soil sources within water samples may be related to site specific
372 characteristics such as relief or soil texture, which has been shown with TSS fluxes on a global
373 scale (54). Thus, the degree to which the underlying sediment community mixes with the
374 overlaying water is likely site specific. Interestingly, even though freshwater contained a
375 significant amount of sediment source taxa, no sediment source was observed at the estuary and
376 marine beach sites through the SourceTracker analysis. This difference could indicate that rapid
377 sedimentation happens during transit to and within the estuary and at the estuarine turbidity
378 maximum zone (55). TSS concentrations and the ratio of particle-associated to total enterococci
379 concentrations, however, showed no differences between freshwater and estuary/marine sites.
380 This could be related to the separate and quite different hydrodynamics within these different
381 water systems. The percent of sediment source in the freshwater samples observed here might

382 also be an over-estimate/over fit from SourceTracker given the limited number of potential
383 sources used, but results consistently showed an elevated presence of sediment in all freshwater
384 samples in this study. SourceTracker analysis also revealed that the freshwater source was
385 significant (35% or more) in estuary and marine beach water samples, suggesting that fresh
386 water is a significant conduit for microbial, and fecal contamination, transport from the
387 watershed to the estuary and marine beach.

388

389 The use of predictive models for water quality has been a focus in the field in parallel with the
390 adoption of bacterial indicator organisms as the gold standard for water quality determination.
391 The goal of this research was to identify significant influences on enterococci concentrations by
392 measuring a wide variety of variables. To distill this information, we used a PSLR model, which
393 has been shown to out-perform similar multiple linear regression and principle components
394 regression analyses (56) and has gained popularity in the water quality field (57, 58). Results
395 from the PLSR analysis in this study showed that particle-associated enterococci and
396 concentrations of mammal fecal sources were the driving force behind variation in enterococci
397 concentrations, as described by both PLSR models constructed. Other factors were found to
398 influence enterococci concentrations, however, these differed between the freshwater and
399 estuary/marine beach models. For example, TSS concentration as well as the percent of both
400 freshwater sediment and unknown sources positively influenced enterococci concentrations at
401 freshwater sites. This indicates that sediment is a likely source of enterococci that influences
402 concentrations measured in the water. Positive influences from the unidentified source taxa
403 suggests that there is either an alternative source (not measured in this study) within the
404 watershed that also influences enterococci concentrations or that SourceTracker could simply not

405 resolve all the potential sources we used. This finding is not surprising given the vast number of
406 potential sources of fecal pollution within a watershed and that fecal sources were not a part of
407 the SourceTracker analysis. Results from the estuary and marine beach model returned a two-
408 factor regression, with each factor essentially being the inverse of each other. Specifically, it
409 highlighted freshwater being a major conduit for microbial transport to and through the estuary.
410 Negative influences from the unknown source reaffirms this finding, along with positive
411 influences from the previous high tide height. The second factor explained approximately 15% of
412 the variation in enterococci concentration, therefore its importance must be weighed
413 proportionately to factor one, which explained almost 50% of the variation. However, positive
414 loadings from previous high tide height and percent of estuarine sediment indicate estuarine
415 sediment could be a source of enterococci whose influence is dependent on tide height. The
416 negative loadings from mammal fecal source(s) may indicate that enterococci originating from
417 the estuarine sediment are not from mammal fecal sources.

418

419 Overall, the results from this study demonstrated that concentrations of enterococci in the coastal
420 estuarine/marine beach study area were largely controlled by particle-associated enterococci and
421 mammal fecal source input. The influence of these factors is likely universal across freshwater
422 and estuarine environments, however other ecosystem factors likely play a role as well. For
423 freshwater portions of the coastal watershed, sediment may act as a significant enterococci
424 reservoir that is frequently re-suspended within the water column. Freshwater itself could act as a
425 major conduit for bacterial transport to an estuary and marine beach area where other
426 environmental factors (water temperature and high tide height) can influence enterococci
427 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 = 1 marine beach) as well as twelve soil, twelve fresh-water sediment and four estuarine sediment sampling sites. Data for air temperature and rainfall amount for the 48 h prior to sampling were obtained from Weather Underground (https://www.wunderground.com/cgibin/findweather/getForecast?query=Wells,%20ME) and characteristics of tides during sampling were obtained from US Harbors (www.meusharbors.com).

Water sampling. Surface water samples were collected weekly from June to September 2017 (n = 117). Sampling started two hours before low tide to maximize the potential impacts of freshwater pollution sources, and samples from all estuary and marine beach sites were collected before the slack tide. Water samples were collected in autoclaved 1L Nalgene™ Wide-Mouth Lab Quality PPCO bottles (Thermo Fisher Scientific, Waltham, MA, USA), and environmental parameters were measured with a YSI Pro2030® dissolved oxygen, conductivity, and salinity Instrument (YSI Incorporated, Yellow Springs, Ohio, USA). A field replicate was collected at a different site for each sampling event.

450 **Soil, sediment, and marine sediment collection.** Environmental sources were collected twice
451 throughout the sampling season to build source libraries that were "finger-printed" with 16S
452 sequencing and SourceTracker analysis. Six soil and sediment samples were collected upstream
453 of both freshwater sites (Webhannet and Depot; Figure 1). Soil samples were collected at the
454 crest of the stream embankment, where a 10 x 10 cm a plastic square template was placed down
455 and all soil (O-horizon) within the template at a 2 cm depth was collected. Samples were sieved
456 (USA Standard No. 5) to remove any loose-leaf litter and roots to only sample smaller soil
457 particles and their microbes. Underlying stream sediments were collected using a Van Veen
458 sediment sampler from depositional sites chosen based on the presence of fine grain sediments.
459 One grab sample was collected for each site and then the top 2 cm of sediment was subsampled
460 for analysis. Sediments were sieved (USA Standard No. 45) to remove coarse grain and gravel
461 size particles. Estuarine sediments were collected during low tide when intertidal sediments were
462 exposed using the Van Veen sampler, and the top 2 cm were again collected for analysis.

463

464 **Enterococci and total suspended solids quantification.** Total and particle-associated
465 enterococci were enumerated using the EPA Method 1600 membrane filtration protocol (59) and
466 particle-associated enterococci were determined via filtration through a 0.47 mm diameter 3.0
467 µm pore size polycarbonate filter (Millipore™, Darmstadt, Germany) as first reported by Crump
468 et al. (60). The filters were rolled onto plates containing mEI agar and incubated at 41°C ±
469 0.5°C; representative colonies were counted in 24 ± 2 hours. Total suspended solids (TSS) were
470 measured using EPA method 160-2, where 500 ml of the water sample was used to determine
471 TSS concentrations (61).

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 recovery (62–64). To minimize additional DNA loss during the extraction process solutions C2 and C3 (from manufacturer's protocol) were halved in volume and combined into a single step. DNA extraction from soil, freshwater sediment, and marine sediment were conducted per the manufacture's protocol.

Microbial source tracking (MST) PCR and qPCR assays. MST PCR assays that target Mammals (Bac32; 65), Humans (HF183; 9), Gulls (Gull2; 66), Dogs (DF475; 10) and Ruminants (CF128; 9) were used to determine the presence of fecal sources in water samples. Positive control plasmids were created for each PCR assay from fresh fecal samples that came from each target organism (Human, Gull, Dog, and Cow). The TOPO™ TA™ Cloning Kit was used (Invitrogen, Carlsbad, CA, USA), with a blue/white screen of *E. coli* transformants on kanamycin (50 µg/mL) selective TSA plates. Positive *E. coli* colonies were screened with their respective PCR assay, and PCR positive colonies were then grown in TSB and extracted with the PureLink® Quick Plasmid Miniprep Kit (Invitrogen, Carlsbad, CA, USA). PCR assays were run on a T100™ Thermal Cycler (BioRad, Hercules, CA, USA) with the GoTaq® Green MasterMix (Promega, Madison, WI, USA). Cycling conditions and amplification protocols for each assay

targeted the different source specific markers and followed protocols delineated by different studies: Bac32 (67) and HF183 (67), CF128 (68), DF475 (69), and Gull2 (66). Quantitative PCR assays were also run to determine fecal source strength for Mammals (AllBac; 70), Humans (HF183; 71), and Birds (GFD; 72). All qPCR assays were run on a Mx3000P cycler (Agilent Technologies, Santa Clara, CA, USA), TaqMan assays used the PerfecCTa® FastMix® II (QuantaBio, Beverly, MA, USA) master mix and the SYBR green assay used the FastSYBR™ Green Master Mix (Applied Biosystems, Foster City, CA, USA). A standard curve ranging from 10^6-10^2 copies (Mammal assay) or 10^5-10^1 copies (Human & Bird assay) was also run for each experimental run with the limit of quantification (LOQ) being 100 copies (Mammal) or 10 copies (Human & Bird) per PCR. The Ct values, amplification efficiency, slope, and R^2 values for each standard curve were compared to previously run standard curves, to ensure satisfactory performance before being used to calculate copy numbers for that run. Each environmental sample was diluted 1:10 and run in triplicate and the reaction volume (25 µl) contained a final concentration of 0.2 mg/ml BSA. Amplification/cycling conditions were preformed per published protocols for AllBac (73), HF183 (73), and GFD (16). TaqMan assays were run with an internal amplification control (74) with a down-shift of 1 cycle considered inhibition. Samples spiked with a plasmid containing 10^4 copies of GFD amplicon were used as inhibition controls for the SYBR assay, with a recovery of less than 10^4 copies (100%) considered inhibition. For a list of primers, probes, and standard curve performance, see Supplementary Material 1.

16S library preparation. The V4 region of the 16S rRNA gene, using the 515F-806R primer-barcode pairs, was used for amplicon sequencing (75). The Earth Microbiome Project protocol was used for amplification and pooling of samples, with minor modifications (76). The Qubit®

dsDNA HS assay was used to quantify sample concentrations, and 500 ng of DNA was pooled per sample. The pool was then run on a 1.2 % low-melt agarose gel to separate primer-dimers from acceptable product, and bands between 300-350 bps were cut and extracted as described above. The final DNA sample was then run on the Agilent Technologies 2200 TapeStation 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

563 re-run with only explanatory variables that were determined to be significant. To see model
564 validation and diagnostic plots, refer to Supplementary Material 3.

565 **Routine statistical analysis and data visualizations**. All routine statistical analyses were
566 performed in R v3.4.0, Python 3.6.1, or JMP Pro13, while multivariate analyses were performed
567 with PC-ORD v6. Graphing was performed in IPython notebook with matplotlib, seaborn,
568 pandas, and numpy packages. All pairwise comparisons were done using the Kruskal-Wallis
569 nonparametric method, with Dunn's nonparametric multiple comparisons run *post hoc* using a
570 Bonferroni correction.

571

572 **ACKNOWLEDGEMENTS**

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

580

581 **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.

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

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

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

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

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

19. Flood C, Ufnar J, Wang S, Johnson J, Carr M, Ellender R. 2011. Lack of correlation 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.

670 25. Haller L, Amedegnato E, Poté J, Wildi W. 2009. Influence of freshwater sediment
671 characteristics on persistence of fecal indicator bacteria. Water Air Soil Pollut 203:217-
672 227. 10.1007/s11270-009-0005-0.

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

676 27. Jeng HC, Sinclair R, Daniels R, Englande AJ. 2005. Survival of Enterococci facalis in
677 estuarine sediments. Int J Environ Stud 62:283–291. 10.1080/00207230420002751 32.

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

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

684 30. Niewolak S. 1998. Total viable count and concentration of enteric bacteria in bottom
685 sediments from the Czarna Hancza River, Northeast Poland. Polish J Environ Stud
686 7:295–306.

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

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

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

696 34. Fujioka R, Sian-Denton C, Borja M, Castro J, Morphew K. 1998. Soil: the environmental
697 source of Escherichia coli and Enterococci in Guam’s streams. Environ Toxicol 6:185-
698 195.

699 35. Jamieson RC, Joy DM, Lee H, Kostaschuk R, Gordon RJ. 2005. Resuspension of
700 sediment-associated Escherichia coli in a natural stream. J Environ Qual 34:581–589.
701 10.2134/jeq2005.0581.

702 36. Hathaway JM, Hunt WF. 2011. Evaluation of first flush for indicator bacteria and total
703 suspended solids in urban stormwater runoff. Water Air Soil Pollut 217:135–147.
704 10.1007/s11270-010-0574-y.

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

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

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

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

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

721 42. Characklis GW, Dilts MJ, Simmons OD, Likirdopulos CA, Krometis LAH, Sobsey MD.
722 2005. Microbial partitioning to settleable particles in stormwater. Water Res 39:1773–
723 1782. 10.1016/j.watres.2005.03.004

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

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

731 45. Talke SA, Swart HE De, Schuttelaars HM. 2009. Feedback between residual circulations
732 and sediment distribution in highly turbid estuaries: an analytical model. Cont Shelf Res
733 29:119–135. 10.1016/j.csr.2007.09.002

734 46. Schinner T, Letzner A, Liedtke S, Castro FD, Eydelnant IA, Tufenkji N. 2010. Transport
735 of selected bacterial pathogens in agricultural soil and quartz sand. Water Res 44:1182–
736 1192. 10.1016/j.watres.2008.11.038.

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

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

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

748 50. Boehm AB, Soller JA, Shanks OC. 2015. Human-associated fecal quantitative
749 polymerase chain reaction measurements and simulated risk of gastrointestinal illness in
750 recreational waters contaminated with raw sewage. Environ Sci Technol Lett 2:270–275.
751 10.1021/acs.estlett.5b00219

752 51. Wong M, Kumar L, Jenkins TM, Xagoraraki I, Phanikumar MS, Rose JB. 2009.
753 Evaluation of public health risks at recreational beaches in Lake Michigan via detection
754 of enteric viruses and a human-specific bacteriological marker. Water Res 43:1137–1149.
755 10.1016/j.watres.2008.11.051

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

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

762 54. Meybeck M, Laroche L, Dürr HH, Syvitski JPM. 2003. Global variability of daily total
763 suspended solids and their fluxes in rivers. Glob Planet Change 39:65–93.
764 10.1016/S0921-8181(03)00018-3

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

768 56. Carrascal LM, Galván I, Gordo O. 2009. Partial least squares regression as an alternative
769 to current regression methods used in ecology. Oikos 118:681–690. 10.1111/j.1600-
770 0706.2008.16881.x

771 57. Thoe W, Gold M, Griesbach A, Grimmer M, Taggart ML, Boehm AB. 2014. Predicting
772 water quality at Santa Monica Beach: evaluation of five different models for public

notification of unsafe swimming conditions. Water Res 67:105–117. 10.1016/j.watres.2014.09.001

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

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

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

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

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

63. 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.

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

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

799 66. Lu J, Santo Domingo JW, Lamendella R, Edge T, Hill S. 2008. Phylogenetic diversity
800 and molecular detection of bacteria in gull feces. Appl Environ Microbiol 74:3969-3976.
801 10.1128/AEM.00019-08.

802 67. Harwood VJ, Brownell M, Wang S, Lepo J, Ellender RD, Ajidahun A, Hellein KN,
803 Kennedy E, Ye X, Flood C. 2009. Validation and field testing of library-independent
804 microbial source tracking methods in the Gulf of Mexico. Water Res 43:4812–4819.
805 10.1016/j.watres.2009.06.029.

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

809 69. Jokinen CC, Schreier H, Mauro W, Taboada E, Isaac-Renton JL, Topp E, Edge T,
810 Thomas JE, Gannon VPJ. 2010. The occurrence and sources of Campylobacter spp.,
811 Salmonella enterica and Escherichia coli O157:H7 in the Salmon River, British
812 Columbia, Canada. J Water Health 8:374–386. 10.2166/wh.2009.076.

813 70. Layton A, McKay L, Williams D, Garrett V, Gentry R, Sayler G. 2006. Development of
814 Bacteroides 16S rRNA gene taqman-based real-time PCR assays for estimation of total,

815 human, and bovine fecal pollution in water. Appl Environ Microbiol 72:4214–4224.
816 10.1128/AEM.01036-05.

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

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

823 73. Haugland RA, Varma M, Sivaganesan M, Kelty C, Peed L, Shanks OC. 2010. Evaluation
824 of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection
825 of selected Bacteroidales species and human fecal waste by qPCR. Syst Appl Microbiol
826 33:348-357. 10.1016/j.syapm.2010.06.001.

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

831 75. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ,
832 Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of
833 sequences per sample. Proc Natl Acad Sci U S A 108:4516–4522.
834 10.1073/pnas.1000080107.

835 76. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM,
836 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.

77. 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.

78. 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

79. 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

80. 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

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

858 **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%

859

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

864

865

866

867

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.

879 **Figure Legends**

880 **Figure 1: Wells Maine Study area and sampling sites.** All water collection sites are marked
881 with a dark grey circle. Sites that correspond to fresh water are indicated with a (1), estuary (2),
882 estuary beach (3), and marine beach (4).

883

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

891

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

896

897 **Figure 4: 16S Taxa Profiles and the Top Three Most Abundant Bacterial Classes in All**
898 **Source and Sink Samples.** Stacked bar plots represent percentages of the class level
899 composition of the microbial communities. Source corresponds to environmental sources that

were finger-printed with the SourceTracker program, and then used to determine their presence within water (sink) samples. The table represents the top three classes for each group of samples and * corresponds to phylum level. For a complete list of all taxa assignments refer to Supplementary material 4.

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

Figure 1: Wells Maine Study area and sampling sites. All water collection sites are marked with a dark grey circle. Sites that correspond to fresh water are indicated with a (1), estuary (2), estuary beach (3), and marine beach (4).

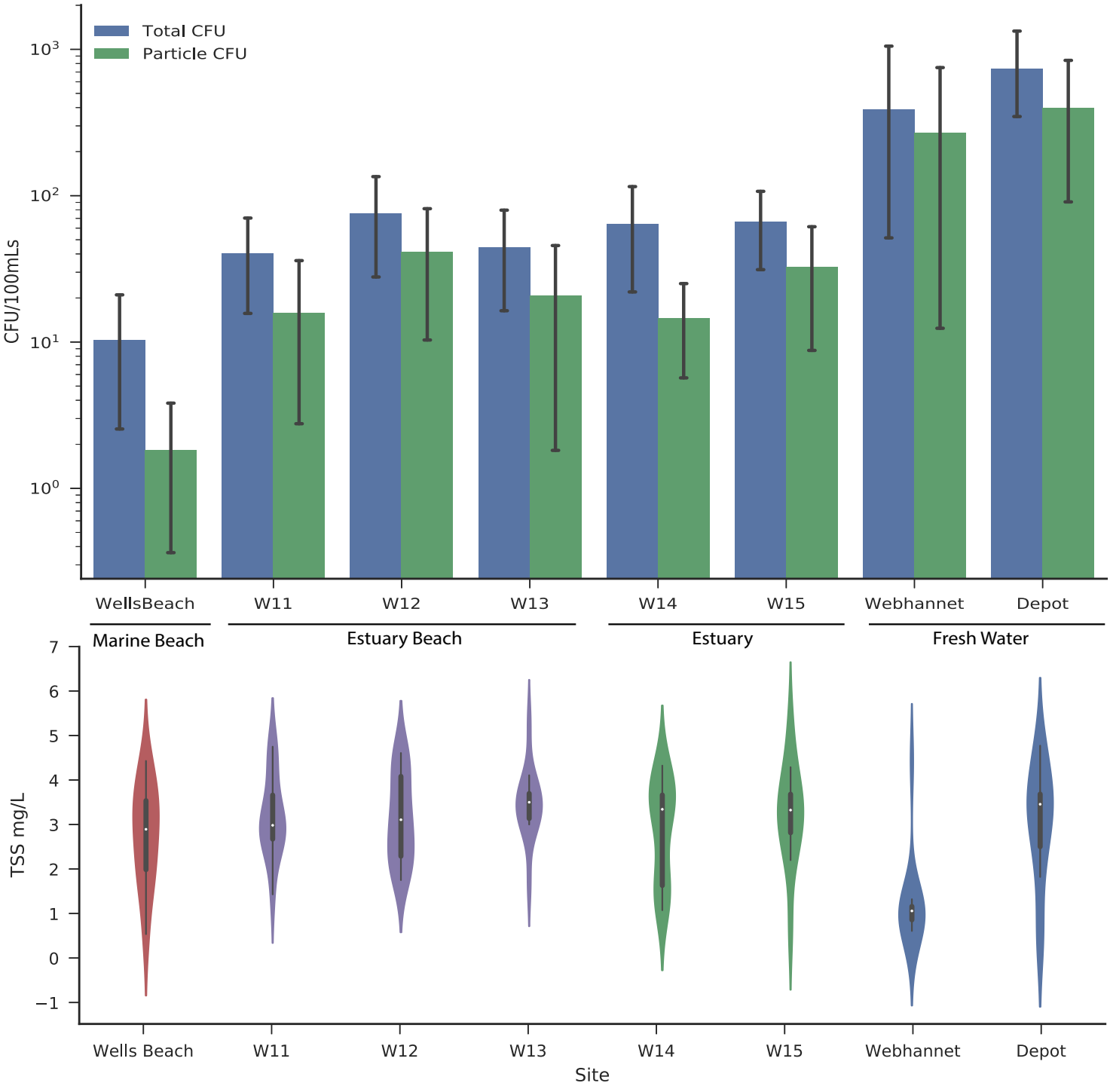

Figure 2: Geometric Mean Concentrations of Total and Particle Associated Enterococci and Average Total Suspended Solids Concentrations at the Eight Study Sites. (A) Total enterococci concentrations are represented with the blue bar, and particle associated enterococci concentrations correspond to the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

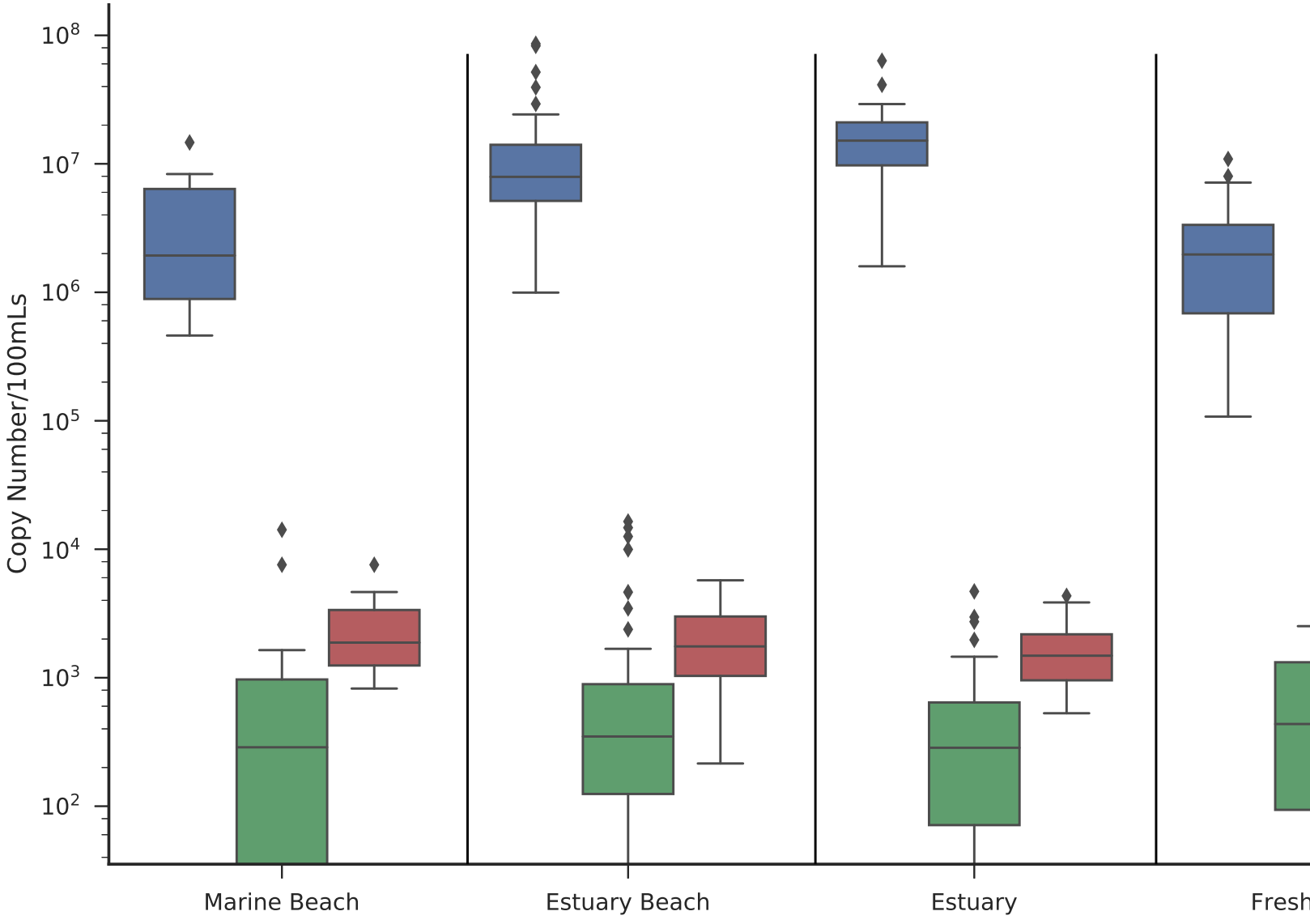

Figure 3: Relative Levels of Mammal, Human, and Bird Fecal Source at the Different Types of Study Sites. Box p
microbial source tracking markers at marine beach (Wells Beach), estuary beach (W11, W12, W13), estuary (W1
(Webhannet & Depot). Outlier data are represented with a black diamond.

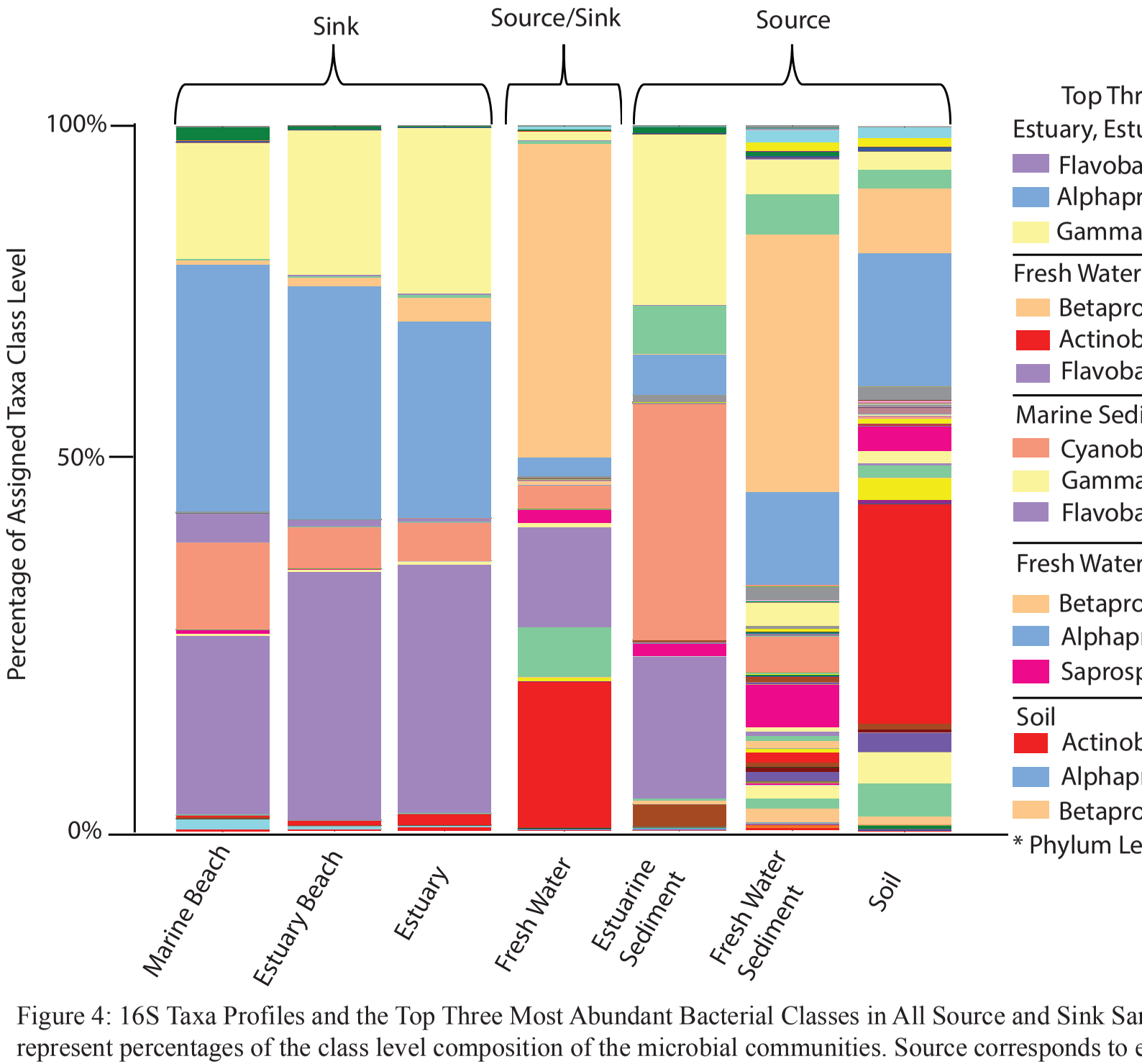

Figure 4: 16S Taxa Profiles and the Top Three Most Abundant Bacterial Classes in All Source and Sink Sam represent percentages of the class level composition of the microbial communities. Source corresponds to en were finger-printed with the SourceTracker program, and then used to determine their presence within water represents the top three classes for each group of samples and * corresponds to phylum level. For a complete refer to Supplementary material 4.

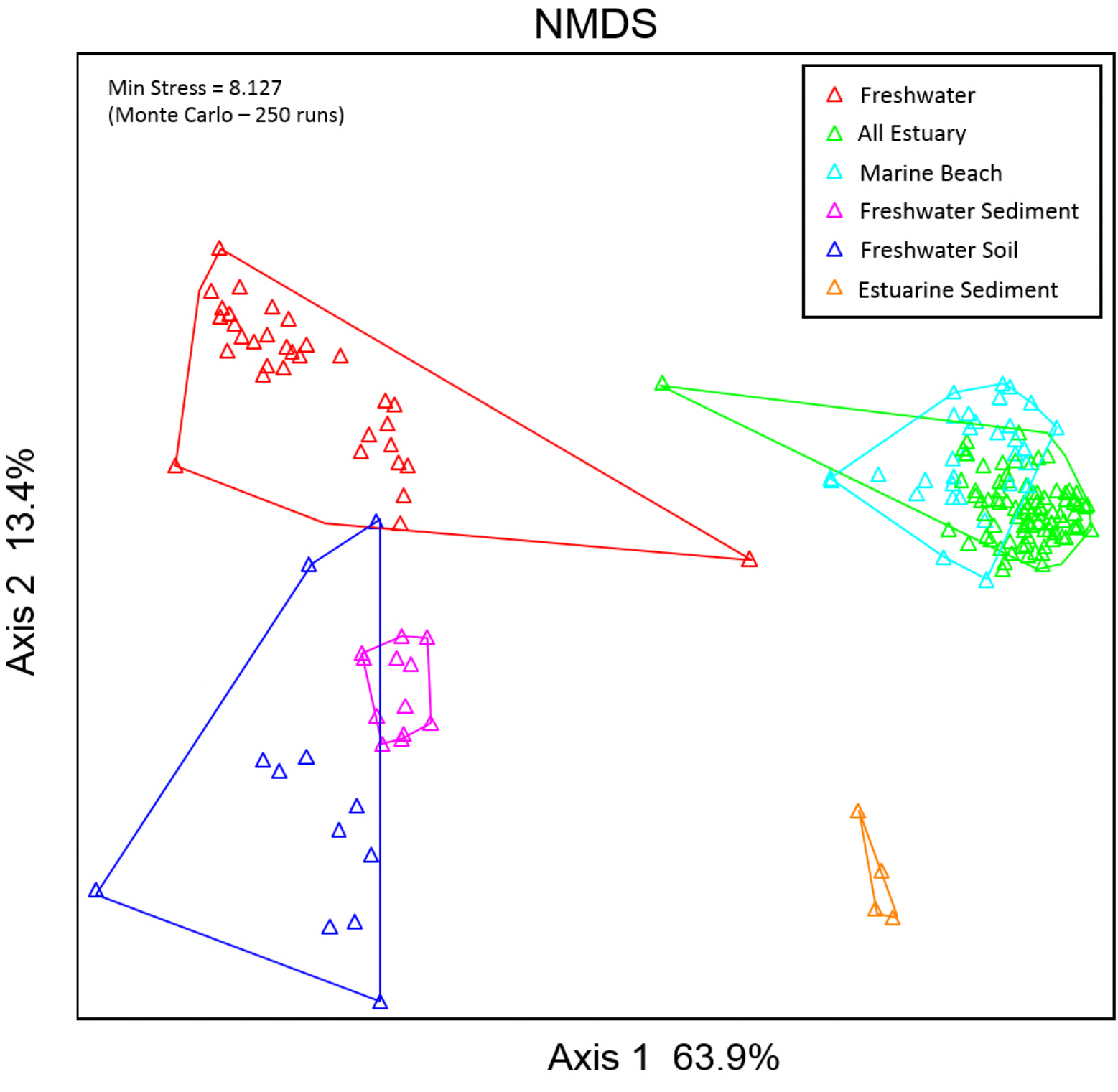

Figure 5. Differences Between Microbial Communities from Different Source Materials. Samples are color-coded based on sample matrix (i.e. soil, fresh water, etc.). Percent of variation explained are displayed on the x and y axis and the minimum stress of the ordination is shown in the top left corner.