# Comparison of qPCR and amplicon sequencing based methods for fecal source tracking in a mixed land use estuarine watershed

Sarah Phelan[1], Disha Soni[1], William R. Morales-Medina[2], N.L. Fahrenfeld[1*]

[1]Civil & Environmental Engineering, Rutgers, The State University of New Jersey, 500 Bartholomew Dr, Piscataway, NJ 08854
[2]Microbiology & Molecular Genetics, Rutgers, The State University of New Jersey

*nfahrenf@rutgers.edu, 848-445-8416

## Supplemental Information

### Recent History of Navesink Fecal Contamination

A 2006 study by the NJDEP classified a portion of the Navesink as impaired and classified 152 acres of the upper estuary as "prohibited" for shellfish harvesting.[1] TMDLs were established for the Navesink River to regulate permitted stormwater discharges for pollutant sources to the river. For the shellfish impaired portion of the Navesink (Navesink Estuary B), the annual TMDL was set at $1.26 \times 10^{15}$ cfu/yr. To achieve this limit, a 93% reduction in stormwater outputs from agricultural, marina, and urban sources was required.[2] In 2008, the NJDEP bureau of Marine Water Monitoring (BMWM) produced a report on storm studies, identifying that rain events correlated with high coliform results, indicating stormwater runoff was a source of coliform pollution, highest in the upstream portion of the river.[2] The investigation by the DEP in 2008 included microbial source tracking using Multiple Antibiotic Resistance Sampling (MAR), optical brightening agents, and F+ RNA coliphage. The study identified multiple sources of elevated coliform including both human and wildlife at various sites along the river.[2] The Monmouth county health department coordinated with municipalities along the Navesink to identify and address many sources. In 2015 after additional data review, an additional 565.7 acres of river were downgraded from "Special Restricted" to "Prohibited" for shellfish harvesting. The Navesink River, in addition to the downstream Shrewsbury River, support almost all soft clam fishery in New Jersey this downgrade has potential economic impacts within the state of New Jersey. In addition to these restrictions, the Navesink is listed in a 2012 Water body report as impaired for Fish Consumption, Primary Contact Recreation, and Aquatic Life.

**Table S1**: Fecal spike sample location and description for samples used in the creation of the fecal library.

| Sample type | Sample Location | Description |
|---|---|---|
| Horse | Horse Farm in Middlesex County | Fresh Manure from 2 horses from two stables |
| Horse | Rutgers Horse Farm | Fresh Manure Samples from multiple horses in farm |
| Goose | Park in Red Bank, New Jersey | Multiple fresh samples collected adjacent to park pond |
| Goose | Business Park in Somerset, New Jersey | Fresh samples collected from park |
| Dog | 4 Domestic dogs from 3 owners – mixed breeds | Fresh samples provided by owners |

**Table S2:** Fecal library composition by volume (surface water and wastewater) and weight (fecal material spikes).

| Sample name | Contents | Surface Water | Wastewater | Goose | Horse | Dog |
|---|---|---|---|---|---|---|
| **Sa, Sb** | Surface Water | 900 mL | 0 mL | 0 g | 0 g | 0 g |
| **SWa, SWb** | Surface Water, Wastewater | 900 mL | 100 mL | 0 g | 0 g | 0 g |
| **SWGa, SWGb** | Surface Water, Wastewater, Goose | 900 mL | 100 mL | 1 g | 0 g | 0 g |
| **SWHa, SWHb** | Surface Water, Wastewater, Horse | 900 mL | 100 mL | 0 g | 1 g | 0 g |
| **SWDa, SWDb** | Surface Water, Wastewater, Dog | 900 mL | 100 mL | 0 g | 0 g | 1 g |
| **SWGHa, SWGHb** | Surface Water, Wastewater, Goose, Horse | 900 mL | 100 mL | 1 g | 1 g | 0 g |
| **SWGDa, SWGDb** | Surface Water, Wastewater, Goose, Dog | 900 mL | 100 mL | 1 g | 0 g | 1 g |
| **SWDHa, SWDHb** | Surface Water, Wastewater, Dog, Horse | 900 mL | 100 mL | 0 g | 1 g | 1 g |
| **SWGHDa, SWGHDb** | Surface Water, Wastewater, Goose, Horse, Dog | 900 mL | 100 mL | 1 g | 1 g | 1 g |
| **SDGHa, SDGHb** | Surface Water, Dog, Goose, Horse | 900 mL | 0 mL | 1 g | 1 g | 1 g |
| **B** | De-Ionized Water | 0 mL | 0 mL | 0 g | 0 g | 0 g |

**Table S3:** Primer sequences, annealing temperatures, amplicon lengths, and references.

| Target | Primer or Probe | Sequence 5' to 3' | Ta (°C) | Amplicon length (bp) | Reference(s) |
|---|---|---|---|---|---|
| Horse | Bac708R | CAATCGGAGTTCTTCGTG | 53 | 111 | Dick et al.[3] |
| | HoF597F | CCAGCCGTAAAATAGTCGG | | | |
| Human | HF183F | ATCATGAGTTCACATGTCCG | 53 | 59 | Bernhard & Field[4]; Seurinck et al.[5] |
| | Bac242R | TACCCCGCCTACTATCTAATG | | | |
| Human | BacHum160f | TGAGTTCACATGTCCGCATGA | 60 | 81 | Kildare et al.[6] |
| | BacHum241r | CGTTACCCCGCCTACTATCTA | | | |
| | Probe | /56-FAM/TCC GGT AGA CGA TGG GGA TGC GTT /36-TAMSp/ | | | |
| Bacteria | 16S rRNA forward | CCTACGGGAGGCAGCAG | 65 | 202 | Muyzer al.[7] |
| | 16S rRNA reverse | ATTACCGCGGCTGCTGG | | | |

For both the horse primer set and HF183, qPCR was performed by denaturing at 95°C for 10 min then 40 cycles of melting at 95°C for 15 sec and annealing at 53°C for 30 sec. For BacHum, qPCR was performed by heating samples to 50°C for 2 min, denaturing at 95°C for 10 min followed by 40 cycles of melting at 95°C for 15 sec and annealing at 60°C for 30 sec.

**Table S4:** Library and field sample sequencing information. Subsampling was performed resulting in 32,404 sequences per sample before calculation of evenness, Margalef's richness, Shannon diversity index, Simpson diversity index at the species level.

| Navesink library samples | | | | | | Navesink River samples | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample Name | Number of Sequences | Evenness | Richness | Shannon Index $\log_e$ | Simpson | Sample Name | Number of Sequences | Evenness | Richness | Shannon Index $\log_e$ | Simpson |
| S-a | 60512 | 0.53 | 47 | 3.3 | 0.93 | S-10 wet a | 38997 | 0.43 | 64 | 2.8 | 0.86 |
| S-b | 48839 | 0.54 | 51 | 3.4 | 0.93 | S-10 wet b | 46356 | 0.42 | 64 | 2.8 | 0.86 |
| SW-a | 38070 | 0.60 | 73 | 4.0 | 0.94 | S-10 dry a | 62272 | 0.49 | 68 | 3.2 | 0.87 |
| SW-b | 46261 | 0.58 | 74 | 3.9 | 0.94 | S-10 dry b | 54542 | 0.50 | 71 | 3.3 | 0.88 |
| SWG-a | 51676 | 0.56 | 62 | 3.7 | 0.93 | S-14 wet a | 60325 | 0.52 | 53 | 3.3 | 0.88 |
| SWG-b | 50644 | 0.57 | 68 | 3.7 | 0.94 | S-14 wet b | 61086 | 0.54 | 57 | 3.5 | 0.91 |
| SWH-a | 63107 | 0.55 | 70 | 3.6 | 0.91 | S-14 dry a | 49723 | 0.54 | 48 | 3.4 | 0.92 |
| SWH-b | 53462 | 0.60 | 72 | 4.0 | 0.95 | S-14 dry b | 52282 | 0.51 | 45 | 3.1 | 0.88 |
| SWD-a | 60816 | 0.59 | 62 | 3.8 | 0.94 | S-34 wet a | 53117 | 0.52 | 49 | 3.3 | 0.92 |
| SWD-b | 53998 | 0.57 | 67 | 3.7 | 0.92 | S-34 wet b | 53775 | 0.53 | 46 | 3.2 | 0.93 |
| SWGH-a | 49075 | 0.62 | 68 | 4.1 | 0.96 | S-34 dry a | 50952 | 0.55 | 63 | 3.6 | 0.94 |
| SWGH-b | 68312 | 0.57 | 58 | 3.7 | 0.93 | S-34 dry b | 47716 | 0.54 | 59 | 3.5 | 0.93 |
| SWGD-a | 54991 | 0.64 | 62 | 4.1 | 0.97 | S-52 wet a | 46295 | 0.54 | 60 | 3.5 | 0.94 |
| SWGD-b | 51986 | 0.61 | 64 | 3.9 | 0.95 | S-52 wet b | 51613 | 0.56 | 62 | 3.6 | 0.95 |
| SWDH-a | 64088 | 0.56 | 56 | 3.5 | 0.93 | S-52 dry a | 54426 | 0.54 | 59 | 3.5 | 0.91 |
| SWDH-b | 48282 | 0.51 | 53 | 3.2 | 0.90 | S-52 dry b | 40910 | 0.54 | 56 | 3.5 | 0.90 |
| SWGDH-a | 69636 | 0.53 | 52 | 3.3 | 0.91 | S-56 wet a | 39462 | 0.56 | 48 | 3.4 | 0.94 |
| SWGDH-b | 54634 | 0.57 | 58 | 3.7 | 0.94 | S-56 wet b | 38070 | 0.57 | 45 | 3.5 | 0.94 |
| SDGH-a | 57054 | 0.56 | 48 | 3.5 | 0.94 | S-56 dry a | 54252 | 0.53 | 67 | 3.5 | 0.91 |
| SDGH-b | 62959 | 0.55 | 55 | 3.5 | 0.92 | S-56 dry b | 58018 | 0.56 | 67 | 3.6 | 0.93 |
| Trip Blank 1* | 54511 | 0.49 | 33 | 2.8 | 0.90 | S-58 dry a | 48014 | 0.52 | 47 | 3.2 | 0.91 |
| | | | | | | S-58 dry b | 32404 | 0.50 | 45 | 3.1 | 0.89 |

*Trip Blank 1 was composed of autoclaved DI water and was used during library creation. Trip Blank 1 had on average 4.2 ± 0.7 log units less 16S rRNA gene copies and the other library samples.

**Table S5:** Wet and dry weather coliform results and relative percent differences (RPD) for replicate samples

| **Navesink River wet weather** | | |
|---|---|---|
| Sample ID | CFU/100 mL | RPD (%) |
| S-10 Wet | 3100 | 71 |
| S-14 Wet | 260 | 15 |
| S-34 Wet | 190 | 11 |
| S-52 Wet | 200 | 40 |
| S-56 Wet | 950 | 53 |
| **Navesink River dry weather** | | |
| Sample ID | CFU/100 mL | RPD (%) |
| S-10 Dry | 2800 | N/A* |
| S-14 Dry | TNTC** | N/A |
| S-34 Dry | 203 | N/A* |
| S-52 Dry | 180 | 105 |
| S-56 Dry | TNTC** | N/A |
| S-58 Dry | 87 | 15 |
| TB-2 | 0 | N/A |
| **Navesink Library** | | |
| S*** | 600 | N/A |

** RPD could not be calculated because duplicate sample was TNTC. Reported results are single sample*

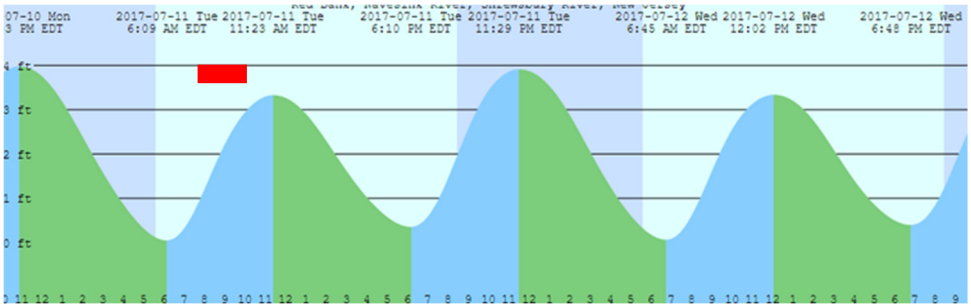
*** TNTC: Too numerous to count*

****S: Surface water sample, mixture of sample from S-34 and S-58*

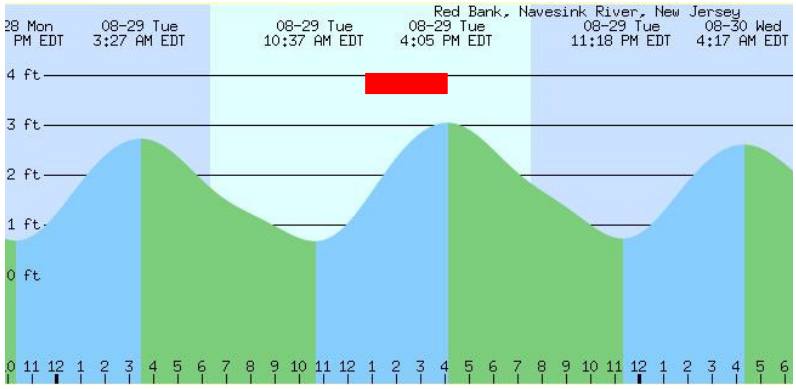
**Table S6**. Average percent of all SourceTracker assignments to the Australian fecal library. The relative percent difference for replicates across all assignments was 49±41% for the library and 28±25% for the river samples.

| Sample | Human/sewage | | | | Horse | Avian | | | | Dog | Bat | Cat | Agricultural mammal | | | | Wild mammal | | | Unknown |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | human | raw sewage | septic | second-ary sewage | | bird | duck | chicken | common-miner | | | | cow | goat | pig | sheep | possum | rabbit | rat | |
| S | 0.4 | 1.2 | 0.7 | 7.3 | 0.7 | 0.7 | 1.0 | 1.1 | 0.7 | 0.4 | 3.1 | 0.3 | 0.3 | 0.2 | 0.3 | 0.3 | 0.2 | 0.6 | 0.4 | 80 |
| SW | 4.1 | 26 | 1.8 | 22 | 1.8 | 1.4 | 7.0 | 2.2 | 2.5 | 1.0 | 20 | 1.0 | 1.0 | 1.1 | 1.2 | 1.4 | 0.4 | 1.5 | 1.6 | 0.7 |
| SWG | 0.6 | 5.6 | 1.0 | 10 | 2.3 | 27 | 5.6 | 27 | 3.0 | 0.9 | 12 | 0.4 | 0.6 | 0.6 | 0.6 | 0.8 | 0.3 | 0.9 | 0.9 | 0.2 |
| SWH | 4.3 | 18 | 1.8 | 16 | 3.6 | 1.8 | 7.7 | 2.6 | 2.1 | 1.2 | 27 | 1.5 | 1.5 | 1.8 | 1.7 | 2.2 | 0.9 | 2.0 | 2.2 | 0.6 |
| SWD | 2.8 | 5.1 | 1.9 | 13 | 1.9 | 1.8 | 7.0 | 2.5 | 2.1 | 2.8 | 40 | 4.3 | 1.9 | 1.8 | 1.6 | 2.1 | 2.0 | 2.6 | 2.3 | 0.3 |
| SWGH | 1.5 | 3.0 | 0.8 | 25 | 3.8 | 11 | 4.3 | 23 | 2.1 | 1.3 | 16 | 0.9 | 1.1 | 0.8 | 0.9 | 1.2 | 0.6 | 1.1 | 1.3 | 0.2 |
| SWGD | 1.8 | 5.1 | 1.1 | 16 | 3.1 | 11 | 7.5 | 21 | 3.6 | 2.2 | 17 | 1.5 | 1.3 | 1.1 | 1.2 | 1.5 | 0.9 | 1.5 | 1.7 | 0.1 |
| SWDH | 1.4 | 3.1 | 0.8 | 14 | 3.3 | 26 | 5.3 | 20 | 2.3 | 1.9 | 11 | 1.4 | 1.3 | 1.4 | 1.1 | 1.9 | 0.8 | 1.0 | 1.2 | 1.2 |
| SWGHD | 1.0 | 4.1 | 0.8 | 27 | 2.9 | 19 | 3.6 | 26 | 1.8 | 1.3 | 4.1 | 0.6 | 0.7 | 0.8 | 0.7 | 1.2 | 0.5 | 1.1 | 1.0 | 2.0 |
| SDGH | 1.2 | 2.2 | 0.7 | 22 | 3.1 | 9.5 | 3.7 | 34 | 2.0 | 1.2 | 12 | 1.1 | 0.6 | 0.8 | 0.7 | 1.0 | 0.5 | 0.7 | 1.0 | 2.0 |
| Blank | 1.5 | 3.1 | 1.4 | 8.6 | 4.4 | 4.6 | 4.1 | 6.2 | 1.7 | 1.8 | 4.0 | 2.4 | 2.4 | 1.9 | 1.9 | 1.6 | 5.1 | 3.6 | 1.8 | 38 |
| S-10 Wet | 0.4 | 1.2 | 0.8 | 71 | 0.7 | 0.3 | 0.4 | 0.4 | 0.3 | 0.2 | 0.5 | 0.3 | 0.3 | 0.2 | 0.3 | 0.2 | 0.2 | 0.3 | 0.3 | 21 |
| S-10 Dry | 0.7 | 1.5 | 0.8 | 28 | 0.8 | 0.7 | 0.7 | 0.9 | 0.5 | 0.3 | 0.9 | 0.4 | 0.4 | 0.2 | 0.3 | 0.3 | 0.3 | 0.4 | 0.3 | 61 |
| S-14 Wet | 0.3 | 1.0 | 0.7 | 30 | 0.7 | 0.8 | 0.7 | 0.7 | 0.4 | 0.2 | 1.5 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.1 | 0.4 | 0.2 | 61 |
| S14-Dry | 0.2 | 4.0 | 0.8 | 28 | 0.8 | 0.9 | 5.2 | 1.3 | 1.1 | 0.3 | 33 | 0.1 | 0.2 | 0.1 | 0.1 | 0.2 | 0.1 | 0.4 | 0.2 | 22 |
| S-34 Wet | 0.3 | 1.1 | 0.7 | 6.9 | 0.8 | 0.8 | 1.4 | 0.9 | 0.7 | 0.3 | 3.2 | 0.3 | 0.3 | 0.2 | 0.2 | 0.3 | 0.2 | 0.5 | 0.4 | 81 |
| S-34 Dry | 0.6 | 1.6 | 0.8 | 11 | 1.3 | 0.7 | 2.2 | 1.2 | 0.9 | 0.4 | 8.6 | 0.3 | 0.4 | 0.4 | 0.4 | 0.5 | 0.2 | 0.5 | 0.5 | 68 |
| S-52 Wet | 0.5 | 1.2 | 0.6 | 11 | 0.9 | 0.7 | 1.3 | 1.0 | 0.9 | 0.4 | 4.1 | 0.3 | 0.4 | 0.2 | 0.3 | 0.4 | 0.2 | 0.5 | 0.5 | 74 |
| S-52 Dry | 0.3 | 0.9 | 0.6 | 26 | 0.7 | 0.6 | 0.5 | 0.5 | 0.4 | 0.2 | 1.2 | 0.3 | 0.2 | 0.1 | 0.2 | 0.2 | 0.1 | 0.3 | 0.2 | 66 |
| S-56 Wet | 0.4 | 1.1 | 0.7 | 8.1 | 0.8 | 0.7 | 1.6 | 1.1 | 0.8 | 0.3 | 10 | 0.2 | 0.2 | 0.1 | 0.2 | 0.2 | 0.2 | 0.7 | 0.3 | 72 |
| S-56 Dry | 0.3 | 5.3 | 1.0 | 47 | 1.1 | 1.0 | 6.0 | 1.6 | 1.1 | 0.4 | 32 | 0.2 | 0.2 | 0.2 | 0.3 | 0.3 | 0.1 | 0.7 | 0.3 | 1.1 |
| S-58 Dry | 0.3 | 2.1 | 0.7 | 5.7 | 0.8 | 0.7 | 3.6 | 1.3 | 1.1 | 0.4 | 34 | 0.1 | 0.2 | 0.1 | 0.2 | 0.2 | 0.1 | 0.4 | 0.3 | 47 |

a.




b.




c.

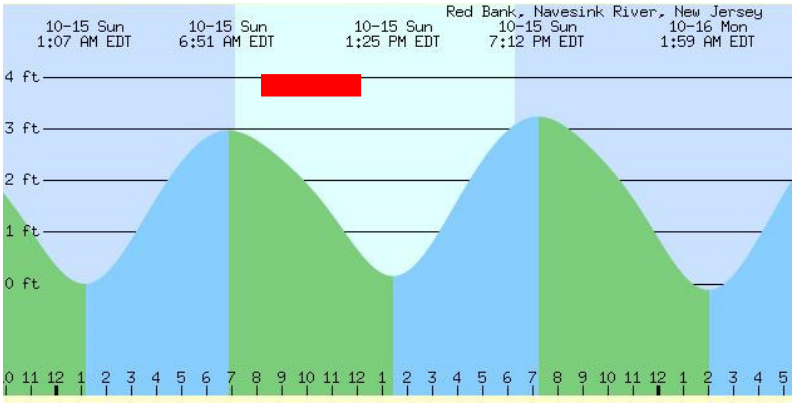



**Figure S1:** Tidal Cycles for (a) library (collected 8:30-9:50), (b) wet (collected 1-4pm), and (c) dry weather (collected 8am-12pm) sampling events. Tidal data was collected from tides.mobilegeographics.com.

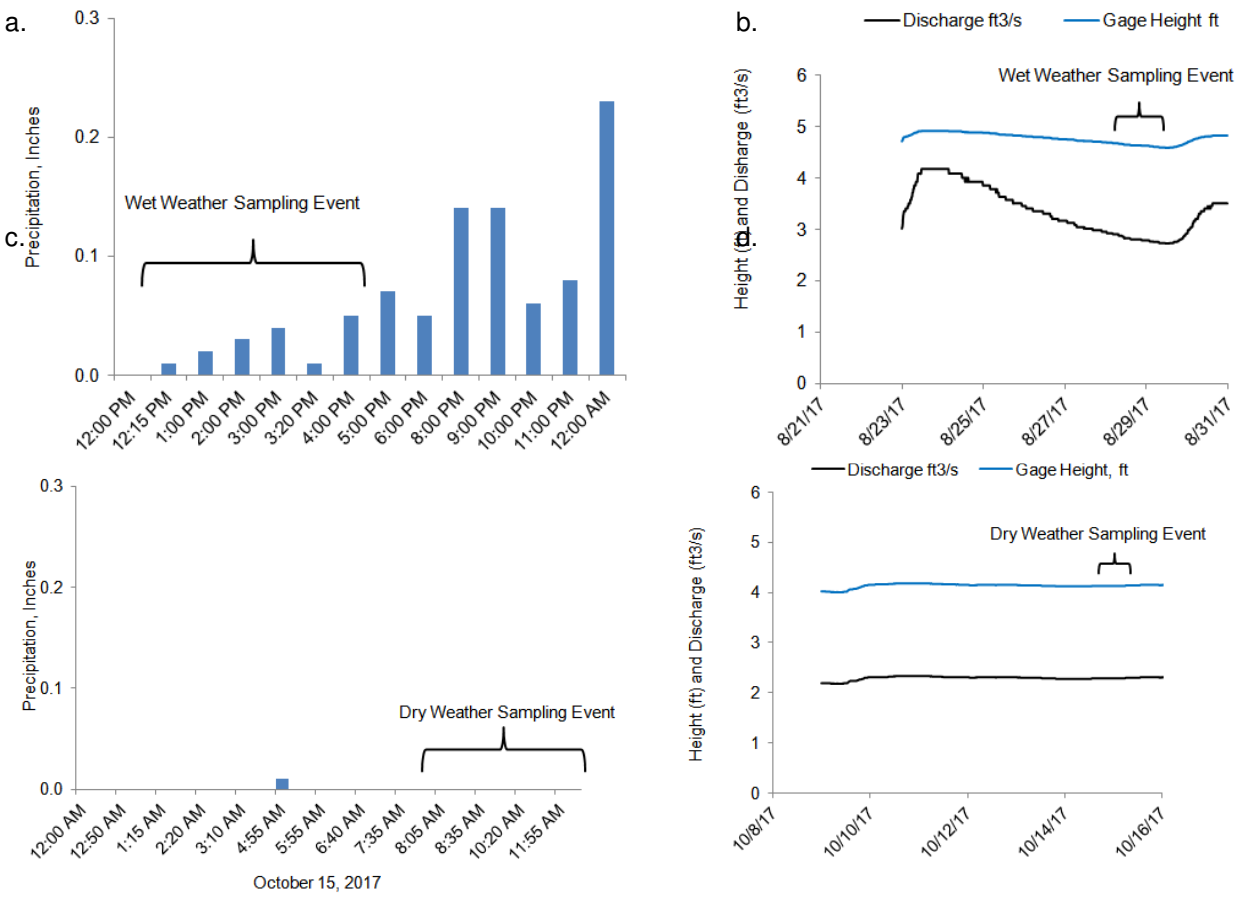



**Figure S2:** Precipitation (a., c.), discharge, and gage height (b., d.) data for Red Bank New Jersey during wet weather (a., b.) and dry weather (c., d) sampling events. Precipitation from Weather Underground, discharge and gage height from USGS *Station 01407500 Swimming River near Red Bank NJ*

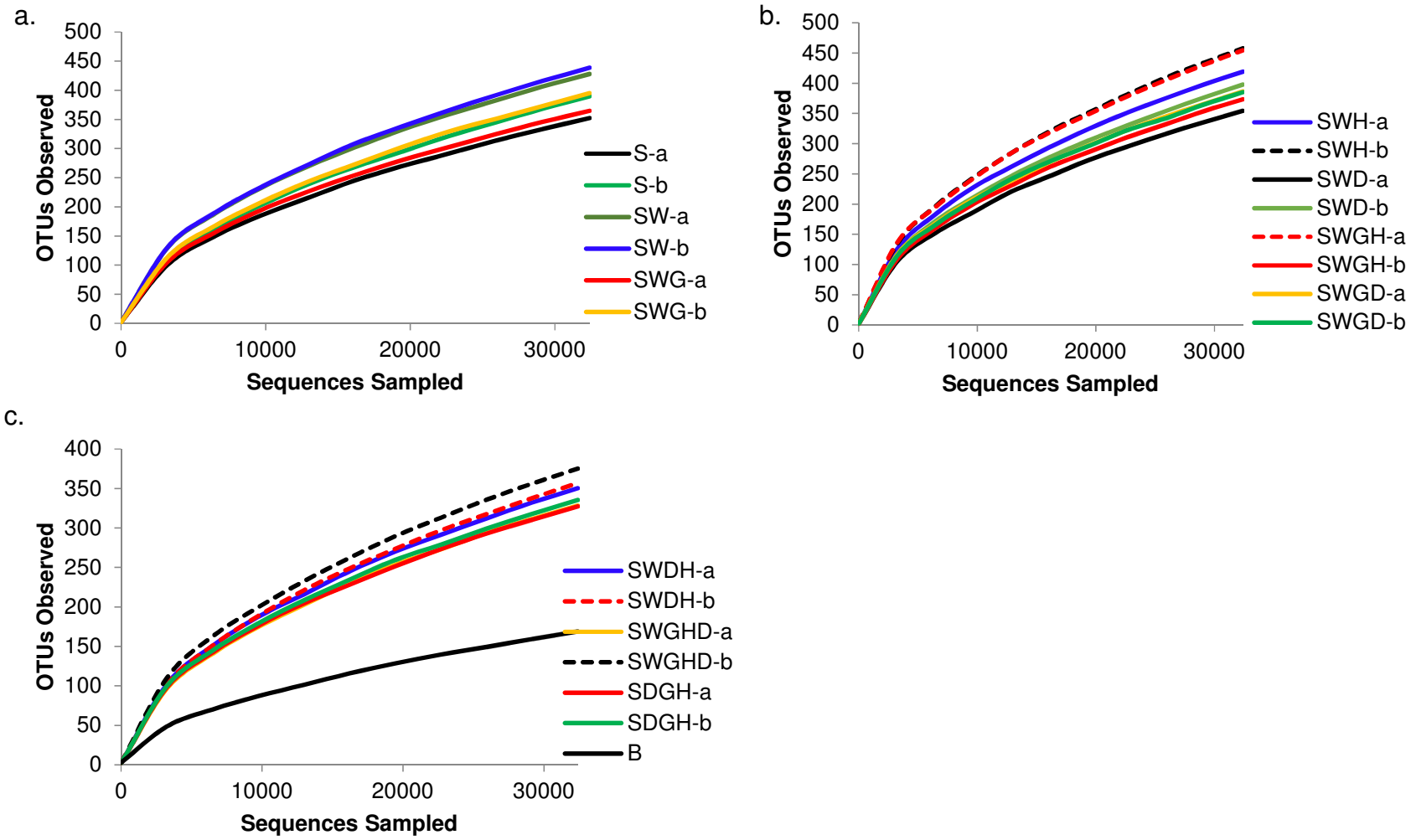



**Figure S3: a-c** Rarefaction curves for Navesink library samples at the class level

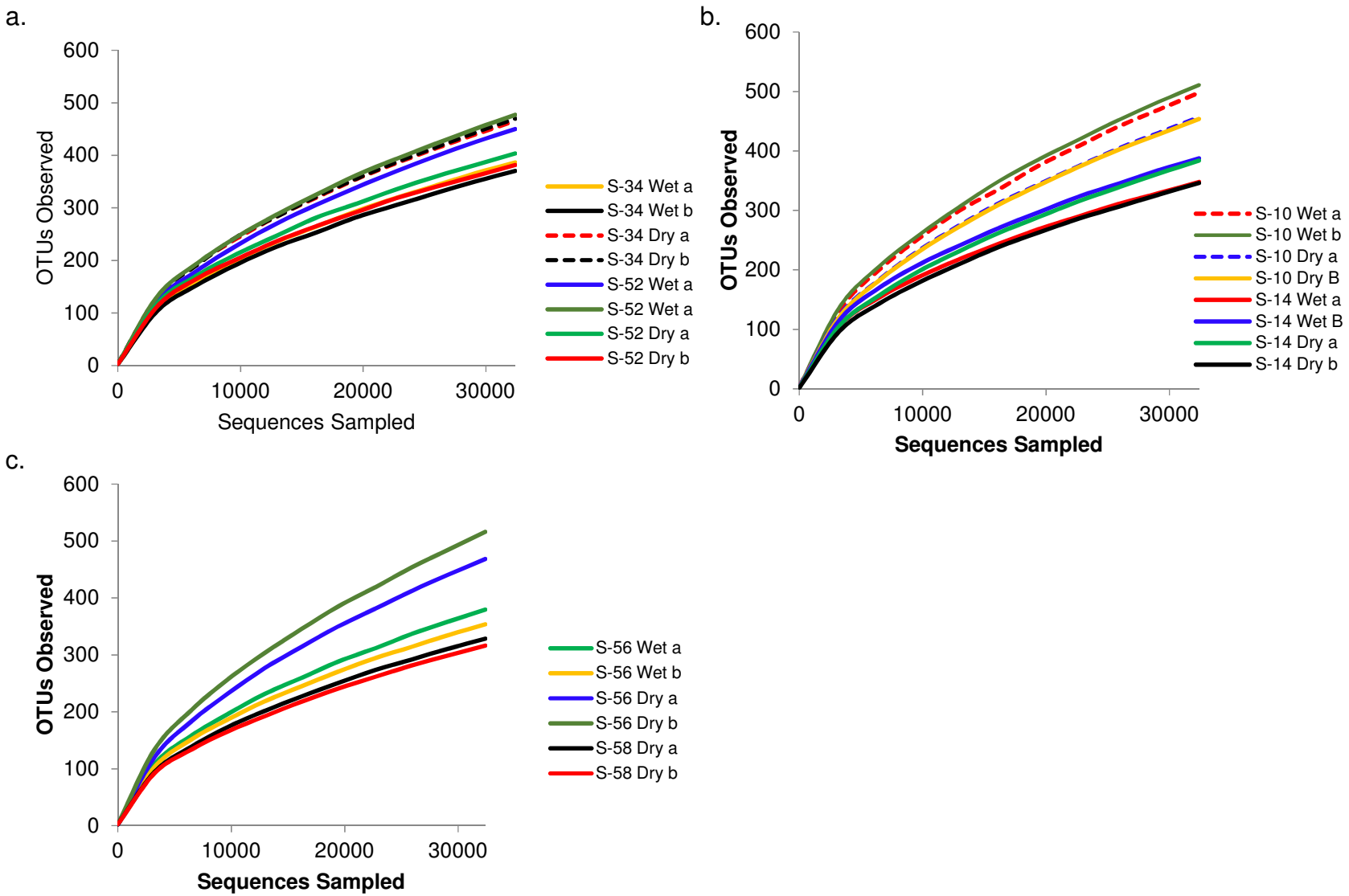



**Figure S4:** a-c Rarefaction curves for Navesink River samples

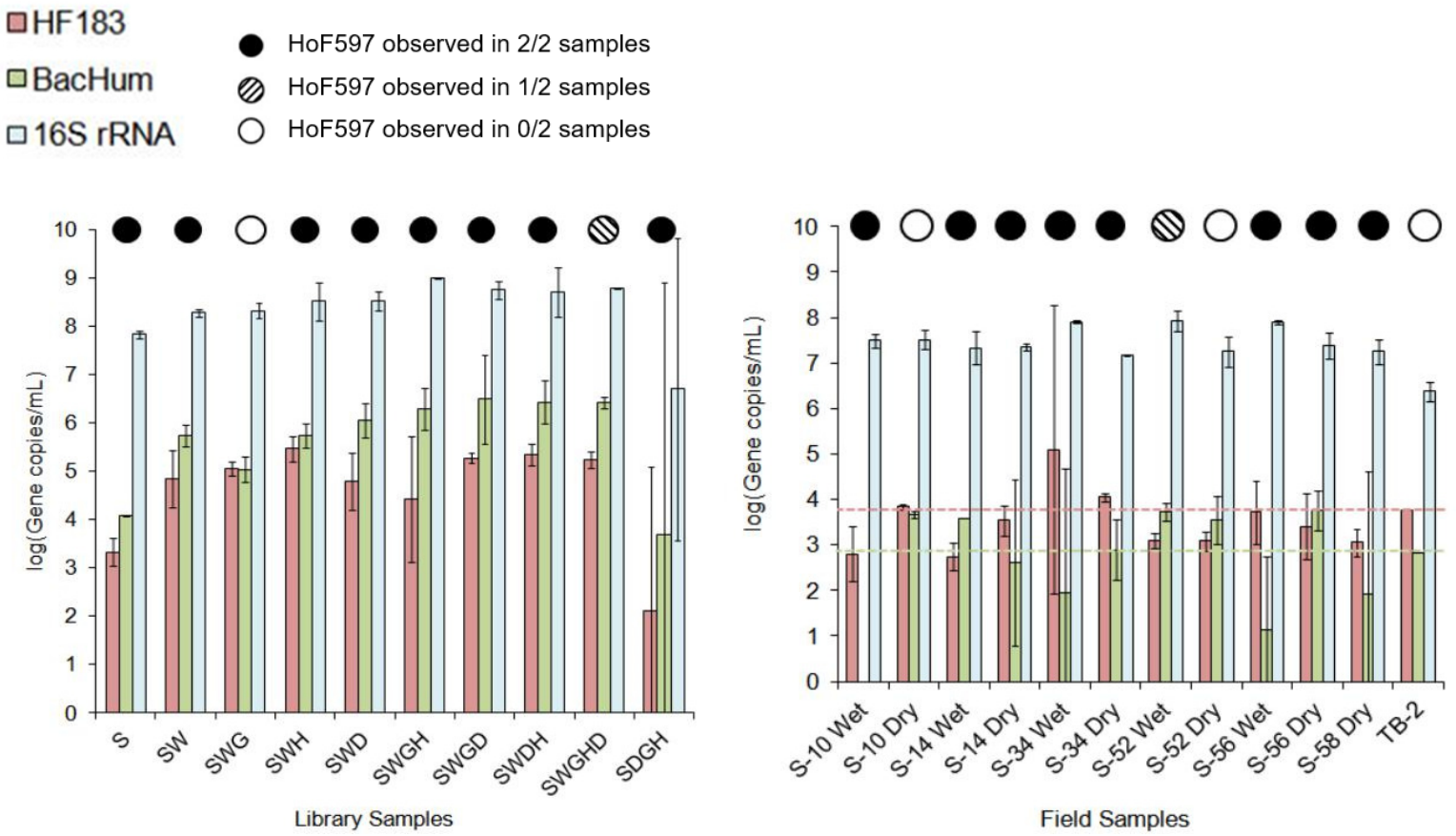



**Figure S5** qPCR results for a. library and b. field samples using human fecal marker genes HF183 and BacHum and 16S rRNA gene for total bacterial population. Error bars represent high and low values of replicate (N=2) samples. Presence of the horse fecal marker gene HOF597 is indicated by solid black circles, observation in one of two replicates with half-filled circles, and absence by white circles. Dashed lines represent trip blank (TB-2) concentrations of HF183 and BacHum.

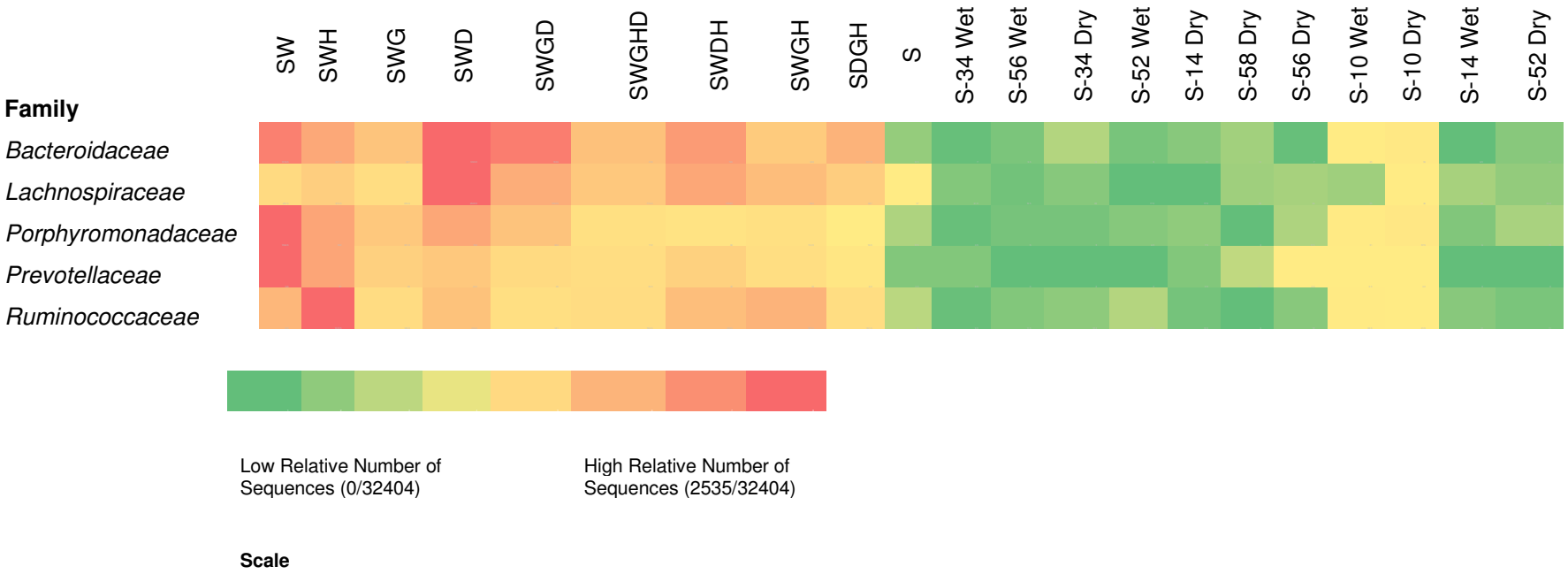



**Figure S6 a** Heatmap showing relative number of sequences for bacterial families previously reported to be associated with fecal material: *Bacteroidaceae, Lachnospiraceae, Ruminococcaceae, Porphyromonadaceae, and Prevotellaceae*

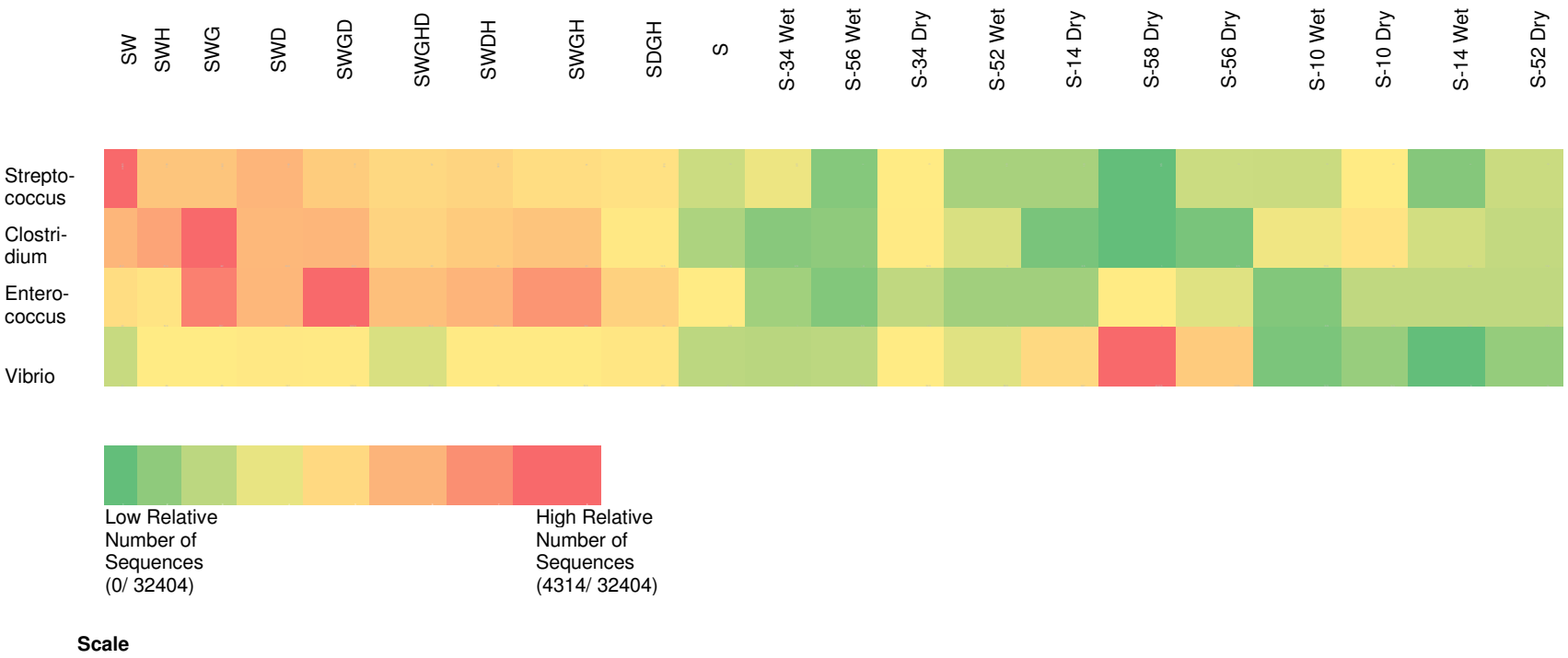
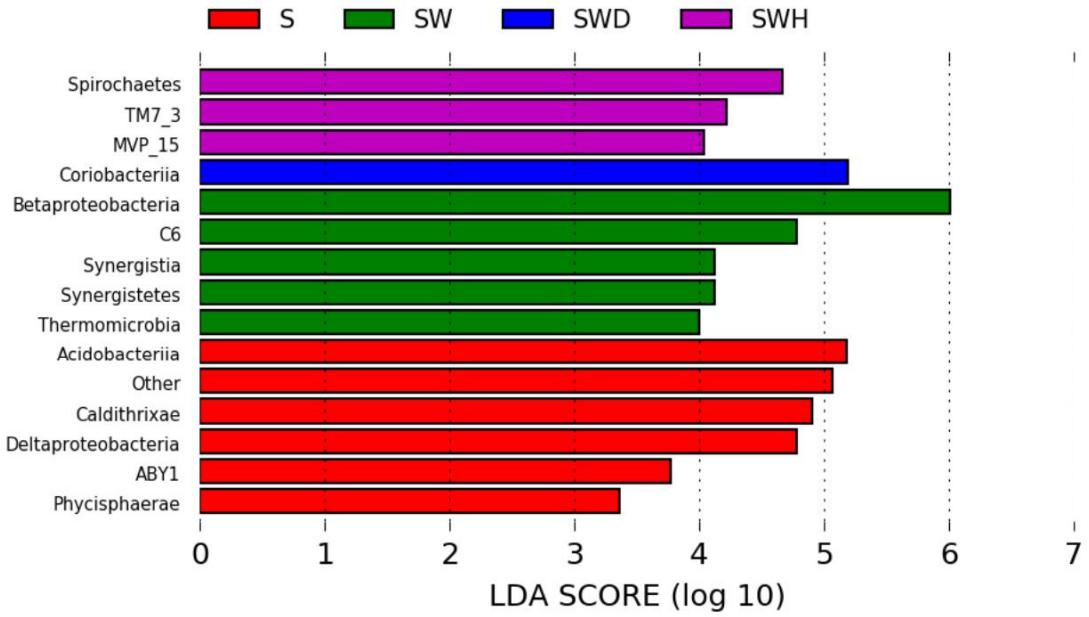



**Figure S6b:** Heatmap showing relative number of sequences for bacterial genera containing microbes used as fecal indicators: *Streptococcus, Clostridium, Enterococcus, Vibrio.*

a.




b.

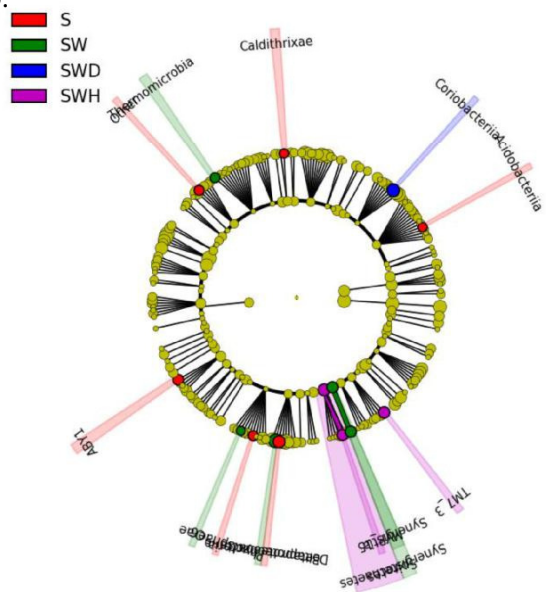



**Figure S7:** Linear discriminant analysis results from galaxy LEfSE analysis tool for class level data. a. Biomakers for library samples. b. Cladogram visually illustrates demonstrates relationship between biomarker species identified in figure a.

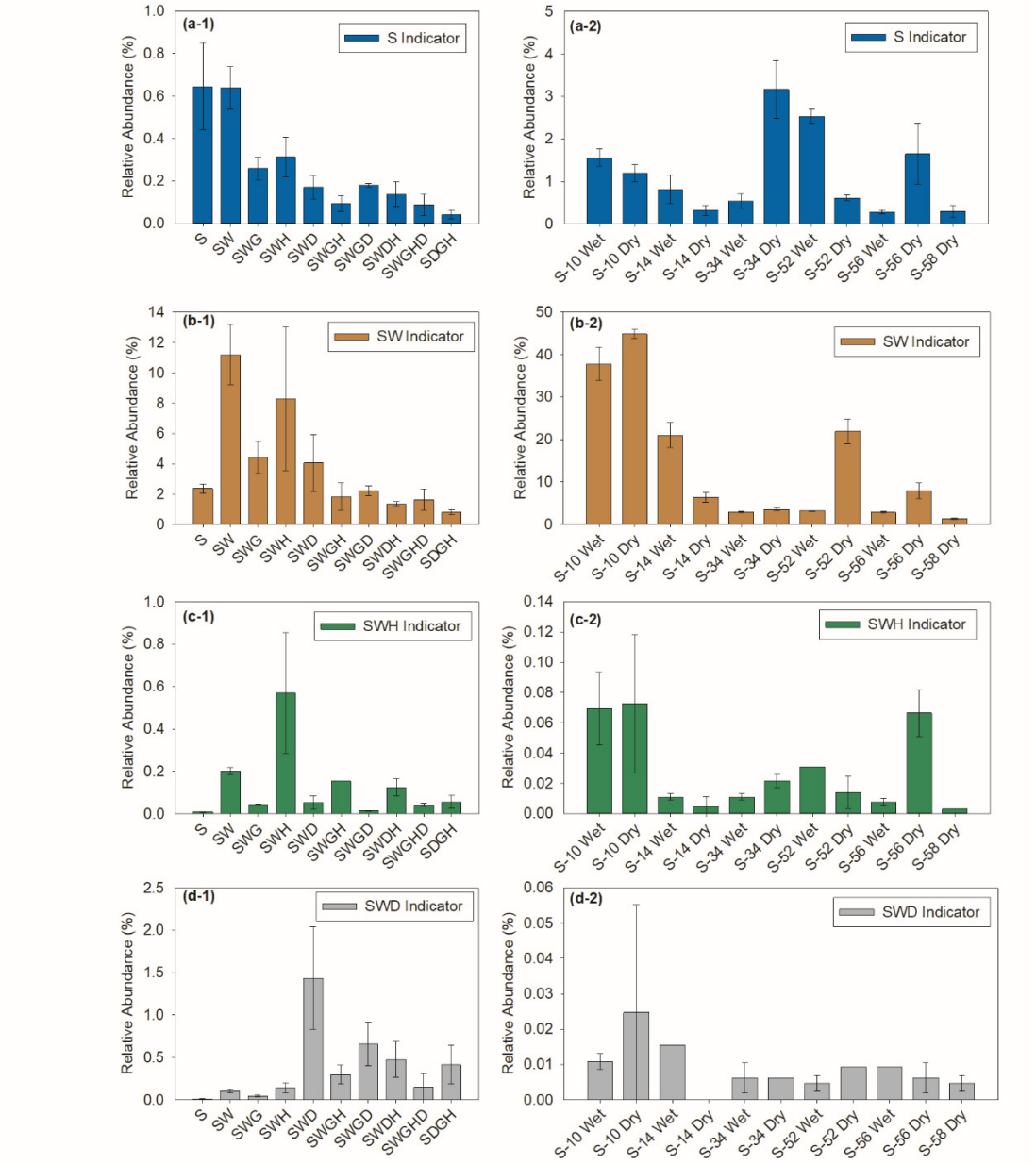



**Figure S8:** Relative abundance of indicator bacteria for surface water* in (a-1) library and (a-2) field samples, wastewater spiked surface water** in (b-1) library and (b-2) field samples, horse and wastewater spiked surface water*** in (c-1) library and (c-2) field samples, and dog and wastewater spiked surface water**** in (d-1) library and (d-2) field samples. Error bars represent high and low values of replicate samples (N = 2).

**Acidobacteriia*, *Caldithrixae*, ABY1, *Phycisphaerai*, *Deltaproteobacteria*; ***Thermomicrobia*, C6, *Betaproteobacteria*, *Synergistia*; *** MVP-15, *Spirochaetes*, and TM7-3; **** *Coriobacteriia*.

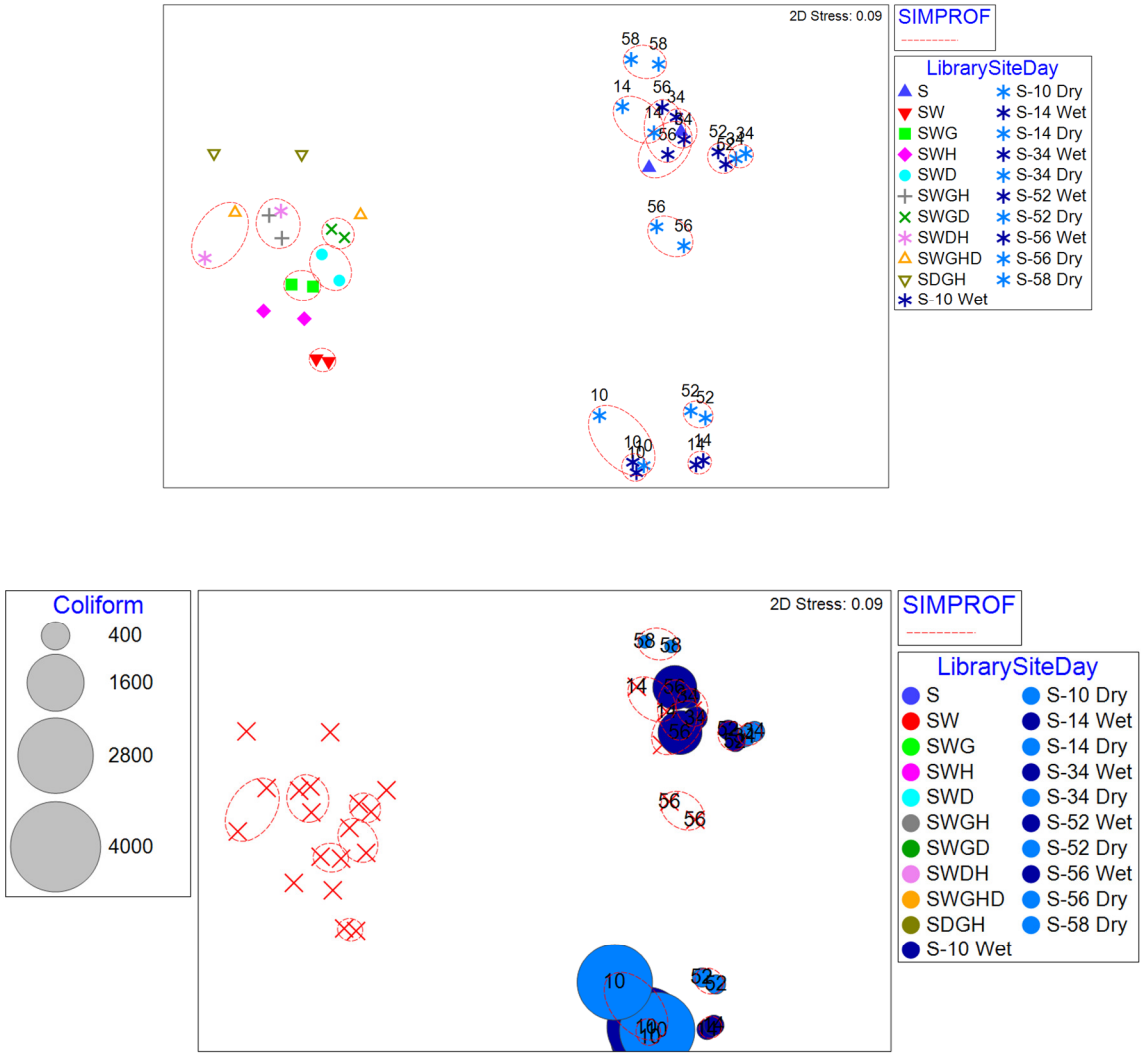



**Figure S9:** a. Nonmetric multidimensional scaling of microbial communities in field and library samples at the class level. Results of SIMPROF test showing no significant differences is overlaid. b. nDMS bubble plot for fecal coliform CFU/100mL (TNTC were assigned 4000 CFU/100 ML) to facilitate illustration.

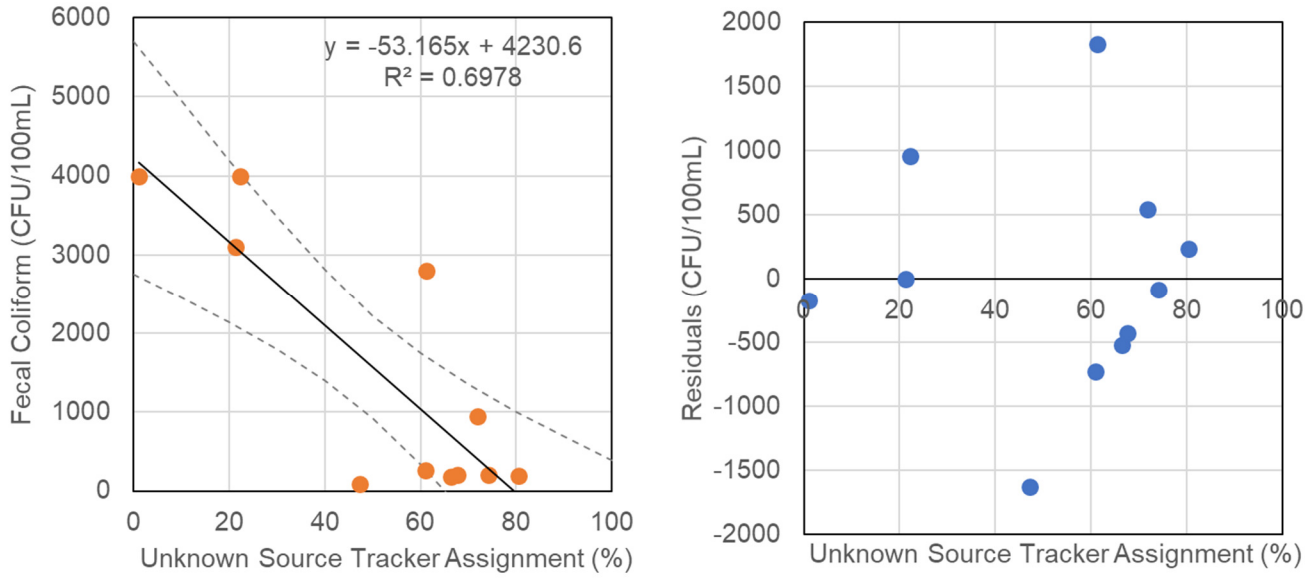



**Figure S10 a.** Fecal coliform versus unknown source tracker assignment. Solid line is linear regression, equation shown is the equation of this line, and dotted lines represent 95% confidence intervals. b. Residuals (modeled fecal coliform value less the actual fecal coliform value) versus unknown SourceTracker assignment.

**References**


1. Monmouth County Planning Board and M. C. H. Department, Restoring the shell-fish impaired waters of the Navesink Estuary. 2010.
2. NJDEP Bureau of Marine Water Monitoring, *Upper Navesink River Stormwater Study: Microbial Source Tracking*, 2008.
3. L. K. Dick, A. E. Bernhard, T. J. Brodeur, J. W. Santo Domingo, J. M. Simpson, S. P. Walters and K. G. Field, host distributions of uncultivated fecal *Bacteroidales* bacteria reveal genetic markers for fecal source identification, *Appl Environ Microbiol*, 2005, **71**, 3184.
4. A. E. Bernhard and K. G. Field, A PCR Assay To Discriminate Human and Ruminant Feces on the Basis of Host Differences in <em>Bacteroides-Prevotella</em> Genes Encoding 16S rRNA, *Appl Environ Microbiol*, 2000, **66**, 4571.
5. S. Seurinck, T. Defoirdt, W. Verstraete and S. D. Siciliano, Detection and quantification of the human-specific HF183 *Bacteroides* 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater, *Environ Microbiol*, 2005, **7**, 249-259.
6. B. J. Kildare, C. M. Leutenegger, B. S. McSwain, D. G. Bambic, V. B. Rajal and S. Wuertz, 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal *Bacteroidales*: a Bayesian approach, *Water Res.*, 2007, **41**, 3701-3715.
7. G. Muyzer, E. C. de Waal and A. G. Uitterlinden, Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA, *Appl Environ Microbiol.*, 1993, **59**, 695-700.