Frances Janz

June 2, 2016

Metagenome File Merge Report

The samples used for this study were collected from ten of NEON’s sites. Three types of samples were collected by field technicians per NEON’s data collection processes (National Ecological Observatory Network, 2016)[[1]](#footnote-1): individual samples for 16S marker genome testing, individual samples for metagenomic (MG) sampling, and composite (comp) samples for MG testing.

The data was collected before a standardized sample ID format was decided upon, thus much of the data was initially incomparable. To address this difficulty, I wrote two scripts in the R programming language to automate the process of correcting the sample IDs. The scrips reordered the ID’s to comply with the current accepted standard: a four letter site code and three-digit plot number separated with an underscore, an M or O to indicate horizon type (Mineral, Organic), a two-digit X-coordinate within plot, a two-digit Y-coordinate within plot, and an eight-digit date with dashes between each piece of information (e.g. HARV\_001-M-01-01-20100101). The phrase “DataMissing” was added to the end of an ID for ID’s without a date. ID’s that could not be corrected due to a substantial amount of missing information, were relabeled with an error message. This allowed all three types of data across all sites to be merged together into one file.

Once the ID’s had been corrected, the second script filtered the results of the first script to find only the sample ID’s that matched based on site code, plot number, horizon, and date. This provided the sample results below.

|  |  |  |  |
| --- | --- | --- | --- |
| **Site** | **16S per Site** | **Comp per Site** | **Comp with MG** |
| BART | 3 | 0 | 0 |
| SPER | 49 | 17 | 11 |
| DSNY | 43 | 14 | 11 |
| HARV | 13 | 2 | 2 |
| JERC | 9 | 3 | 0 |
| OSBS | 39 | 13 | 7 |
| SCBI | 12 | 4 | 0 |
| STER | 42 | 14 | 6 |
| TALL | 27 | 8 | 1 |
| WOOD | 27 | 9 | 0 |
| **Total** | 264 | 84 | 38 |

1. National Ecological Observatory Network. (2016). *TOS Science Design for Terrestrial Microbial Diversity.* [↑](#footnote-ref-1)