Recommendation:

1. Blake’s Recommendation for Date Format: *YYYYMMDD*
2. Recommendation to change the name of the Sample\_ID column to make sure that it’s consistent across the data table.
3. Recommendation to change the data file name:
   1. For qPCR datafile, each file should start with “qpcr\_\*” and follow with what’s necessary to distinguish that file from other qpcr files. End with date (yyyymmdd) if needed.
   2. For taxonomical profile, genome coverage and metadata, if the pool number is critical, start with that and add unique labels to separate it from other files.
      1. i.e:
         1. p1647\_genome\_coverage\_general\_report.all\_samples.100windows.mean\_cov”
         2. p1647\_taxonomical\_profile\_general\_report.coverm.combined.tax

**Data Definitions:**

**Metadata Folder:**

This folder contains data from Baylor. The prefix “P-XXXX” in the file names indicate different batch numbers (pool number).

***Columns*:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Folder | Column Names | Data Type | Description | ­Comments |
| Metadata | Sample\_ID | String, TEXT | These are the labels of the bottle samples from WWTP sites, BARCODE |  |
| Site | String, TEXT |  |  |
| City | String, TEXT |  |  |
| Date | Date  Format: YYYY-MM-DD | YYYY-MM-DD |  |
| Flow | Real Number, Float, Decimal |  |  |
| PoolID­ | String, TEXT | Group of sample libraries loaded together on a sequencer |  |

**qPCR Folder:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Folder | Column Names | Data Type | Description | Comments |
| qPCR | LocationAbbr | String, TEXT | Abbreviation for WW treatment plant |  |
| CMMR\_Barcode | String, TEXT |  |  |
| Target | String, TEXT |  |  |
| Ct | Real Number,  Double |  |  |
| copiesperml | Real Number, Float, Decimal |  |  |
| SampleName | String, TEXT |  |  |
| date\_of\_collection | Date  Format: m/dd/yyyy |  |  |
|  |  |  |  |

**Genome\_coverage Folder:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Folder | Column Names | Data Type | Description | Comments |
| Genome\_coverage | Sample\_ID | String, TEXT | BARCODE.POOLID |  |
| accession | String, TEXT |  |  |
| Start\_base | Integer |  |  |
| End\_base | Integer |  |  |
| Mean\_depth | Real Number, Float, Decimal |  |  |

**Site\_coding Folder:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Folder | Column Names | Data Type | Description | Comments |
| Site\_coding | Name | String, TEXT |  |  |
| Code | String, TEXT |  |  |

**Taxonomical\_profiles Folder:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Folder | Column Names | Data Type | Description | Comments |
| taxonomical\_profiles | accession | String, TEXT |  |  |
| reference\_length | Real Number,  Integer |  |  |
| covered\_base | Real Number,  Integer |  |  |
| reads\_aligned | Real Number,  Integer |  |  |
| mean\_coverage | Real Number,  double |  |  |
| RPKMF | Real Number,  double |  |  |
| sequence\_name | String, TEXT |  |  |
| taxid | String, TEXT |  |  |
| kingdom | String, TEXT |  |  |
| phylum | String, TEXT |  |  |
| class\_var | String, TEXT |  |  |
| order\_var | String, TEXT |  |  |
| family | String, TEXT |  |  |
| genus | String, TEXT |  |  |
| species | String, TEXT |  |  |
| subspecies | String, TEXT |  |  |
| strain | String, TEXT |  |  |
| sample\_ID | String, TEXT | BARCODE.POOLID | The PoolID field has been removed |
| total\_filtered\_reads\_in\_sample | Real Number,  double |  |  |
|  |  |  |  |

**WWTP\_Location Folder:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Folder | Column Names | Data Type | Description | Comments |
| Wwtp\_location | State | String, TEXT |  |  |
| County | String, TEXT |  |  |
| City | String, TEXT |  |  |
| WWTP | String, TEXT |  |  |
| lat | Real Number, Float, Decimal |  |  |
| lon | Real Number, Float, Decimal |  |  |
| county\_centroid\_lat | Real Number, Float, Decimal |  |  |
| county\_centroid\_lon | Real Number, Float, Decimal |  |  |
| city\_centroid\_lat | Real Number, Float, Decimal |  |  |
| city\_centroid\_lon | Real Number, Float, Decimal |  |  |