

Table 2. Ten randomly chosen urban microbiome samples from the MetaSUB New York dataset along with metadata such as their dates and locations of origin.

Sample ID	Full Sample Name	Sampling Time	Train Line	Transit (Borough)	Station	Swabbed Surface	Num. Read Pairs
0	haib17CEM5106_ HCY5JCCXY_SL269539	Winter 2014	D	NA		Subway train seat	5933694
1	haib17CEM5106_ HCV72CCXY_SL269726	Winter 2014	6	59 St. (Manhattan)		NA	2002519
2	haib17CEM5106_ HCVMTCCXY_SL269540	Winter 2014	D	NA		Subway train seat	2864273
3	haib17CEM5106_ HCV72CCXY_SL269724	Winter 2014	6	Grand Central (Manhattan)		NA	1814764
4	haib17DB4959_ H3MGVCCXY_SL259896	June 2017 (City Sampling Day)	NA	Van Siclen (Brooklyn)	Av.	Transit station bench	8285289
5	haib17CEM5106_ HCV72CCXY_SL269741	Winter 2014	6	NA		Subway train ceiling handrail	3864641
6	haib17CEM5106_ HCY5JCCXY_SL269601	Winter 2014	NA	36th St. (Queens)		Transit station turnstile	3760124
7	haib17CEM4890_ H2NYMCCXY_SL254808	June 2016 (City Sampling Day)	NA	Halsey (Brooklyn)	St.	Transit station turnstine	23076691
8	haib17DB4959_ H3MGVCCXY_SL259845	June 2017 (City Sampling Day)	NA	Flushing-Main (Queens)	St.	Transit station escalator handle	5349073
9	haib17CEM4890_ HMC MJCCXY_SL335779	June 2016 (City Sampling Day)	NA	33rd (Manhattan)	St.	Transit station kiosk	4834819

- **2 samples from D train**
3 samples from 6 train

and **Major train stations**

- Grand central
- Flushing-Main St

and **Minor train stations**

- Van Siclen
- Halsey St
- 36th St
- 33rd St

Note: Samples 4 & 7 have a high number of reads compared to the rest of the samples

Quality control and preprocessing of sequencing data. **(A)** Number of reads and **(B)** Bases retained

