**Table S1.** Raw and high quality read and base statistics for the three metagenomic data sets used in this study.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metagenome** | **Raw reads** | **Raw bases** | **High quality reads** | **High quality bases** |
| Simulated metagenome | 53,334,008 | 4,000,050,600 | 30,427,779 | 1,503,286,668 |
| Mock community | 6,562,065 | 492,154,875 | 3,943,759 | 206,197,036 |
| 124 gut metagenomes (average) | 45,411,302 | 3,239,648,242 | 35,789,998 | 2,049,553,760 |