Table S1. Sample environments and expected corresponding MIxS checklists/packages used in the generation of Figure 2.

| Sample environment* | Expected MIxS checklist or package | Metagenome Taxonomy ID | ENA Checklist ID |
|---------------------|-----------------------------------|------------------------|-----------------|
| air, aerosol, cloud | GSC MIxS air                      | 655179, 1708701, 1643811 | ERC000012       |
| indoor              | GSC MIxS built environment        | 1256227                | ERC000031       |
| human gut           | GSC MIxS human gut                | 408170                 | ERC000015       |
| human oral, human saliva | GSC MIxS human oral | 447426, 1679718         | ERC000016       |
| human skin          | GSC MIxS human skin               | 539655                 | ERC000017       |
| human vaginal       | GSC MIxS human vaginal            | 1632839                | ERC000018       |
| microbial mat, biofilm | GSC MIxS microbial mat biofilm | 527640, 718308          | ERC000019       |
| plant, root, shoot, moss, leaf, flower | GSC MIxS plant associated | 1297885, 1118232, 1006967, 1675540, 1297858, 1385665 | ERC000020 |
| sediment, alkali sediment, freshwater sediment, marine sediment | GSC MIxS sediment | 749907, 1680915, 556182, 412755 | ERC000021 |
| soil, soil crust, compost | GSC MIxS soil | 410658, 1710357, 702656 | ERC000022 |
| wastewater, sludge | GSC MIxS wastewater sludge       | 527639, 1592332         | ERC000023       |
| aquatic, freshwater, ballast water, marine, groundwater, lake water, rock porewater, seawater, wastewater | GSC MIxS water | 1169740, 449393, 1954210, 408172, 717931, 1647806, 1531857, 1561972, 527639 | ERC000024 |

* metagenome name