Table S3. Relative abundances of three known repair genes in different metagenomic datasets.

| Samples | Sample size[a] | MutT(198aa)[b] | MutM(260aa) | MutY(202aa) |
|---|---|---|---|---|
| North Pacific Ocean | 3 785 773 | 3476(0.092%)[c] | 784(0.021%) | 1462(0.039%) |
| Sargasso sea | 1 001 987 | 808(0.081%) | 516(0.051%) | 614(0.061%) |
| Freshwater | 2 546 538 | 492(0.019%) | 249(0.0098%) | 395(0.016%) |
| Peru Margin sediment | 237 680 | 52 (0.022%) | 32(0.013%) | 27(0.011%) |
| Minnesota Farm | 184 374 | 94(0.051%) | 48(0.026%) | 44(0.024%) |
| FACE | 13 881 342 | 1707(0.012%) | 2057(0.015%) | 1676(0.012%) |

**Note:** [a]Sample size stands for the total number of assembled peptides in each dataset; [b]Number in parentheses indicates the average protein length encoded by each repair gene; [c]The integer indicates number of peptide classified for each repair gene; the decimal in parentheses, which can be used as a measure of the relative abundance of each repair gene in each dataset, is calculated by corresponding number of protein dividing the total sample size.