**Table S5**.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sample** | **Normal (3) a** |  | **Colorectal cancer (3) a** |  | **Ulcerative colitis (3) a** |
| Search method | Strategy 0 | Strategy 1 | Strategy 2 |  | Strategy 0 | Strategy 1 | Strategy 2 |  | Strategy 0 | Strategy 1 | Strategy 2 |
| Endogenous Sequence | 12.62% | 39.94% | 58.71% |  | 11.62% | 37.75% | 56.45% |  | 12.58% | 40.49% | 60.15% |
| Bacteria Sequence b | 18.83% | 10.10% | 7.03% |  | 20.37% | 11.87% | 8.20% |  | 18.49% | 9.74% | 6.33% |
| Fungi Sequence b | 37.20% | 25.32% | 14.98% |  | 32.10% | 22.05% | 13.48% |  | 32.62% | 22.02% | 12.83% |
| Other Sequence b | 8.55% | 4.30% | 3.06% |  | 13.60% | 6.27% | 4.04% |  | 13.45% | 5.38% | 3.17% |
| Unmapped Sequence | 22.80% | 20.34% | 16.22% |  | 22.31% | 22.06% | 17.83% |  | 22.85% | 22.38% | 17.52% |

1. Numbers in parentheses represent number of samples in each group.
2. To increase the sequence mapping accuracy, we did not allow any sequence mismatch except in the endogenous sequence search step.