

### Supplementary material

**Table 1:** Predictions of the RICHEST programs on 16S rRNA sequence data

<b>N</b>	<b>Bayesian</b>	<b>PNPML</b>	<b>Good-Toumlin</b>
400 (86)	114.7 (6.1)	113.0 (6.4)	114.1 (7.3)
500 (97)	114.2 (4.2)	108.2 (4.7)	114.0 (4.6)
600 (106)	114.0 (2.8)	112.5 (3.2)	114.0 (2.8)

The first column provides the size of initial sampling of the 16S rRNA accompanied by the observed number of microbial species in parenthesis. The remaining columns provide microbial species richness predictions for the 690 16S rRNA sequences in the study from Bayesian, PNPML, and Good-Toumlin methods, respectively. The predictions are accompanied by standard deviations in parenthesis computed from repeating the experiment 10 times. The total number of observed species (i.e., the truth) is 114 [8].