Supplementary Material 1

1 The NCBI taxonomy versions used in experiments
   + MEGAN (downloaded on February 13th, 2015) uses the NCBI taxonomy downloaded on November 15, 2014 from the NCBI ftp (from release notes of MEGAN)
   + SOrt-ITEMS (downloaded on November 29th, 2014) uses the NCBI taxonomy updated in June, 2014 (by personal communication with the authors of SOrt-ITEMS).

2 The impact of parameters on the classification performance of SeMeta
   This section presents the impact of parameters min-score $s_{\text{min}}$, top-percent $p_{\text{top}}$, and max-occur $o_{\text{max}}$ on the classification performance of SeMeta. Those experiments are conducted on dataset $ds2$.

2.1 Parameter min-score $s_{\text{min}}$

![Figure S1](image1)

Figure S1 The sensitivity and precision of SeMeta with different values of parameter min-score $s_{\text{min}}$ for the scenario of known species.

![Figure S2](image2)

Figure S2 The sensitivity and precision of SeMeta with different values of parameter min-score $s_{\text{min}}$ for the scenario of unknown species.

2.2 Parameter top-percent $p_{\text{top}}$
2.3 Parameter max-occur $o_{\text{MAX}}$
2.4 The impact of parameters on SeMeta in the aspect of assigning to correct taxa exactly at the lowest levels

2.4.1 Parameter min-score $s_{min}$

2.4.2 Parameter top-percent $p_{top}$

2.4.3 Parameter max-occur $o_{max}$
3 The effect of the usage of cluster cores on SeMeta

Figure S9 The sensitivity$_B$ and precision$_B$ of SeMeta (assigning exactly at the lowest levels) with different values of parameter max-occ$\_o$ for the scenario of unknown species.

Figure S10 The sensitivity$_A$ and precision$_A$ of SeMeta and its variant which does not use cluster cores on dataset $ds_2$, the scenario of known species.

Figure S11 The sensitivity$_A$ and precision$_A$ of SeMeta and its variant which does not use cluster cores on dataset $ds_2$, the scenario of unknown species.
Figure S12 The $sensitivity_B$ and $precision_B$ of SeMeta and its variant which does not use cluster cores on dataset $ds2$, the scenario of unknown species, and for the aspect of assigning exactly at lowest levels.