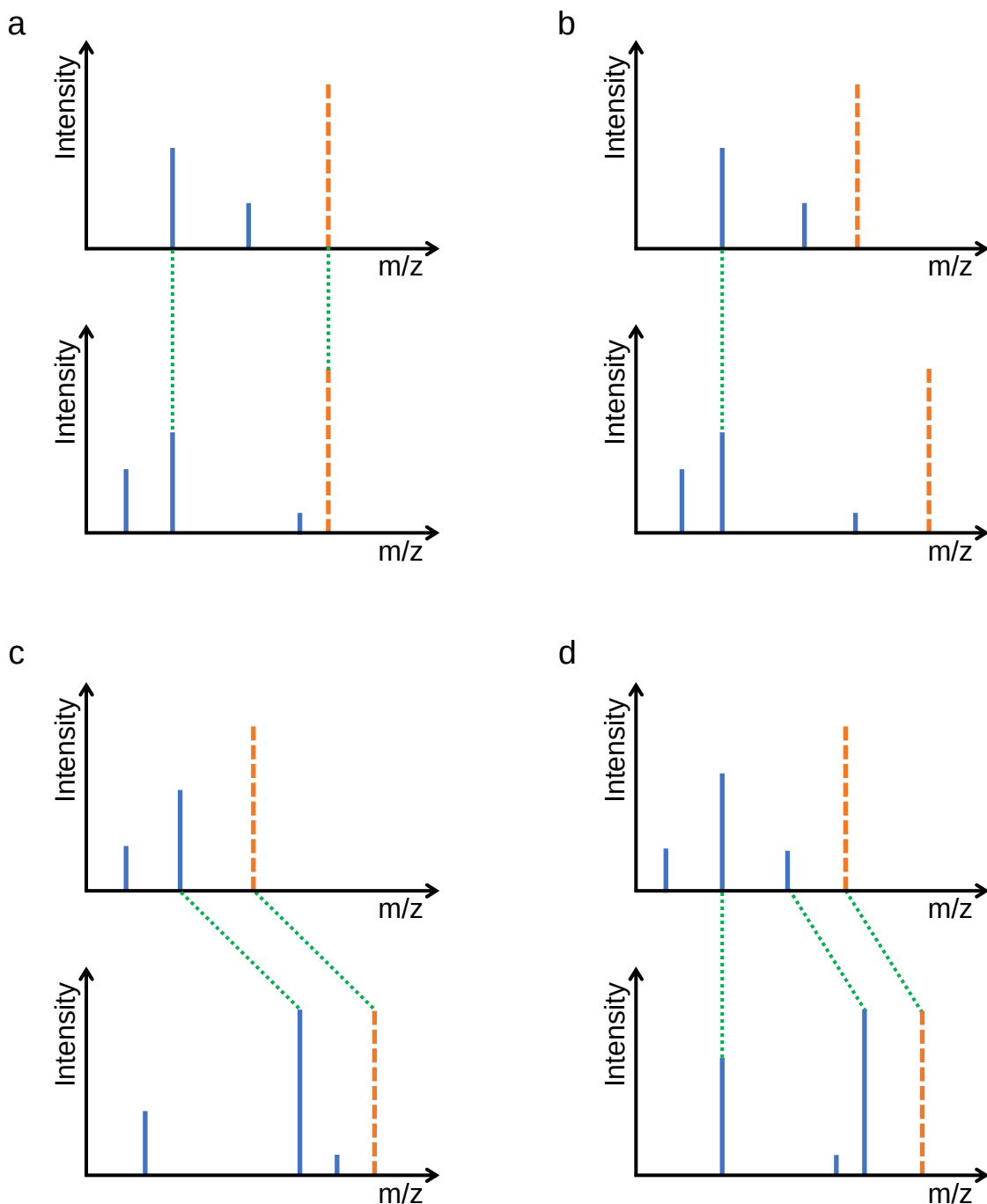


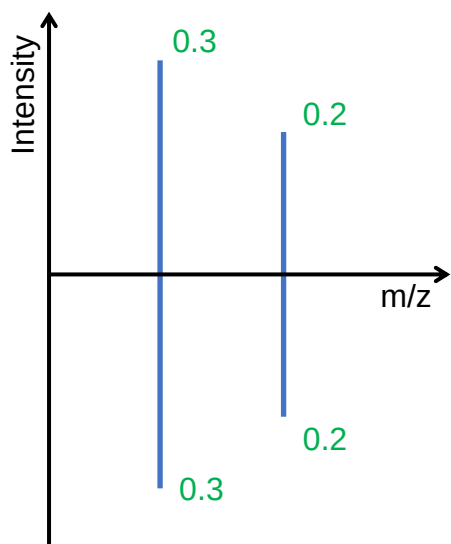


Supplementary Figure 1.

Schematic diagrams to perform MS/MS similarity searches.

(a) identity search, **(b)** open search, **(c)** neutral loss search, and **(d)** hybrid search. Fragment ions are shown in blue; precursor ions are represented by orange dashed lines. Comparisons performed by the algorithms are symbolized by dashed green lines.

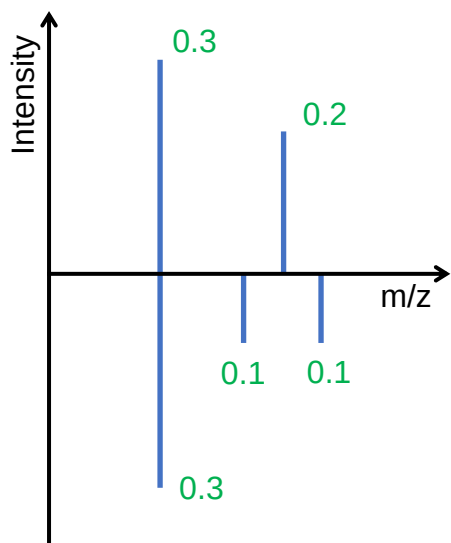
a



Entropy similarity

$$\begin{aligned}
 &= [f(0.3 + 0.3) - f(0.3) - f(0.3)] + \\
 &\quad [f(0.2 + 0.2) - f(0.2) - f(0.2)] \\
 &= [f(0.6) - f(0.3) - f(0.3)] + \\
 &\quad [f(0.4) - f(0.2) - f(0.2)] \\
 &= [0.6 \log_2 0.6 - 0.3 \log_2 0.3 - 0.3 \log_2 0.3] + \\
 &\quad [0.4 \log_2 0.4 - 0.2 \log_2 0.2 - 0.2 \log_2 0.2] \\
 &= 1
 \end{aligned}$$

b



Entropy similarity

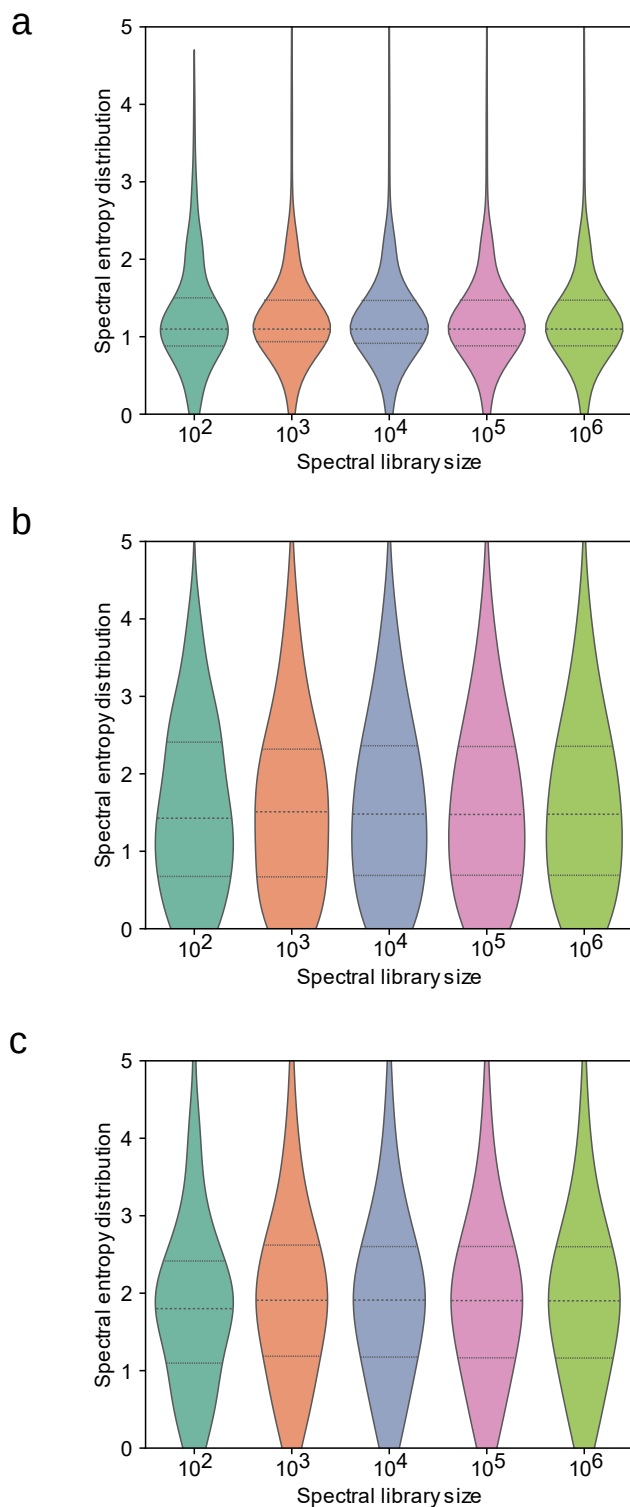
$$\begin{aligned}
 &= f(0.3 + 0.3) - f(0.3) - f(0.3) \\
 &= f(0.6) - f(0.3) - f(0.3) \\
 &= 0.6 \log_2 0.6 - 0.3 \log_2 0.3 - 0.3 \log_2 0.3 \\
 &= 0.6
 \end{aligned}$$

Supplementary Figure 2.

Examples for calculating Flash entropy similarity.

(a) Example when all ions match between query spectrum (top) and library spectrum (bottom). In the two spectra are matched.

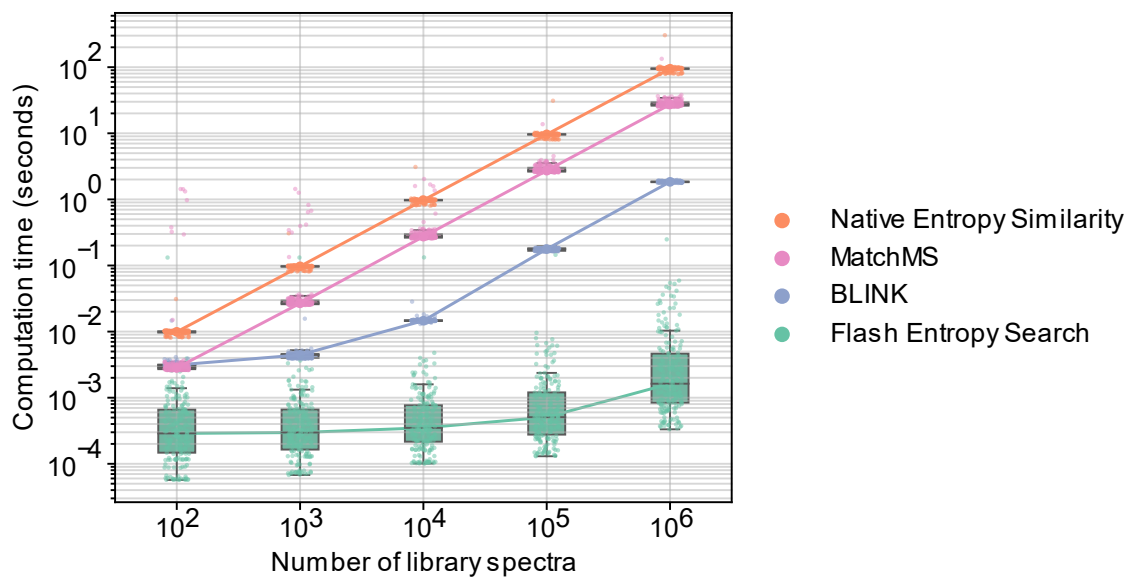
(b) Example when only one pair of ions matches between query and library spectra. Note that the sum intensities of ion abundances in each spectra are normalized to equal 0.5 (see Supplementary Note for equations). Hence, mismatched ions do not contribute themselves into the calculations, but are considered during the normalization process.



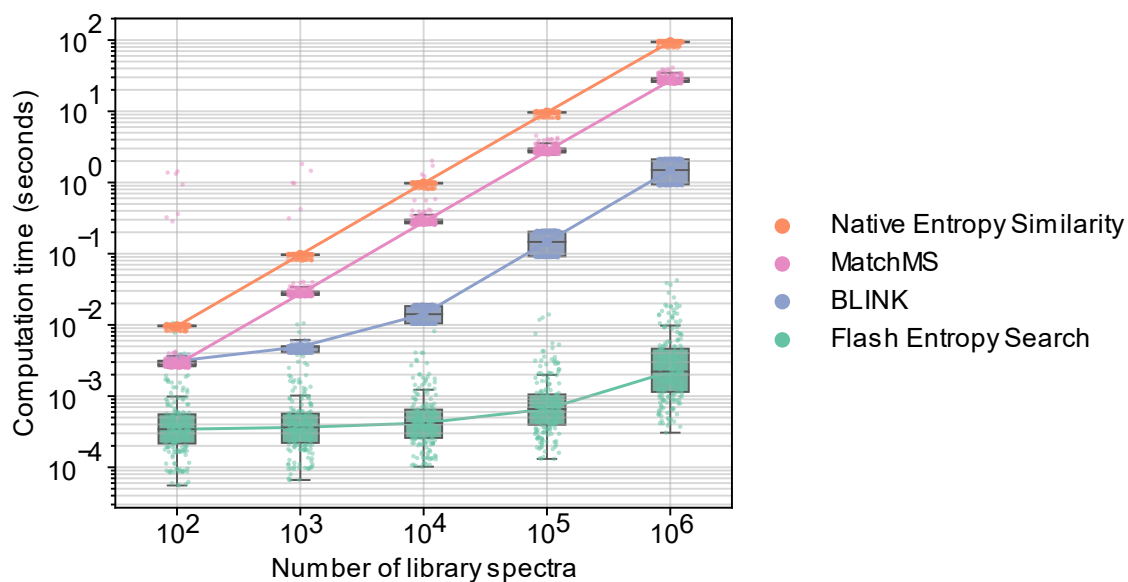
Supplementary Figure 3.

Distributions of spectral entropies when sampling spectra from different MS/MS repositories for benchmarking studies. **(a)** MassBank.us, **(b)** GNPS for annotated compounds (library), **(c)** all combined experimental public MS/MS repositories including MassIVE/GNPS, MetaboLights, MetabolomicsWorkbench, and West Coast Metabolomics Center.

a

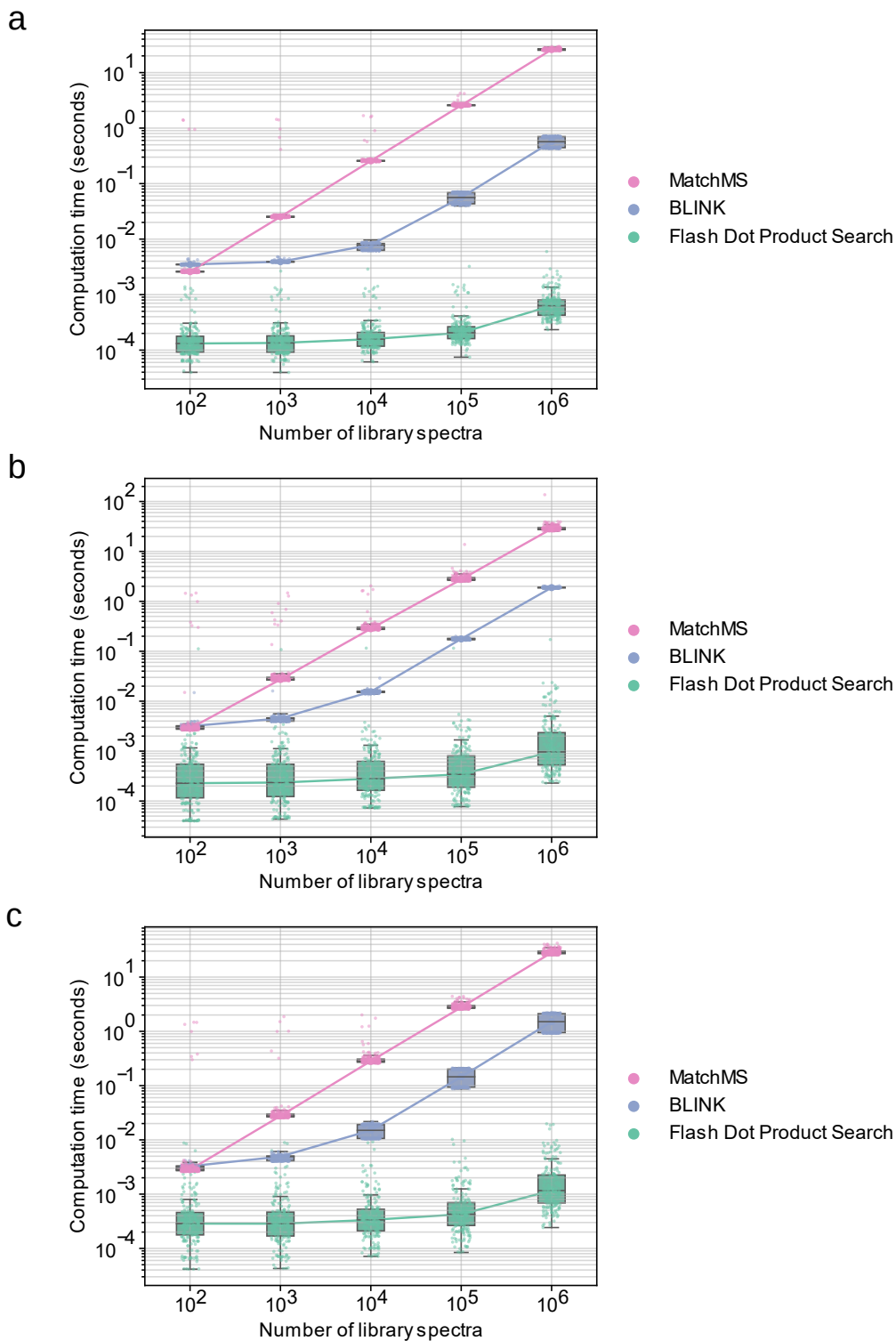


b



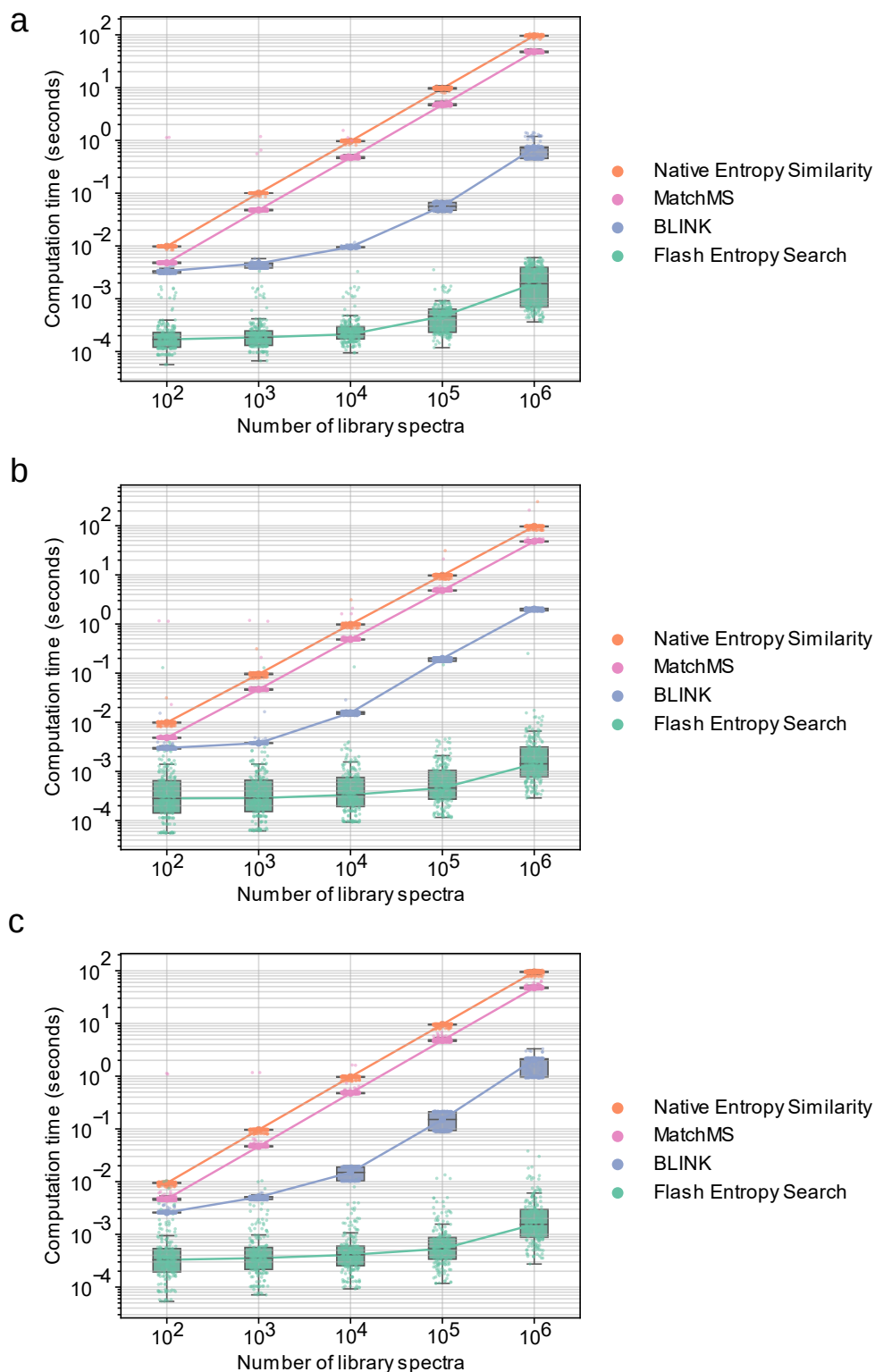
Supplementary Figure 4.

Computation time required to perform 'open search' queries using **entropy similarity** for 100 positive ESI and 100 negative ESI mass spectra against spectral libraries of different sizes. MS/MS spectra were sampled from (a) GNPS (b) public repositories.



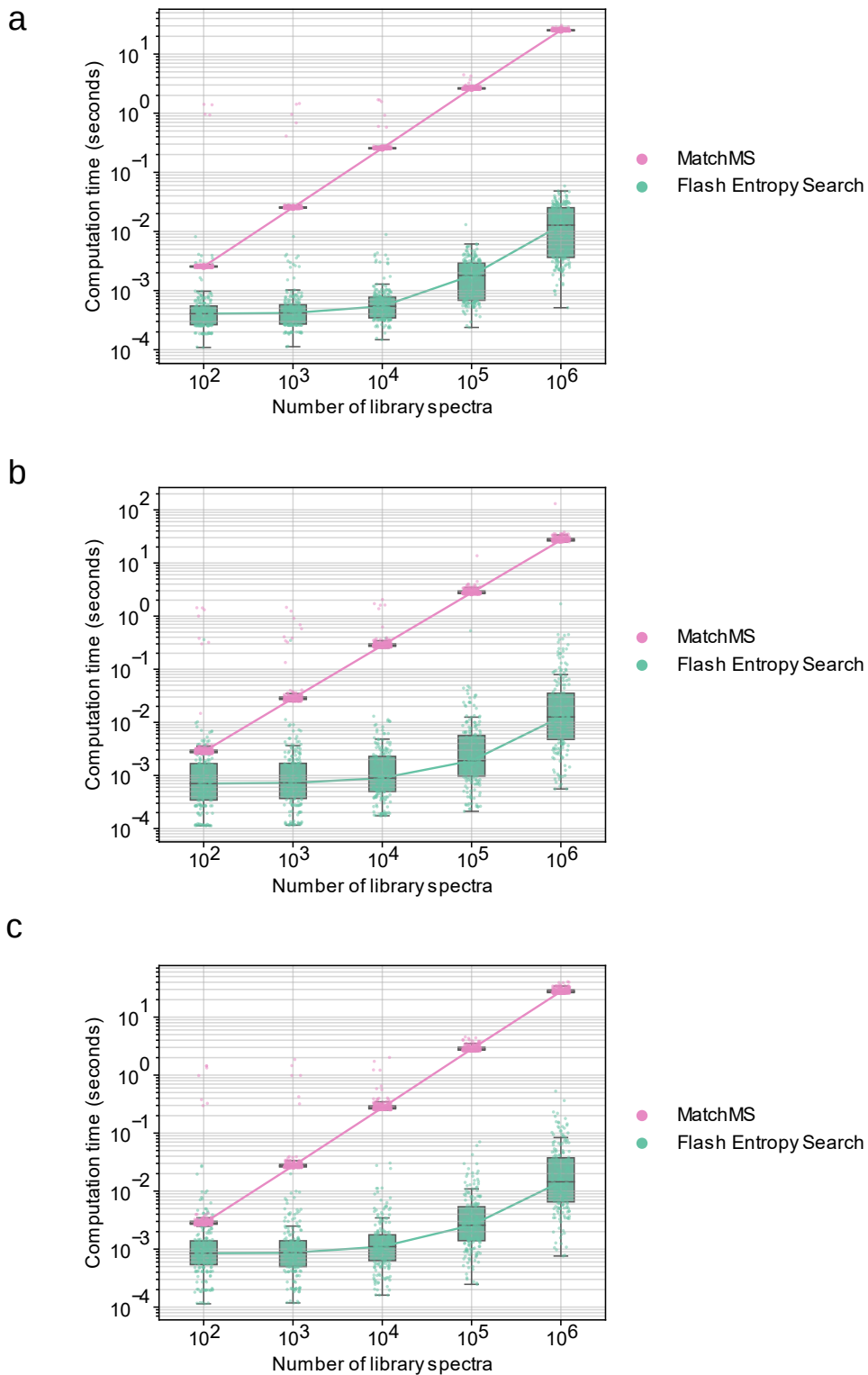
Supplementary Figure 5.

Computation time required to perform ‘**open search**’ queries using **dot product similarity** for 100 positive ESI and 100 negative ESI mass spectra against spectral libraries of different sizes. MS/MS spectra were sampled from (a) MassBank.us, (b) GNPS, (c) public repositories.



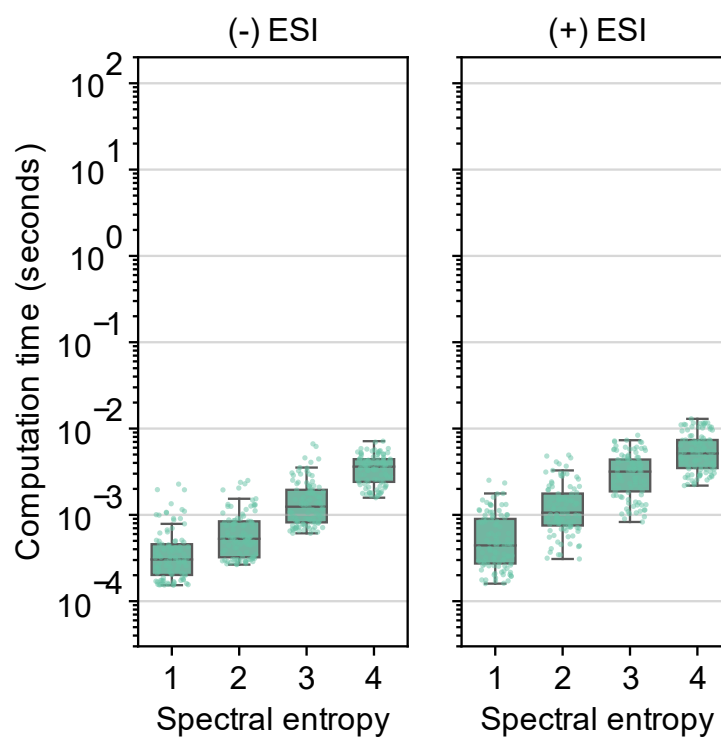
Supplementary Figure 6.

Computation time required to perform ‘**neutral loss**’ searches with **entropy similarity** for 100 positive ESI and 100 negative ESI mass spectra against spectral libraries of different sizes. MS/MS spectra were sampled from (a) MassBank.us, (b) GNPS, (c) public repositories.



Supplementary Figure 7.

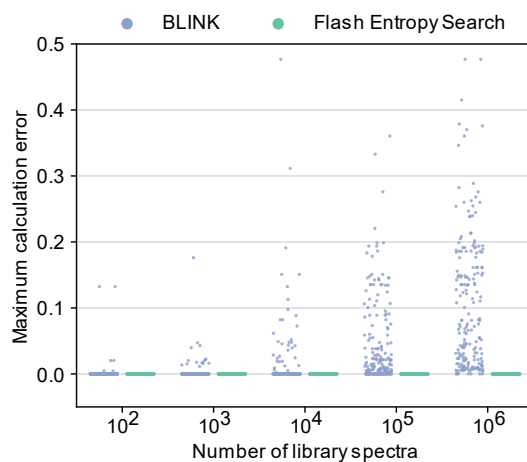
Computation time required to perform ‘**hybrid searches**’ with **entropy similarity** for 100 positive ESI and 100 negative ESI mass spectra against spectral libraries of different sizes. MS/MS spectra were sampled from (a) MassBank.us, (b) GNPS, (c) public repositories.



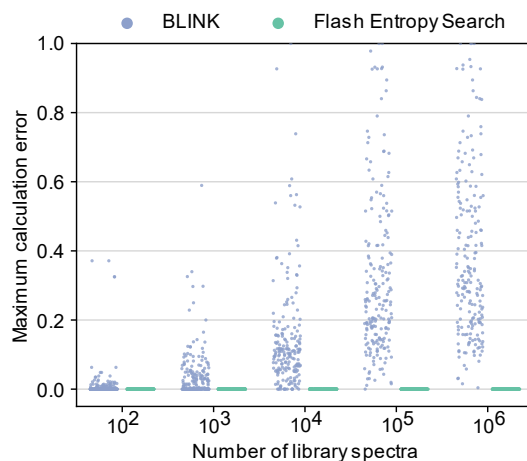
Supplementary Figure 8.

Calculation time to search 100 positive ESI and 100 negative ESI MS/MS spectra at different spectral entropy levels against randomly picked samples from the MassBank.us library.

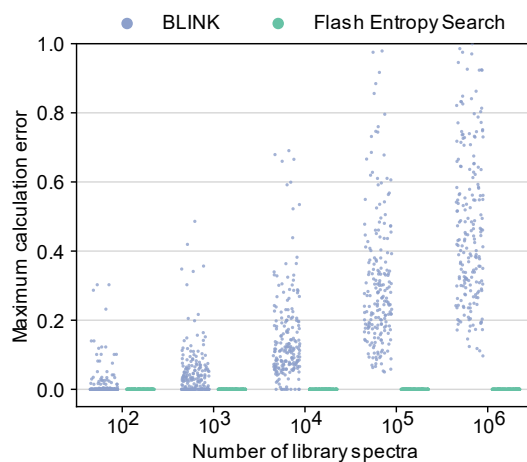
a



b



c



Supplementary Figure 9.

Comparison of the accuracy of similarity query results between Flash entropy search and BLINK. Each dot shows the maximum similarity difference between the fast algorithms and their classic algorithm counterparts. 100 positive ESI and 100 negative ESI MS/MS spectra were sampled from (a) MassBank.us, (b) GNPS, (c) public repositories.