Text S1: Calculation of the MSA Neff value.

In order to quantify the quality of an MSA, we define the number of effective sequences (Neff) as follows:

\[
Neff = \frac{1}{\sqrt{L}} \sum_{n=1}^{N} \frac{1}{1 + \sum_{m=1, m \neq n}^{N} I[S_{m,n} \geq 0.8]}
\]

where \( L \) is the length of a query protein, \( N \) is the number of sequences in the MSA, \( S_{m,n} \) is the sequence identity between the \( m \)-th and \( n \)-th sequences, and \( I[\ ] \) represents the Iverson bracket, which means \( I[S_{m,n} \geq 0.8] = 1 \) if \( S_{m,n} \geq 0.8 \) or 0 otherwise.