Resource Summary Report

Generated by nidm-terms on Jul 28, 2026

SABmark

RRID:SCR_011817
Type: Tool

Proper Citation

SABmark (RRID:SCR_011817)

Resource Information

URL: http://bioinformatics.vub.ac.be/databases/databases.html

Proper Citation: SABmark (RRID:SCR_011817)

Description: Downloadable data set designed to assess the performance of both multiple and pairwise (protein) sequence alignment algorithms, and is extremely easy to use. Currently, the database contains 2 sets, each consisting of a number of subsets with related sequences. It''s main features are: * Covers the entire known fold space (SCOP classification), with subsets provided by the ASTRAL compendium * All structures have high quality, with 100% resolved residues * Structure alignments have been derived carefully, using both SOFI and CE, and Relaxed Transitive Alignment * At most 25 sequences in each subset to avoid overrepresentation of large folds* Automated running, archiving and scoring of programs through a few Perl scripts The Twilight Zone set is divided into sequence groups that each represent a SCOP fold. All sequences within a group share a pairwise Blast e-value of at least 1, for a theoretical database size of 100 million residues. Sequence similarity is thus very low, between 0-25% identity, and a (traceable) common evolutionary origin cannot be established between most pairs even though their structures are (distantly) similar. This set therefore represents the worst case scenario for sequence alignment, which unfortunately is also the most frequent one, as most related sequences share less than 25% identity. The Superfamilies set consists of groups that each represent a SCOP superfamily, and therefore contain sequences with a (putative) common evolutionary origin. However, they share at most 50% identity, which is still challenging for any sequence alignment algorithm. Frequently, alignments are performed to establish whether or not sequences are related. To benchmark this, a second version of both the Twilight Zone and the Superfamilies set is provided, in which to each alignment problem a number of false positives, i.e. sequences not related to the original set, are added. Database specifications: * Current version: 1.65 (concurrent with PDB, SCOP and ASTRAL) * Twilight Zone set (with false positives): 209 groups, 1740 (3280) sequences, 10667 (44056) related pairs * Superfamilies set (with false positives): 425 groups, 3280 (6526) sequences, 19092 (79095) related pairs

Abbreviations: SABmark

Synonyms: SABmark - Sequence and structure Alignment Benchmark, Sequence Alignment Benchmark, Sequence and structure Alignment Benchmark

Resource Type: data or information resource, data set

Defining Citation: PMID:15333456

Resource Name: SABmark

Resource ID: SCR_011817

Alternate IDs: OMICS_00988

Record Creation Time: 20220129T080306+0000

Record Last Update: 20260728T043358+0000

Ratings and Alerts

No rating or validation information has been found for SABmark.

No alerts have been found for SABmark.

Data and Source Information

Source: SciCrunch Registry

Usage and Citation Metrics

We found 8 mentions in open access literature.

Listed below are recent publications. The full list is available at nidm-terms .

Zhai Y, et al. (2025) ReAlign-P: a vertical iterative realignment method for protein multiple sequence alignment. Bioinformatics (Oxford, England), 41(8).

Sejour R, et al. (2020) Sirt4 Modulates Oxidative Metabolism and Sensitivity to Rapamycin Through Species-Dependent Phenotypes in Drosophila mtDNA Haplotypes. G3 (Bethesda, Md.), 10(5), 1599.

Chen W, et al. (2020) pmTM-align: scalable pairwise and multiple structure alignment with Apache Spark and OpenMP. BMC bioinformatics, 21(1), 426.

Keul F, et al. (2017) PFASUM: a substitution matrix from Pfam structural alignments. BMC bioinformatics, 18(1), 293.

Rivas E, et al. (2015) Parameterizing sequence alignment with an explicit evolutionary model. BMC bioinformatics, 16, 406.

Wright ES, et al. (2015) DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment. BMC bioinformatics, 16, 322.

Bawono P, et al. (2015) Quantifying the displacement of mismatches in multiple sequence alignment benchmarks. PloS one, 10(5), e0127431.

Pei J, et al. (2006) MUMMALS: multiple sequence alignment improved by using hidden Markov models with local structural information. Nucleic acids research, 34(16), 4364.