

Resource Summary Report

Generated by [Kravitz](#) on Sep 10, 2026

CD-HIT

RRID:SCR_007105

Type: Tool

Proper Citation

CD-HIT (RRID:SCR_007105)

Resource Information

URL: <http://weizhong-lab.ucsd.edu/cd-hit/>

Proper Citation: CD-HIT (RRID:SCR_007105)

Description: THIS RESOURCE IS NO LONGER IN SERVICE. Documented on February 28,2023. Software program for clustering biological sequences with many applications in various fields such as making non-redundant databases, finding duplicates, identifying protein families, filtering sequence errors and improving sequence assembly etc. It is very fast and can handle extremely large databases. CD-HIT helps to significantly reduce the computational and manual efforts in many sequence analysis tasks and aids in understanding the data structure and correct the bias within a dataset. The CD-HIT package has CD-HIT, CD-HIT-2D, CD-HIT-EST, CD-HIT-EST-2D, CD-HIT-454, CD-HIT-PARA, PSI-CD-HIT, CD-HIT-OTU and over a dozen scripts. * CD-HIT (CD-HIT-EST) clusters similar proteins (DNAs) into clusters that meet a user-defined similarity threshold. * CD-HIT-2D (CD-HIT-EST-2D) compares 2 datasets and identifies the sequences in db2 that are similar to db1 above a threshold. * CD-HIT-454 identifies natural and artificial duplicates from pyrosequencing reads. * CD-HIT-OTU cluster rRNA tags into OTUs The usage of other programs and scripts can be found in CD-HIT user's guide. CD-HIT was originally developed by Dr. Weizhong Li at Dr. Adam Godzik's Lab at the Burnham Institute (now Sanford-Burnham Medical Research Institute)., THIS RESOURCE IS NO LONGER IN SERVICE. Documented on September 16,2025.

Abbreviations: CD-HIT

Synonyms: CD-HIT Program

Resource Type: data processing software, software application, software resource, source code

Defining Citation: [PMID:20053844](#), [PMID:16731699](#), [DOI:10.1093/bioinformatics/btl158](#)

Keywords: cluster, protein, sequence, classification, domain, analysis, nucleotide sequence, dna, protein sequence, bio.tools, FASEB list

Funding: NCRR 1R01RR025030

Availability: THIS RESOURCE IS NO LONGER IN SERVICE

Resource Name: CD-HIT

Resource ID: SCR_007105

Alternate IDs: OMICS_05157, biotools:cd-hit, nif-0000-30240

Alternate URLs: <http://cd-hit.org>, <https://code.google.com/p/cdhit/>, <https://bio.tools/cd-hit>, <https://sources.debian.org/src/cd-hit/>

Old URLs: <http://bioinformatics.ljcrf.edu/cd-hi/>

Record Creation Time: 20220129T080239+0000

Record Last Update: 20260905T132606+0000

Ratings and Alerts

No rating or validation information has been found for CD-HIT.

No alerts have been found for CD-HIT.

Data and Source Information

Source: [SciCrunch Registry](#)

Usage and Citation Metrics

We found 3223 mentions in open access literature.

Listed below are recent publications. The full list is available at [Kravitz](#).

Yang X, et al. (2026) LCPBert: ProtBERT-based early-stage lung cancer prediction from T cell receptor beta sequences. iScience, 29(5), 115581.

Begnel ER, et al. (2026) Comparison of gut virome in Kenyan infants born to women with and without HIV. iScience, 29(3), 114900.

Zheng W, et al. (2026) Mutation elevation and functional alterations in Escherichia coli are pertinent to the onset of gestational diabetes mellitus. Cell reports, 45(4), 117143.

Bjornson S, et al. (2026) Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes. Current biology : CB, 36(7), 1659.

Sharma M, et al. (2026) Improved reference assembly and core collection resequencing to facilitate exploration of important agronomical traits for the improvement of oilseed crop, Carthamus tinctorius L. GigaScience, 15.

Rusiñol M, et al. (2026) Livestock aggregated samples for monitoring viruses infecting animals and potentially zoonotic viral pathogens. *One health* (Amsterdam, Netherlands), 22, 101340.

Doré G, et al. (2026) First draft genome of the decaploid species, *Ludwigia grandiflora* subsp. *hexapetala*, validated through gene expression. *GigaByte* (Hong Kong, China), 2026, gigabyte176.

Rakesh S, et al. (2026) Expanding the landscape of BREX diversity: uncovering multi-layered functional frameworks and identification of novel BREX-related defense systems. *Nucleic acids research*, 54(3).

Li H, et al. (2026) Dysbiosis and unsustainable delayed gut microbiota development as non-invasive biomarkers for predicting autism spectrum disorder in Chinese children. *Frontiers in microbiology*, 17, 1753665.

Basharat Z, et al. (2025) Proteome mining of *Yersinia Enterocolitica* for drug targets and computational inhibitor identification with ADMET, anti-inflammation potential and formulation characteristics. *BioData mining*, 18(1), 68.

Li W, et al. (2025) Benchmarking and optimizing qualitative and quantitative pipelines in environmental metatranscriptomics using mixture controlling experiments. *ISME communications*, 5(1), ycaf090.

Gao Y, et al. (2025) Benchmarking short-read metagenomics tools for removing host contamination. *GigaScience*, 14.

Márquez R, et al. (2025) A draft genome assembly for the dart-poison frog *Phylllobates terribilis*. *GigaByte* (Hong Kong, China), 2025, gigabyte157.

van Hooff JJE, et al. (2025) Repeated duplications and losses shaped SMC complex evolution from archaeal ancestors to modern eukaryotes. *Cell reports*, 44(7), 115855.

Azam S, et al. (2025) Constructing a draft Indian cattle pangenome using short-read sequencing. *Communications biology*, 8(1), 605.

Chen XX, et al. (2025) Microbial dysbiosis with tryptophan metabolites alteration in lower respiratory tract is associated with clinical responses to anti-PD-1 immunotherapy in advanced non-small cell lung cancer. *Cancer immunology, immunotherapy* : CII, 74(4), 140.

Xue HC, et al. (2025) A unified cell atlas of vascular plants reveals cell-type foundational genes and accelerates gene discovery. *Cell*.

Naktang C, et al. (2025) Chromosome-level assemblies of two hexaploid bamboos, *Thyrsostachys oliveri* and *Thyrsostachys siamensis*, provide a foundation for functional and comparative genomics studies. *GigaScience*, 14.

Cumming J, et al. (2025) Dissecting FAP+ Cell Diversity in Pancreatic Cancer Uncovers an Interferon-Response Subtype of Cancer-Associated Fibroblasts with Tumor-Restraining Properties. *Cancer research*, 85(13), 2388.

Durán-Fuentes JA, et al. (2025) Repeatome diversity in sea anemone genomics (Cnidaria: Actiniaria) based on the Actiniaria-REPLib library. BMC genomics, 26(1), 473.