

Resource Summary Report

Generated by [FDI Lab - SciCrunch.org](#) on Apr 20, 2025

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).

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: 20250420T014352+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 3147 mentions in open access literature.

Listed below are recent publications. The full list is available at [FDI Lab - SciCrunch.org](#).

Magyar LB, et al. (2025) Pore-Forming Toxin-Like Proteins in the Anti-Parasitoid Immune Response of Drosophila. Journal of innate immunity, 17(1), 10.

- Essien C, et al. (2025) Predicting the location of coordinated metal ion-ligand binding sites using geometry-aware graph neural networks. *Computational and structural biotechnology journal*, 27, 137.
- Yi W, et al. (2025) In silico characterization of defense system hotspots in *Acinetobacter* spp. *Communications biology*, 8(1), 39.
- Chen Z, et al. (2025) UniAMP: enhancing AMP prediction using deep neural networks with inferred information of peptides. *BMC bioinformatics*, 26(1), 10.
- Sun L, et al. (2025) Shallow-water mussels (*Mytilus galloprovincialis*) adapt to deep-sea environment through transcriptomic and metagenomic insights. *Communications biology*, 8(1), 46.
- Aichelman HE, et al. (2025) Cryptic coral diversity is associated with symbioses, physiology, and response to thermal challenge. *Science advances*, 11(3), eadr5237.
- Dai W, et al. (2025) Metagenomic Insights into Pigeon Gut Microbiota Characteristics and Antibiotic-Resistant Genes. *Biology*, 14(1).
- Wu H, et al. (2025) The Protective Effects of Melatonin on Hainan Black Goats Under Heat Stress: Understanding Its Actions and Mechanisms. *Antioxidants (Basel, Switzerland)*, 14(1).
- Zeng X, et al. (2025) Effects of Different Nitrogen Fertilizer Application Rates on Soil Microbial Structure in Paddy Soil When Combined with Rice Straw Return. *Microorganisms*, 13(1).
- Tang G, et al. (2025) Metagenomic estimation of absolute bacterial biomass in the mammalian gut through host-derived read normalization. *bioRxiv : the preprint server for biology*.
- Nayfach S, et al. (2025) Engineering of CRISPR-Cas PAM recognition using deep learning of vast evolutionary data. *bioRxiv : the preprint server for biology*.
- Li S, et al. (2025) Influences of fluctuating nutrient loadings on nitrate-reducing microorganisms in rivers. *ISME communications*, 5(1), ycae168.
- Wang F, et al. (2025) The impact of straw and its post-pyrolysis incorporation on functional microbes and mineralization of organic carbon in yellow paddy soil. *PLoS one*, 20(1), e0314984.
- Yi L, et al. (2025) Chromosome-level genome assemblies of sunflower oilseed and confectionery cultivars. *Scientific data*, 12(1), 24.
- Tong C, et al. (2025) Pangenome and pantranscriptome as the new reference for gene-family characterization: A case study of basic helix-loop-helix (bHLH) genes in barley. *Plant communications*, 6(1), 101190.

Li D, et al. (2025) Seed microbiomes promote *Astragalus mongholicus* seed germination through pathogen suppression and cellulose degradation. *Microbiome*, 13(1), 23.

Kiouri DP, et al. (2025) Structure-Based Approaches for Protein-Protein Interaction Prediction Using Machine Learning and Deep Learning. *Biomolecules*, 15(1).

Du B, et al. (2025) Gut microbiota and plasma metabolites in pregnant mothers and infant atopic dermatitis: A multi-omics study. *The World Allergy Organization journal*, 18(1), 101017.

Wu X, et al. (2025) A legume-enriched diet improves metabolic health in prediabetes mediated through gut microbiome: a randomized controlled trial. *Nature communications*, 16(1), 942.

Arce-Aceves MF, et al. (2025) Fitness costs of *Mycobacterium tuberculosis* resistant to rifampicin is compensated by rapid Th2 polarization mediated by early and high IL-4 production during mice infection. *Scientific reports*, 15(1), 2811.