CD-HIT
RRID:SCR_007105
Type: Tool

Proper Citation

CD-HIT (RRID:SCR_007105)

Resource Information

**URL:** [http://weizhong-lab.ucsd.edu/cd-hit/](http://weizhong-lab.ucsd.edu/cd-hit/)

**Description:** 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-CR, 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).

**Resource Name:** CD-HIT

**Proper Citation:** CD-HIT (RRID:SCR_007105)

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

**Keywords:** cluster, protein, sequence, classification, domain, analysis, nucleotide sequence, dna, protein sequence
Resource ID: SCR_007105

Parent Organization: Google Code, University of California at San Diego; California; USA

Funding Agency: NCRR

References: PMID: 20053844, PMID: 16731699

Website Status: Last checked down

Alternate IDs: nif-0000-30240


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

Abbreviations: CD-HIT

Mentions Count: 1225

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 1225 mentions in open access literature.

Listed below are recent publications. The full list is available at FDI Lab - SciCrunch Infrastructure.


Mary Mathew K, et al. (2020) Data on large cardamom transcriptome associated with Chirke
disease. Data in brief, 29, 105047.

the global dissemination of carbapenemase-producing Escherichia coli. Genome medicine,
12(1), 10.

Gu X, et al. (2020) Comparative genomics and association analysis identifies virulence
genes of Cercospora sojina in soybean. BMC genomics, 21(1), 172.

Li S, et al. (2020) Characterization of wheat homeodomain-leucine zipper family genes and
functional analysis of TaHDZ5-6A in drought tolerance in transgenic Arabidopsis. BMC plant
biology, 20(1), 50.

Hunnicutt KE, et al. (2020) Comparative Genomic Analysis of the Pheromone Receptor
Class 1 Family (V1R) Reveals Extreme Complexity in Mouse Lemurs (Genus, Microcebus)
and a Chromosomal Hotspot across Mammals. Genome biology and evolution, 12(1), 3562-
3579.


Niu X, et al. (2020) Longitudinal analysis of the antibody repertoire of a Zika virus-infected
patient revealed dynamic changes in antibody response. Emerging microbes & infections,
9(1), 111-123.

Denecke S, et al. (2020) A transcriptomic and proteomic atlas of expression in the Nezara
viridula (Heteroptera: Pentatomidae) midgut suggests the compartmentalization of xenobiotic
metabolism and nutrient digestion. BMC genomics, 21(1), 129.


Velez G, et al. (2020) Structural Insights into the Unique Activation Mechanisms of a Non-

64.

Chu YW, et al. (2020) miRgo: integrating various off-the-shelf tools for identification of
microRNA-target interactions by heterogeneous features and a novel evaluation indicator.
Scientific reports, 10(1), 1466.

Klimenko NS, et al. (2020) Co-occurrence patterns of bacteria within microbiome of Moscow
subway. Computational and structural biotechnology journal, 18, 314-322.

Arce-Leal ÁP, et al. (2020) assembly and functional annotation oftranscriptome
from Liberibacter asiaticus infected and non-infected trees. Data in brief, 29, 105198.

