

Resource Summary Report

Generated by T1D on Aug 2, 2026

HPC-CLUST

RRID:SCR_005052

Type: Tool

Proper Citation

HPC-CLUST (RRID:SCR_005052)

Resource Information

URL: <http://meringlab.org/software/hpc-clust/>

Proper Citation: HPC-CLUST (RRID:SCR_005052)

Description: A set of tools designed to cluster large numbers (>1 million) of pre-aligned nucleotide sequences. It performs the clustering of sequences using the Hierarchical Clustering Algorithm (HCA). There are currently three different cluster metrics implemented: single-linkage, complete-linkage, and average-linkage. In addition, there are currently four sequence distance functions implemented, these are: identity (gap-gap counting as match), nogap (gap-gap being ignored), nogap-single (like nogap, but consecutive gap-nogap"s count as a single mismatch), tamura (distance is calculated with the knowledge that transitions are more likely than transversions). One advantage that HCA has over other algorithms is that instead of producing only the clustering at a given threshold, it produces the set of merges occurring at each threshold. With this approach, the clusters can afterwards very quickly be reported for every arbitrary threshold with little extra computation. This approach also allows the plotting of the variation of number of clusters with clustering threshold without requiring the clustering to be run for each threshold independently. Another feature of the way HPC-CLUST is implemented is that the single-, complete-, and average-linkage clusterings can be computed in a single run with little overhead.

Abbreviations: HPC-CLUST

Resource Type: software resource

Defining Citation: [PMID:24215029](https://pubmed.ncbi.nlm.nih.gov/24215029/)

Keywords: c++, mpi

Resource Name: HPC-CLUST

Resource ID: SCR_005052

Alternate IDs: OMICS_01446

Record Creation Time: 20220129T080228+0000

Record Last Update: 20260801T190248+0000

Ratings and Alerts

No rating or validation information has been found for HPC-CLUST.

No alerts have been found for HPC-CLUST.

Data and Source Information

Source: [SciCrunch Registry](#)

Usage and Citation Metrics

We found 5 mentions in open access literature.

Listed below are recent publications. The full list is available at [T1D](#).

Gable AL, et al. (2022) Systematic assessment of pathway databases, based on a diverse collection of user-submitted experiments. Briefings in bioinformatics, 23(5).

Wei ZG, et al. (2021) Comparison of Methods for Picking the Operational Taxonomic Units From Amplicon Sequences. Frontiers in microbiology, 12, 644012.

Bao K, et al. (2020) Proteome and Microbiome Mapping of Human Gingival Tissue in Health and Disease. Frontiers in cellular and infection microbiology, 10, 588155.

Tackmann J, et al. (2019) Rapid Inference of Direct Interactions in Large-Scale Ecological Networks from Heterogeneous Microbial Sequencing Data. Cell systems, 9(3), 286.

Matias Rodrigues JF, et al. (2014) HPC-CLUST: distributed hierarchical clustering for large sets of nucleotide sequences. Bioinformatics (Oxford, England), 30(2), 287.