

Resource Summary Report

Generated by [dkNET](#) on Sep 20, 2026

PCPipe: Protein clustering with SIMAP annotations

DOI:10.17504/protocols.io.ehfb3n

Type: Protocol

Proper Citation

Bonnie Hurwitz, Ken Youens-Clark 2016. PCPipe: Protein clustering with SIMAP annotations. protocols.io dx.doi.org/10.17504/protocols.io.ehfb3n

Protocol Information

URL: <https://dx.doi.org/10.17504/protocols.io.ehfb3n>

Authors: Bonnie Hurwitz, Ken Youens-Clark

Group: VERVE Net, Hurwitz Lab

Summary: PCPipe is a protein-clustering tool. The input is a set of ORFs and a FASTA file with already clustered ORFs. The process entails: Use cd-hit-2d to compare the input peptides to previously clustered proteins. The result is a file with input proteins that clustered to existing clusters and those that did not. Use the unclustered peptides and self-cluster them via cd-hit. Take a representative sequence from each novel cluster, and use "blastp" to compare to SIMAP. Use the resulting SIMAP "feature_id" to look up the SIMAP features, merging the query results with the protein ID into a tab-delimited annotations file. Provide the user with two cluster files and the annotations for the new clusters based on the representative sequence. Code is freely available at Github.

Affiliations: University of Arizona, University of Arizona

Version: 1

Publication Date: 2016

Proper Citation: Bonnie Hurwitz, Ken Youens-Clark 2016. PCPipe: Protein clustering with SIMAP annotations. protocols.io dx.doi.org/10.17504/protocols.io.ehfb3n

Record Creation Time: 20210329T220545+0000

Record Last Update: 20210329T221154+0000

Data and Source Information

Source: