CAZy-parser a way to extract information from the Carbohydrate-Active enZYmes Database

Rodrigo V Honorato[1]

[1]Brazilian Center for Research in Energy and Materials (CNPEM) / Brazilian Biosciences National Laboratory

18 August 2016

Summary

The Carbohydrate-Active enZYmes Database (CAZy) provides access to a sequence based classification of enzyme that are responsible for the assembly, modification and breakdown of oligo and polysaccharides (Lombard et al. 2013). This database has been online for eighteen years providing relevant genomic, structural and biochemical data on carbohydrate-active enzymes, such as glycoside hydrolases, glycosyl transferases, polysaccharide lyases, carbohydrate esterases and similar enzymes with auxiliary activities. The database is organized and presented to the user as a series of highly annotated HTML tables. In this paper a script to extract information from the CAZy according to user needs is presented. To this date, there is no way to obtain specific subsets from the database. Therefore the scripts presented in this paper will enable researchers to carry out a wide range of complex scientific tasks such as; phylogenetic analyses, multiple sequence alignment and cross-family comparisons.

References

Lombard, Vincent, Hemalatha Golaconda Ramulu, Elodie Drula, Pedro M. Coutinho, and Bernard Henrissat. 2013. "The Carbohydrate-Active Enzymes Database (CAZy) in 2013." *Nucleic Acids Research* 42 (D1). Oxford University Press (OUP): D490–D495. doi:10.1093/nar/gkt1178.