TITLE: METAGENOMIC SEQUENCE ANALYSES FROM AMAZON RIVER AND TERMITE GUT USING A BIOINFORMATICS SOFTWARE PLATFORM FOR DETECTION OF POTETIONS CELLULASES FOR BIOTECHNOLOGICAL PURPOSES

AUTHORS: LIMA, R.A.T.; SANTANA, R.H.; THOMPSON, F.L.; KRUGER, R.H

INSTITUTION: UNIVERSIDADE DE BRASÍLIA, BRASÍLIA, DF (CAMPUS DARCY RIBEIRO, CEP 70910-900, BRASÍLIA – DF, BRAZIL).

ABSTRACT:

Enzymes that are involved in the cleavage of complex carbohydrates, as well as those related to their biosynthesis, are known as carbohydrate-active enzymes (CAZYmes). They are produced by a wide spectrum of organisms and are enzymes with environmental and biotechnological relevance. Cellulases (include in glycoside hydrolases group - GHs) are essential enzymes that are widely used in various industrial fields, including the bioethanol, textile, detergent, feedstuff, food, forage, beer-brewing and pulp-and-paper industries. With such uses, cellulases have the potential to become the largest group of industrially used enzymes worldwide. Cellulose can be degraded to glucose through the synergistic action of three classes of GHs: (1) endo-β-1,4-glucanase (EC: 3.2.1.4), (2) the exo-β-1,4-cellobiohydrolases, CBH I and CBH II (EC: 3.2.1.91) and (3) β-glucosidase (EC: 3.2.1.21). In this study, we aimed to select some metagenomic sequences from Amazon River (AR) and Termites Gut (TG) that were deposited and annotated in ggKbase database by our group like one of the enzymes aforementioned and endo-1,4-β-xylanase (EC: 3.2.1.8) too, for a sequencing analysis including detection of families, domains and active sites. EC numbers were used to do the search in database. The sequences (protein) found were imported in FASTA file type to a bioinformatics software platform (Geneious 10.2.2), where sequencing analysis were done by alignment, followed by an editing (when consensus, length and have/haven't a start-stop codon were taken as features for do this), construction of trees, split of the sequences into subgroups according with the results of the trees, BLAST of the subgroups of each group of enzymes, selection of a best hit and addition of your domain and active site (using the UniProt) and finally a realignment. As a result a total of 166 (16 AR; 150 TG), 7 (1 AR; 6 TG), 300 (77 AR; 223 TG) and 423 (1 AR; 422 TG) metagenomic sequences were extract from ggKbase for endo-β-1,4-glucanase, exo-β-1,4-cellobiohydrolases, β-glucosidase and endo-1,4-β-xylanase, respectively. And the families predicted for each ones were GH5 and GH9 (endo-β-1,4-glucanase), GH8 and GH9 (exo-β-1,4-cellobiohydrolases), GH3 (β-glucosidase) and GH10 (endo-1,4-β-xylanase). Thus, the work performed by us allowed the discovery of new enzymatic sequences for cellulase that bring an advantageous gain, since these can become an important industrial tool in biotechnological applications compared to others already used.

Keywords: bioinformatics, biotechnological, carbohydrate-active enzymes, cellulases, metagenomic sequences

Development Agency: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).