Poster A-5

Annotation, comparative and evolutionary analysis of fungal Carbohydrate Active enZymes (CAZymes).

Authors:
Etienne G.J. Danchin (*AFMB, CNRS*)
Pedro M. Coutinho (*AFMB, CNRS*)
Corinne Rancurel (*AFMB, CNRS*)
Bernard Henrissat (*AFMB, CNRS*)

Short Abstract: Carbohydrate Active enZymes (CAZymes) play a central role in the biology of species having an intimate relation with sugars such as Fungi. Moreover, the wealth and biodiversity of fungal genomes already available now allows deciphering the evolutionary history and adaptations to the environment of fungal CAZymes.

Long Abstract:
INTRODUCTION Carbohydrate Active enZymes (CAZymes) have a crucial role in the biology and ecology of organisms that have intimate relations with Carbohydrates such as Fungi. Indeed, CAZymes are involved in important processes for this species such as biosynthesis and integrity of fungal cell wall, phytopathogenecity, ability to degrade carbohydrate from the biomass (starch and cellulose etc…). This latest activities account for industrial applications of enzymes secreted in high yields by Fungi (biofuels, pulp and paper manufacture, human and animal foods etc…). In parallel, Fungi pose a problem of public health as many lineages are human pathogens (for example A.fumigatus and P.cariini can be lethal in immunosupressed patients). Finally phytopathogenous lineages like G.zeae (head blight of wheat) and M.grisea (rice blast) are responsible for several billion dollars damages every year. The involvement of CAZymes in pathogenicity and fungal cell wall integrity places these enzymes as targets for the development of antifungal drugs. In parallel, Fungi represent the most important sources of eukaryotic genomes with more than 40 genomes already available and more than 50 in the pipelines of the global sequencing centers (1). This small sample of the huge Fungal biodiversity (estimated at around 1.5 million species) already includes species with various ecological niches, lifestyles and lifeforms, including both unicellular and multicellular organisms, insects, plant and human pathogens, fungi with asexual and / or sexual reproduction. The availability of these data allows the large scale annotation and comparative analysis of fungal CAZomes (set of CAZymes in a proteome). In the light of the phylogenetic positions and ecology of fungal species, the evolutionary history of CAZymes and their adaptation to the environment can be investigated. RESULTS We currently annotated 15 different fungal CAZomes, including 4 eurotiomycetes, 5 sordariomycetes, 4 saccharomycotina, 1 archiascomycete, and 3 Basidiomycetes. This consisted in the annotation of around 150,000 ORFs of which approximately 4,000 correspond to CAZymes. These enzymes are classified into various different classes and families in the CAZy database(http://afmb.cnrs-mrs.fr/CAZY/) (2). The classification is done according to the characteristic modules found in these proteins. The GH class includes catalytic modules found in Glycosidases and Tansglycosidases and are responsible for carbohydrate hydrolysis and transglycosylation. Glycosyltransferases (GTs) contain modules

responsible for biosynthesis of storage, cell wall carbohydrates and other glycoconjugates. Carbohydrate Binding Modules (CBMs) contain modules that are known to bind carbohydrates and may be associated with other modules. Polysaccharide Lyase (PL) modules are responsible for carbohydrates degradation by β-elimination. Finally the CE class includes modules found in Carbohydrate Esterases that remove carbohydrate ester-type modifications. Each class of modules has been subdivided in families which group enzymes and ancillary modules based upon sequence similarity, a common fold and mechanism of action. However, substrate (and products) specificity can be diverse among a given family (see methods). We thus classified the 4,000 CAZymes in the various different classes and families described in the CAZy db, resulting in approximately 2,600 GHs (in 60 different families); 1,300 GTs (covering 35 different families) and more than 900 CBMS, PLs and CEs. The comparative analysis of 15 fungal CAZomes allowed deciphering the general trends for each class and families of CAZymes at different scales from the individual species to the whole fungal set. At the classes scale, the main finding is that the GHs appear much more diverse both in terms of population size and families distributions than the set of GTs in Fungi. Our hypothesis is that the extracellular GHs due to their tight relation with the environment are subject to extensive adaptation whereas GTs have a more housekeeping-like behaviour. This trend can be noticed at all scales and even closely related species having different ecological behaviour present clearly different sets of GHs while not significantly different in their GT repertoires. Fungal CAZomes comparison also permitted identifying at the families' level some features that appear characteristic both to certain lineages and to certain lifestyles. METHODS Detection of CAZymes from fungal proteomes. The search for CAZyme modules (GHs, GTs, PLs, CEs and CBMs) was performed as for the daily updates of the CAZy db (2). Briefly, the protein sequences in CAZy were cut into their constitutive modules. The resulting fragments were assembled as sequence libraries (3) for BLAST searches (4). Accordingly, each fungal protein model was BLASTed against the library of approximately 50,000 modules. All models that gave an e-value passing the 0.1 threshold were automatically kept and manually analyzed. Manual analysis involved examination of the alignment of the model with the members of each family and a search of the conserved signatures/motifs characteristic of each family. The presence of the catalytic machinery was verified whenever known in the family. CAZomes comparisons. We developed a comparative analysis tool which produces a report of the CAZome content and distribution of several fungal species at a glance. From these reports, we launched several statistical analyses to identify specific features in CAZomes' composition and distribution taking into account both taxonomic and CAZyme families variability. Basically, the approach consists in importing all the CAZomes data to spreadsheets and then apply a Chi square independence test and other statistical analyses to identify the most unexpected points for a given CAZyme family / species according to the general distribution. References 1. J. E. Galagan, M. R. Henn, L.-J. Ma, C. A. Cuomo, B. Birren, Genome Res. 15, 1620 (December 1, 2005, 2005). 2. P. M. Coutinho, B. Henrissat, in Rec. Adv. Carbohydr. Bioeng., H. J. Gilbert, G. J. Davies, B. Henrissat, B. Svensson, Eds. (The Royal Society of Chemistry, Cambridge, 1999), pp. 3-14. 3. M. R. Stam, E. Blanc, P. M. Coutinho, B. Henrissat, Carbohydr Res 340, 2728 (Dec 30, 2005). 4. S. F. Altschul et al., Nucleic Acids Res. 25, 3389 (1997). 5. B. Henrissat, Biochem. J. 280, 309 (1991). 6. P. M. Coutinho, E. Deleury, G. J. Davies, B. Henrissat, J Mol Biol 328, 307 (Apr 25, 2003).