Available online at www.sciencedirect.com
ScienceDirect
Fungal Genetics and Biology 45 (2008) 638–645
FUNGAL GENETICS AND BIOLOGY
www.elsevier.com/locate/yfgbi

FOLy: An integrated database for the classification and functional annotation of fungal oxidoreductases potentially involved in the degradation of lignin and related aromatic compounds

Anthony Levasseur [a,b,*], François Piumi [a], Pedro M. Coutinho [c], Corinne Rancurel [c], Michèle Asther [a], Michel Delattre [a], Bernard Henrissat [c], Pierre Pontarotti [b], Marcel Asther [a], Eric Record [a]

[a] *UMR 1163 INRA de Biotechnologie des Champignons Filamenteux, IFR86-BAIM, Universités de Provence et de la Méditerranée, ESIL, 163 Avenue de Luminy, Case Postale 925, 13288 Marseille Cedex 09, France*
[b] *Laboratoire de Phylogénomique, EA 3781 Evolution Biologique Université de Provence, Case 19, Pl. V. Hugo, 13331 Marseille Cedex 03, France*
[c] *Glycogenomics and Biomedical Structural Biology, AFMB Laboratory—UMR 6098—CNRS Universités d'Aix-Marseille I et II, Case 932, 171 Avenue de Luminy, 13288 Marseille Cedex 09, France*

Received 14 August 2007; accepted 15 January 2008
Available online 26 January 2008

Abstract

The breakdown of lignin by fungi is a key step during carbon recycling in terrestrial ecosystems. This process is of great interest for green and white biotechnological applications. Given the importance of these enzymatic processes, we have classified the enzymes potentially involved in lignin catabolism into sequence-based families and integrated them in a newly developed database, designated Fungal Oxidative Lignin enzymes (FOLy). Families were defined after sequence similarity searches starting from protein sequences and validated by the convergence of results with biochemical experiments reported in the literature. The resulting database was applied as a tool for the functional annotation of genomes from different fungi, namely (i) the Basidiomycota *Coprinopsis cinerea, Phanerochaete chrysosporium* and *Ustilago maydis* and (ii) the Ascomycota *Aspergillus nidulans* and *Trichoderma reesei*. Genomic comparison of the oxidoreductases of these fungi revealed significant differences in the putative enzyme arsenals. Two Ascomycota fungal genomes were annotated and new candidate genes were identified that could be useful for lignin degradation and (or) melanin synthesis, and their function investigated experimentally. This database efforts aims at providing the means to get new insights for the understanding and biotechnological exploitation of the lignin degradation. A WWW server giving access to the routinely updated FOLy classifications of enzymes potentially involved in lignin degradation can be found at http://foly.esil.univ-mrs.fr.

Keywords: Functional classification; Lignin degradation; Fungal genomes; Functional annotation; Database; Green biotechnology

1. Introduction

Lignin biodegradation is a key step during carbon recycling in terrestrial ecosystems. Formed through oxidation and free radical coupling of phenyl alcohol precursors, lignin is an insoluble polymer that is highly resistant to chemical and biological degradation (for a review see Cullen and Kersten, 2004; Martinez et al., 2005). Among the various groups of micro-organisms able to degrade lignin, the naturally-occurring wood-rotting basidiomycetous fungi are the most efficient. However, despite their proficient ability to degrade lignin, white-rot fungi are unable to utilise lignin as the sole carbon or energy source and depolymerisa-

* Corresponding author. Address: UMR 1163 INRA de Biotechnologie des Champignons Filamenteux, IFR86-BAIM, Universités de Provence et de la Méditerranée, ESIL, 163 Avenue de Luminy, Case Postale 925, 13288 Marseille Cedex 09, France.
E-mail address: Anthony.Levasseur@esil.univmed.fr (A. Levasseur).

1087-1845/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.fgb.2008.01.004

tion provides access to the cellulose and hemicellulose encrusted into the lignin matrix.

Enzymatic carbohydrate degradation requires cocktails of very specific extracellular enzymes targeting each of the labile glycosidic bonds present in the substrate. By contrast, lignin degradation makes use of large panels of extracellular oxidative enzymes that initiate the lignin depolymerisation. They are responsible for generating highly reactive non-specific free radicals, which then effect the cleavage of carbon–carbon and ether inter-unit bonds. The non-specificity of these lignin-degrading enzymes (named Lignin Oxidase, LO), coupled to their high oxidation potential, is atypical and has attracted considerable interest for industrial applications. Additional oxidative enzymes act as auxiliary enzymes (named Lignin Degrading Auxiliary enzymes, LDAs), in synergy with the main degrading enzymes and (or) indirectly on lignin. However, lignin degradation is still not completely understood, as wood-decay fungi employ different enzymatic strategies. For instance, lignin degradation by the Basidiomycete *Pycnoporus cinnabarinus* seems to rely solely on laccase activity (Eggert et al., 1997), while the well-known lignin degrader *Phanerochaete chrysosporium* depends exclusively on secreted peroxidases (Kersten and Cullen, 2007). Such diversity is also observed in the genetic repertoire of ligninolytic enzymes: *P. chrysosporium* posseses over a dozen of different peroxidases genes but no true laccases sequences, while *Coprinopsis cinerea* encodes an abundant set of laccases but only one peroxidase (Kilaru et al., 2006, this study). The growing number of genomes sequenced (including Basidiomycota) offers an opportunity to evaluate and compare the potential of fungi for the degradation of lignin and related aromatic compounds. This task requires a rigorous classification of the ligninolytic enzymes in well-defined families with an efficient system for genome annotation. At present, carbohydrate-active enzymes are classified into several sequence- and structure-based family schemes (Henrissat, 1991; Campbell et al., 1997; Coutinho and Henrissat, 1999; Coutinho et al., 2003) grouped in an integrated database (CAZy; http://www.cazy.org). CAZy describes families of structurally related functional modules of enzymes that degrade, modify or create glycosidic bonds and their ancillary carbohydrate-binding modules. This not only provides a means for rationalising enzymatic action on glycosidic bonds (Davies and Henrissat, 1995), but has been applied to describe and elucidate major aspects of the carbohydrate metabolism of fungi (Martinez et al., 2004; Pel et al., 2007) and other organisms (Tuskan et al., 2006; Taylor et al., 2006), with important consequences for both fundamental and applied knowledge. In contrast to the carbohydrate-active enzymes, no systematic integrated effort to classify ligninolytic enzymes has been made to date, which limits knowledge and consequently hinders biotechnological developments.

The importance of wood-rot fungi in the environment and the release of fungal genomes prompted us to develop an integrated database focusing on the enzymes of fungal origin potentially involved in lignin degradation. This work resulted in the classification of 379 full-length and 601 partial sequences into 10 families of oxidative and auxiliary degrading enzymes, which are presented here.

These enzymes and related sequences were classified based on their sequence and structure and integrated into the new Fungal Oxidative Lignin enzymes (FOLy)[1] database. Complete annotations of the potential ligninolytic system of the three Basidiomycota (*C. cinerea*, *P. chrysosporium*, *Ustilago maydis*) and two Ascomycota (*Aspergillus nidulans* and *Trichoderma reesei*) were performed, lending first comparative insight into the diversity of fungal lignin degradation.

2. Materials and methods

2.1. Database architecture and enzyme classification

FOLy is a relational database developed under MySQL interfaced with PHP and Perl as scripts. The database structure was based on that of CAZy (Coutinho and Henrissat, 1999), and relies on family-based management of sequence information, the corresponding accessions in a number of public databases, and its modular description. Each protein sequence was assigned to an organism and tagged using the NCBI taxonomy (Woodsmall and Benson, 1993). Following database and literature searches, only the unequivocal identification of the biochemical characteristics of the corresponding enzyme allowed the allocation of an E.C. number according to the Enzyme Classification nomenclature (http://www.chem.qmul.ac.uk/iubmb/enzyme/). A number of public database accessions were used: (i) the sequence databases GenBank (Benson et al., 2007) and UniProt (http://www.expasy.uniprot.org/), (ii) the structure database PDB (http://www.rcsb.org/pdb/home/home.do), (iii) the biochemical characterisation databases EMP (Burgard and Maranas, 2001) and PMD (Kawabata et al., 1999) and (iv) the literature database PubMed (Woodsmall and Benson, 1993) complemented by other references using Digital Object Identifier (DOI) tags (http://www.doi.org/). The modular description of each sequence described its main characteristics internally: globular domains of catalytic and auxiliary modules, plus other regions such as signal peptides, pro-peptides lost on maturation, linker regions, anchoring transmembrane regions, etc. Such descriptors were obtained from the analysis of (i) protein structures, (ii) protein sequence alignments, (iii) hydrophobic cluster analysis (Gaboriaud et al., 1987) and (iv) other sequence features using external bioinformatics resources such as Pfam (Finn et al.,

[1] *Abbreviations used:* FOLy, Fungal Oxidative Lignin enzymes database; FOLyme, candidate gene belonging to one family; FOLome, whole candidate genes detected in one genome; LO, Lignin Oxidase; LDA, Lignin Degrading Auxiliary enzyme.