ELSEVIER

British Mycological Society promoting fungal science

journal homepage: www.elsevier.com/locate/mycres

An expanded family of fungalysin extracellular metallopeptidases of *Coprinopsis cinerea*

Walt W. LILLY[a,*], *Jason E. STAJICH*[b,†], *Patricia J. PUKKILA*[c], *Sarah K. WILKE*[a,‡], *Noriko INOGUCHI*[a], *Allen C. GATHMAN*[a]

[a]*Department of Biology, Southeast Missouri State University, Cape Girardeau, MO 63701, USA*
[b]*Department of Molecular Genetics and Molecular Biology, Duke University, Durham, NC 27708, USA*
[c]*Department of Biology, University of North Carolina, Chapel Hill, NC 27599, USA*

ARTICLE INFO

Article history:
Received 9 September 2007
Received in revised form
14 November 2007
Accepted 29 November 2007
Corresponding Editor:
Nicholas P. Money

Keywords:
Basidiomycota
Coprinus cinereus
FTP
Fungalysin
Peptidases
Proteases
Thermolysin

ABSTRACT

Proteolytic enzymes, particularly secreted proteases of fungal origin, are among the most important of industrial enzymes, yet the biochemical properties and substrate specificities of these proteins have been difficult to characterize. Genomic sequencing offers a powerful tool to identify potentially novel proteases. The genome of the model basidiomycete *Coprinopsis cinereus* was found to have an unusually high number of metalloproteases that closely match the M36 peptidase family known as fungalysins. The eight predicted *C. cinereus* fungalysins divide into two groups upon comparison with fungalysins from other fungi. One member, CcMEP1, is most similar to the single representative fungalysins from the basidiomycetes *Phanerochaete chrysosporium*, *Cryptococcus neoformans*, and *Ustilago maydis*, and to the fungalysin type-protein from *Aspergillus fumigatus*. The remaining seven *C. cinereus* predicted fungalysins form a group with similarity to three predicted M36 peptidases of *Laccaria bicolor*. All eight of the *C. cinereus* enzymes contain both the signature M36 Pfam domain and the FTP propeptide domain. All contain large propeptides with considerable sequence conservation near a proposed cleavage site. The predicted mature enzymes range in size from 37–46 kDa and have isoelectric points that are mildly acidic to neutral. The proximity of these genes to telomeres and/or to transposable elements may have contributed to the expansion of this gene family in *C. cinereus*.

Introduction

External digestion of macromolecular nutrients is a general feature of fungal growth and is dependent upon the secretion of hydrolytic enzymes. Among the most important of these hydrolases are those with proteolytic activity. Proteolytic enzymes are mechanistically diverse and include both those requiring highly specific peptide bond substrates, and others that can hydrolyse a broad spectrum of peptides. They are divided into seven major mechanistic classes, including aspartyl proteases, cysteine proteases, glutamyl proteases, metalloproteases, serine proteases, threonine proteases, and those of unknown mechanism. Multiple studies have identified proteolytic activities, generally attributable to serine proteases and metalloproteases, in the growth medium supporting actively growing basidiomycetes. The model basidiomycetes

* Corresponding author.
E-mail address: wlilly@semo.edu
† Current Address: Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA.
‡ Current Address: Department of Biology, University of North Carolina, Chapel Hill, NC 27599, USA.
0953-7562/$ – see front matter © 2007 The British Mycological Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.mycres.2007.11.013

Schizophyllum commune (Hummel *et al.* 1998), *Ustilago maydis* (Hellmich & Schauz 1988), *Agaricus bisporus* (Burton *et al.* 1997), and *Coprinopsis cinerea* (Kalisz *et al.* 1987) all show secreted protease activity, and in most cases, the nitrogen status of the nutrient base affects the amounts (and perhaps types) of protease activity secreted. These older biochemical studies suffered from a lack of understanding of the number of protease genes present in fungal organisms, and often focused on measuring the total activity of an entire class of proteases, rather than the individual contributions of specific enzymes. The lack of resolution of these biochemical analyses also tended to underestimate the total number of enzymes contributing to extracellular proteolysis. In the only published biochemical study of *C. cinerea* proteases, Kalisz *et al.* (1987) identified only five proteases of all classes in culture filtrates using gelatin-containing polyacrylamide gel electrophoresis. We encountered similar resolution difficulties in *S. commune* (Hummel *et al.* 1998) and in *C. cinerea* (unpubl.). Such poor understanding of the extracellular enzyme activities has made it virtually impossible to gain insight into substrate specificity and regulation of individual enzymes.

Recently, genome sequencing of fungi has revealed a large number of predicted fungal extracellular proteases and has provided a basis for investigating the function and specificity of individual enzymes. For example, genomic analysis in the white-rot basidiomycete *Phanerochaete chrysosporium* predicted 52 extracellular peptidases, of which 31 were subsequently found by proteomic methods in ligninolytic cultures (Vanden Wymelenberg *et al.* 2006). In this study, we report the discovery of a large, expanded family of extracellular fungalysin (peptidase M36 family) metalloproteases, based on genome-wide predictions for the mushroom-producing basidiomycete *C. cinerea*. Fungalysins are metalloproteases that are closely related to the bacterial thermolysin family. They were first discovered in *Aspergillus fumigatus* (Markaryan *et al.* 1994), but their role in non-pathogenic fungi is unknown. Here we report the sequence conservation, expression, and genomic location of this gene family.

Materials and methods

Discovery and annotation of the M36 peptidase family

Multiple gene models have been developed for *Coprinopsis cinerea*. A reference set of gene predictions produced using Augustus, GeneZilla, and SNAP predictions has been deposited in GenBank by the Broad Institute. Additional training of GLEAN and GeneZilla software, combined with setting an intron length upper limit of 300 nt, has provided our group with what appears to be a more reliable working set of gene predictions. These are referred to as 'GLEAN_GZ2_Jan06max300' models (subsequently abbreviated Jan06m300). The latter set are downloadable from http://fungal.genome.duke.edu, and most of the available gene models, including the Broad predictions, can be viewed in their genomic context on GBrowse at http://genome.semo.edu/. We screened the *C. cinerea* Jan06m300 predictions by performing a BLASTP search against the MEROPS peptidase database (http://merops.sanger.ac.uk/). Predicted cellular location of these was determined using SignalP 3.0 (Bendtsen *et al.* 2004), and WoLFPSORT (Horton *et al.* 2007). Manual annotation of the M36 peptidase genes was performed by multiple alignment analysis of the Broad predictions, the Jan06m300 models, and available EST data using ClustalW. Where ambiguity of intron splicing was present in the predictions, and EST data were available, the ESTs were taken as correct. In two instances (*CcMEP1* introns 2 and 3) conflicting intron predictions were resolved by PCR analysis. The individual M36 predictions were further compared by alignment to each other and to known M36 sequences from other fungi (particularly, *Aspergillus fumigatus*). The resulting manually annotated genes were then named *CcMEP1* through *CcMEP8*.

Phylogenetic methods

Amino acid sequences were aligned to the Pfam profile HMM of the Peptidase_M36 domain using the hmmalign program of the HMMER package (version 2.3.2; http://hmmer.janelia.org). A Bayesian consensus phylogenetic tree of the sequences was constructed using the MrBayes software package (version 3.1.2) (Ronquist & Huelsenbeck 2003) using mixed aamodel with three runs and four MCMC chains that converged after 800 iterations. ML BS values were calculated for the Bayesian consensus tree with RAxML (version 2.2.3) (Stamatakis 2006) using PROTMIXBLOSUM62 (as BLOSUM was the optimal amino acid module found by MrBayes) and 100 replicates.

Voucher material

Coprinopsis cinerea (syn. *Coprinus cinereus*) str 130, Okayama 7, is available from Patricia Pukkila, Department of Biology, University of North Carolina-Chapel Hill.

Results

Multiple members of family M36 peptidases in Coprinus cinereus

The BLASTP search of *Coprinopsis cinerea* Jan06m300 translated gene predictions against the MEROPS protease database identified 301 unduplicated genes potentially encoding proteolytic enzymes, or non-enzymatic homologues of proteases (E values of 1×10^{-10} or less; data on other proteases will be presented elsewhere). The cellular locations of the predicted protease gene products were then analysed using two methods of location prediction: SignalP 3.0 and WoLFPSORT. These methods provide tests for extracellularity based on presence of a signal sequence (in SignalP 3.0) or the combined features of signal sequence and full protein amino acid composition (in WoLFPSORT). These programs predicted a total of 105 proteases to be extracellular (100 predicted by both programs, three predicted only by SignalP, and two predicted only by WoLFPSORT). Of these 105 putative extracellular proteases, 50 are metalloproteases, 40 are serine proteases, ten are aspartic proteases, three are cysteine proteases, and one is a threonine protease. Members of 21 different MEROPS families are represented; however, six families (M28, M36, M43, S8, S9, and A1) comprise two-thirds of the total predicted