Molecular and structural characterization of cDNA from a cellulase of Moniliophthora perniciosa (Stahel) Aime & Phillips-Mora

https://doi.org/10.56238/uniknowindevolp-144

Cleidineia Souza de Santana
Master's Degree in Biotechnology, State University of Feira de Santana
LATTES: http://lattes.cnpq.br/9431304301043940
E-mail: ssasntanacleide@gmail.com

Alison Borges Vitor
Master's Degree in Plant Genetic Resources, State University of Feira de Santana
LATTES: http://lattes.cnpq.br/4146428710797468
E-mail: alisonborgesvictor@gmail.com

Edjane Bastos Ferreira
Master's Degree in Biotechnology, State University of Feira de Santana
LATTES: http://lattes.cnpq.br/3465031695215053
E-mail: edjanebferreira@gmail.com

Luiz Henrique Machado Oliveira
Master's Degree in Biotechnology, State University of Feira de Santana
LATTES: http://lattes.cnpq.br/8445169113030900
E-mail: henriquebiomol@gmail.com

Jamille de Santana Melo Adorno
Master's Degree in Biotechnology, State University of Feira de Santana
LATTES: http://lattes.cnpq.br/8445169113030900
E-mail: jamilleadorno@yahoo.com.br

Sandra Aparecida de Assis
PhD in Biotechnology, State University of Feira de Santana
LATTES: http://lattes.cnpq.br/7633115237379209
E-mail: sandrinhaassis@yahoo.com.br

Raquel Guimarães Benevides
PhD in Biochemistry, State University of Feira de Santana
LATTES: http://lattes.cnpq.br/7633115237379209
E-mail: raquelgb@gmail.com

ABSTRACT
The basidiomycete fungus Moniliophthora perniciosa (Stahel) Aime & Phillips-Mora, is a phytopathogen, causing witches' broom disease and is known as a potential producer of cellulases enzymes. These enzymes act on the degradation of the plant cell wall and can be produced by numerous microorganisms, with filamentous fungi being the main producers. Cellulases have a wide variety of industrial applications and the production of cellulases from microbial sources is an excellent alternative for the production of cellulolytic enzymes such as endoglucanase. This work presents the expression related to a hypothetical protein of the GH45 family of the fungus Moniliophthora perniciosa, through cDNA analysis and the determination of its 3D structure by comparative modeling, presenting a high quality model, showing a GH45 with a classic structure characteristic of endoglucanases.

Keywords: Fungus, Endoglucanase, Comparative modeling, Protein, Heterologous production.

1 INTRODUCTION

The fungus *Moniliophthora perniciosa* is a phytopathogen, known as the causative agent of witches' broom disease (Basso, 2015; Meinhardt *et al*., 2008). This pathogen affects the *Theobroma cacao* species and is among the most devastating diseases affecting cocoa plantations, occurring mainly in South American countries. In Brazil, it affects cocoa crops in Bahia and the Amazon, and in Bahia the disease has caused the near collapse of the cocoa crop (Barbosa, *et al.*, 2018; Scarpari, 2005).

Pathogenic microorganisms have developed remarkable mechanisms, including the secretion of an arsenal of proteins and small molecules called effectors, which modulate the plant's defense to increase the availability of nutrients to the parasite, allowing infection (Alvim, 2009; Jones & Dangl, 2006; Kamoun, 2007).

For the colonization of plants, the fungus requires the action of secreted hydrolytic enzymes, allowing their passage through the cell wall. This characterizes the initial process of plant infection (Bhat, 2000; Guedes, *et al*., 2014).

This strategy of infection by *M. perniciosa* drew attention to a possible potential in the production of hydrolytic enzymes. These enzymes could hydrolyze the β-1,4-glycosidic bond of the cellulose chain, which is the main component of the cell wall of plant biomass (Guedes *et al.,* 2014; Nimlos, 2012)

Complete hydrolysis of cellulose requires the synergistic cooperation of cellulases, which include: endoglucans (1,4-β-D-endoglucanase EC 3.2.1.4), cellobiohydrolases or exoglucanase (1,4-β-cellobiohydrolase EC 3.2.1.91), and beta-glucosidases (β-D-glucoside glucohydrolase EC 3.2.1.21) (Jorgensen, et al., 2007; Mhlongo, 2015).

Cellulases have a wide variety of industrial applications. In recent years, these enzymes have been increasingly used in the production of biofuels, during the hydrolysis stage of lignocellulosic biomass to convert cellulose into glucose (9). Fungal cellulases have been the most studied and exploited industrially about cellulose degradation and production of cellulolytic enzymes (Esposito; Azevedo, 2004; Srivastava, 2018).

Scientific contributions have been made successively over the years regarding the isolation of cellulase-producing microorganisms. Some of these contributions are related to the increase in the expression of cellulases by genetic engineering, purification and characterization of components of this enzyme complex, cloning and gene expression, determination of three-dimensional structures of cellulases and demonstration of the industrial potential of these enzymes (Castro; Perreira Jr, 2010).

To produce cellulases on a commercial scale, recombinant expression in prokaryotes such as *Escherichia coli and* in eukaryotes such as *Pichia pastoris* allows for robust and cost-effective production (Demain; Vaishnav, 2009; Zeeshan, *et al.*, 2018). In particular, endo-1,4-β-glucanases were produced by cloning and expression in *E. coli*, in studies previously found in the literature (Kleman-Leyer, *et al*. (1996, Tomme, *et al.*, 1988; Woodward, 1982).

In this work we propose, from the cDNA collection and analysis, the verification of the expression of a hypothetical protein of the GH45 family of the fungus *Moniliophthora perniciosa* involved in the degradation of cellulose using cellulolytic residues as substrate, as well as the prediction of its 3D structure by comparative modeling.

2 MATERIALS AND METHODS

The fungus *Moniliophtora perniciosa* CCMB561 was obtained from the Collection of Microorganism Cultures of Bahia (CCMB), reactivated on Potato Dextrose Agar (PDA) and incubated in B.O.D at 27 ºC for 8 days.

To induce the expression of the cellulolytic complex, two different culture media were tested, the semi-solid culture medium WY (40 g of wheat bran; 6 g of yeast extract; 1.0 g of potassium phosphate; 0.2 g of magnesium sulfate; 0.2 g of potassium chloride; 1L of distilled water) and the liquid medium CMC (NH4H2PO4 – 7 g; K2HPO4 – 1.5 g; MgSO4 – 0.5 g; CaCl2 – 0.3g; FeSO4.7H2O – 0.184 g; ZnSO4.7H2O – 0.178 g ; MnCl2.4H2O – 0.158 g; carboxymethylcellulose – 5 g; H2O – 1 L).

2.1 TOTAL RNA EXTRACTION

The TRIZOL Reagent (Invitrogen) was used for RNA extraction. The extracted RNA was analyzed by digital photography visualization (KODAK EDAS 290) after 30 minutes of electrophoretic running at 100 V and 90® mA. For electrophoresis, 1% agarose gel was used in 40 mL of TAE (Tris-Acetate-ETDA) buffer.

After RNA extraction, cDNA synthesis was performed using the enzyme Reverse Transcriptase M-MLV (Promega), according to the manufacturer's recommendations. The cDNA obtained was stored at -20 ºC.

2.2 CDNA AMPLIFICATION - (PCR)

cDNA amplification was performed by reverse transcription polymerase chain reaction (RT-PCR). The first cDNA strand was used as the template strand and the primers designed for the coding sequence of the protein under study were drawn from the nucleotide sequence (464) of a Glycosyl Hydrolase 45.

The primer sequences used for the direct primer were: 5'-GGAATCCATATGATGCCTCACCAAAGCCGTACTT-3' with NdeI cleavage site and 5'-GCGCGGATCCTTATCAGGCAGGCCTCTCTCATAAG-3' reverse primer with BamHI cleavage site.

The reaction was performed using the Qiagen® Master Mix kit: Top taq Master mix enzyme; primers (straight and reverse); cDNA as a template and without Water Nuclease. The amplification conditions for the PCR reaction were made according to the manufacturer's instructions.

The product obtained from the PCR reaction was purified using Invitrogen's PureLink® PCR Purification Kit and subsequently sequenced. The same primers used for amplification were also used

as primers for the sequencing reaction. For the sequencing reaction, the recommendations of Myleus Biotechnology (http://myleus.com/) were followed.

2.3 SEQUENCING ANALYSIS

The obtained electropherogram was analyzed using the Geneious Prime program (version 2020.0.4) and the nucleotide sequence recovered with quality was translated through the bioinformatics resource portal ExPASy (https://www.expasy.org/).

The translations of the sequence were submitted to comparative analysis for similarity with complete sequences of Glycosyl Hydrolases of the 45 family (GH45) deposited in the NCBI, using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and aligned through the online tool Clustal (https://www.ebi.ac.uk/Tools/msa/clustalo/). An alignment was also made using the online tool Clustal, between the nucleotide sequence obtained from the sequencing (Seq_464) and the genome sequence (Endo_464).

In addition, the analysis of the conserved domains and regions (active site), their portions and residues involved in the catalysis of the probable enzyme, identification of the signal peptide and confirmation of other relevant information, such as carbohydrate binding sites and other protein domains if present, was performed, for this, we used the following tools: ProSite Software (Sigrist, *et al.* 2012), Smart Database (Letunic, 2018), Pfam (El-Gebali, S., et al. 2019*), and InterPro (*Mitchell, et al. *2019*).

2.4 STRUCTURAL ANALYSIS BY COMPARATIVE MODELING

As the 3D structure of the sequence under study has not yet been solved experimentally, structural alignment tools were used to obtain the 3D model, in order to compare the structural similarity between the sequence obtained from the cDNA sequencing (Seq_464), the genome sequence of *M. perniciosa* provided by Professor Dr. Gonçalo Amarante Guimarães Pereira from the Laboratory of Genomics and Gene Expression of UNICAMP, (Endo_464) and the sequence found in the GenBank database (EEB93859.1).

2.5 MOLD PROTEIN SELECTION

The prediction of the 3D structure of the GH45 protein sequence of the *fungus M. perniciosa* obtained by cDNA sequencing (Seq_464) was performed by comparative modeling, with a search for molds directly on the SWISSMODEL platform server (Biasini, *et al*., 2014). The same method was applied to the nucleotide sequence (Endo_464) of the genome of *Moniliophthora perniciosa*.

2.6 CONSTRUCTION OF THE 3D MODEL

Once the best proteins were selected as a template, the models were built using the SWISS-MODEL platform. After construction, each model was selected according to its QMEAN/Z-SCORE and GMQE. Based on this information, the best model for statistical evaluation was selected.

2.7 MOLD EVALUATION

The best model was submitted to statistical analysis using the MolProbity (http://molprobity.biochem.duke.edu/) platform, evaluated according to the Ramachandran graph. The models were selected and analyzed using the PyMOL tool (v 1.3); The same tool was used for overlapping analysis of the models and their respective molds.

3 RESULTS AND DISCUSSION

3.1 FUNGAL REACTIVATION

The cultivation of the fungus in potato-dextrose-agar (PDA) medium for 8 days at 27ºC was considered excellent for growth and development in terms of mycelial biomass production; this growth is due to the fact that the PDA medium is rich in nutrients and significant amounts of sugar, evaluated as an ideal culture medium for reactivation, indicated for favoring the mycelial growth of fungi, but it is not considered recommended for verification of enzymatic activity.

Culture media rich in carbohydrates play an important role in the mycelial growth of fungi, which use several nutritional sources for their growth and development, among these, sugars are the most necessary carbohydrates, useful for the growth, maintenance and reactivation of the fungus (Souza, *et. al.*, 2015), thus requiring transfer to a medium that has cellulolytic materials as a substrate for the induction of the enzyme of interest.

3.2 CULTIVATION OF FUNGI IN INDUCTIVE CULTURE MEDIA - WY AND CMC

The fungus *M. perniciosa* when incubated in WY inducing medium and in CMC liquid medium, under favorable conditions for its growth (B.O.D at 27º for 8 days), showed lower growth in terms of mycelial diameter, compared to growth when reactivated in PDA medium, however, more expressive mycelial vigor was observed in WY medium.

Culture media containing lignocellulosic substrates are found in most of the studies reported for induction analysis of cellulolytic enzymes. Agro-industrial residues are substrates with a high capacity to induce the production of cellulolytic enzymes by basidiomycete fungi, bearing in mind that generally the materials used for the production of cellulases are of lignocellulosic or pure cellulosic origin (Carneiro, *et al*., 2017; Castro; Perreira Jr, 2010).

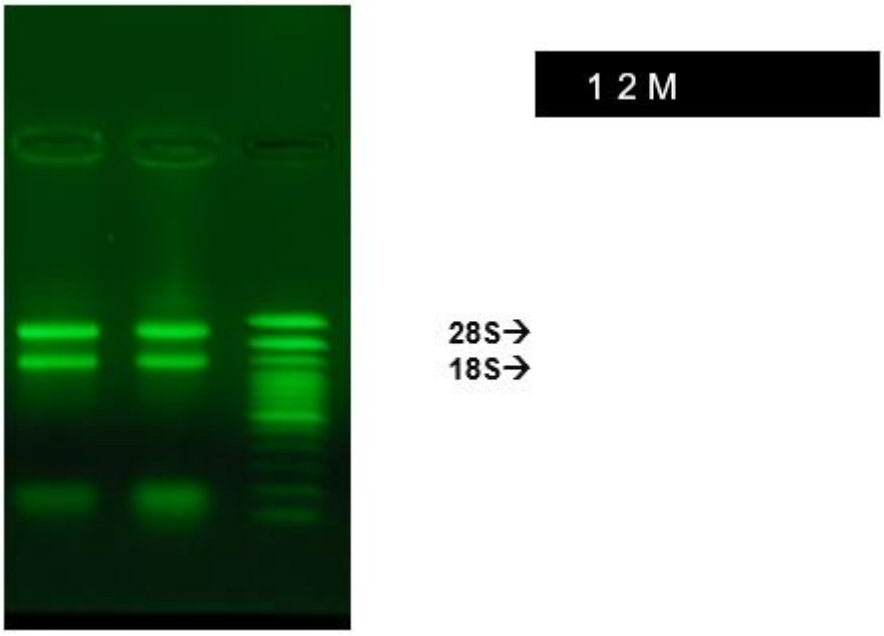
The use of CMC medium for enzymatic induction refers to the fact that this medium contains carboxymethylcellulose as the only carbon source for the growth of the fungus and is described in the literature to induce the cellulolytic complex. Cellulose is a recalcitrant material that is difficult to assimilate nutrients, which can induce the metabolism of fungi to secrete cellulases (Inforsato, 2016).

3.3 TOTAL RNA EXTRACTION

The extraction of the total RNA of the fungus *M. perniciosa* cultured in CMC inducing medium was not considered satisfactory with regard to RNA integrity, it is not possible to verify the 28S and 18S subunits, these ribosomal fractions play an important role in determining the level of degradation of the sample, using RNA integrity assessment (Muelle *et al*., 2004; Muyal, *et al.,* 2009), possibly due to the viscosity of the CMC medium by carboxymethylcellulose, some residues adhered to the mycelium in the removal process interfered in the extraction, therefore, it is suggested an improvement of the extraction method for the fungus cultivated in this culture medium.

The cultivation of the fungus in semisolid medium (WY) was considered satisfactory for cellulase induction analysis, the bands corresponding to the 28S and 18S ribosomal RNA fractions were intact, visualized after electrophoresis in 1% agarose gel (Figure 1) indicating that there was no RNA degradation. The good quality of the extracted RNA is a fundamental step for the continuity of the work. The occurrence of errors in the extraction process is related to many of the problems that can affect subsequent experiments, especially in relation to the efficiency of cDNA amplification.

Figure 1: Elethophoresis agarose gel resulting from the extraction of total RNA from M. perniciosa cultured in WY-inducing medium. 1 and 2: RNA samples obtained with the Trizol (Invitrogen) reagent; M: Invitrogen® 100 molecular weight marker.



3.4 AMPLIFICATION OF M. PERNICIOSA CDNA AND PURIFICATION OF PCR PRODUCT

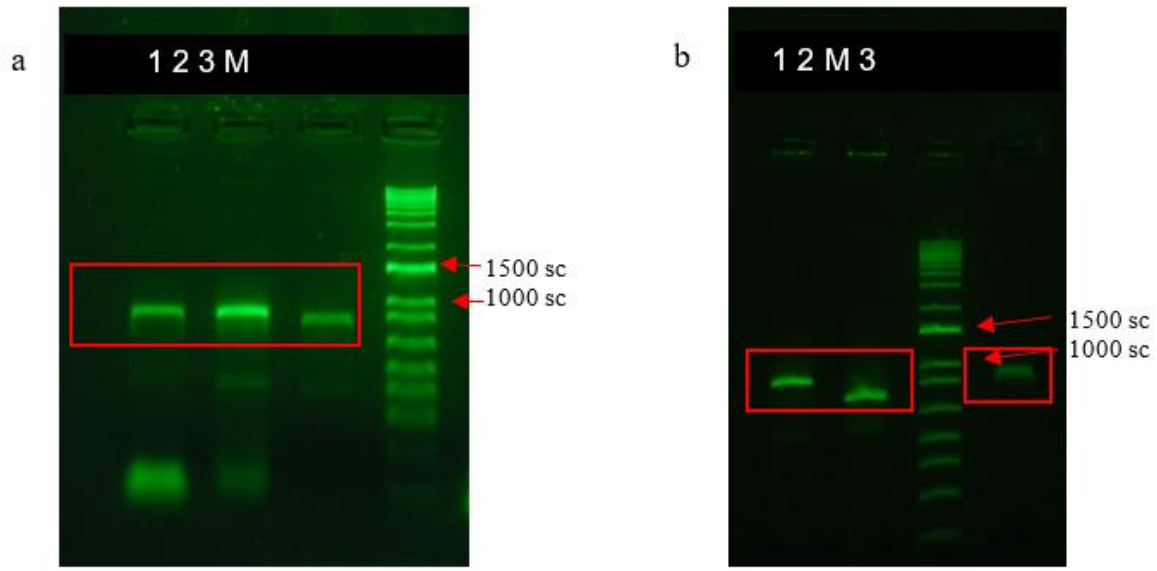
The cDNA amplification showed a positive result when the annealing temperature was optimized to 61 ºC, after testing different temperatures (58 ºC, 59 ºC, 60 ºC and 61 ºC), it was possible to verify the fragment close to the expected size around 800 bp (Figure 2a) after electrophoresis on 1% agarose gel.

These results corroborate data present in the literature, which describe that substrates containing agro-industrial residues have been successfully used to obtain microbial cellulolytic enzymes. Wheat bran is composed of approximately 11% cellulose, 30% hemicellulose, 5% lignin and 17% protein, representing a good inducer for cellulase production, making its use as a substrate advantageous, also the most used carbon source, as it induces a wide variety of cellulolytic enzymes (Rodríguez-Zúñiga, *et al,.* (2011).

The amplified fragment was purified and visualized on 1% agarose gel, where the bands of interest were still present (Figure 2b), and subsequently reverse-sequenced directly.

The result of the sequencing of the purified PCR product was analyzed through the Geneious Prime program (version 2020.0.4), resulting in the recovery of the nucleotide sequence of 87.7% (677 bp), this result refers only to direct sequencing, for reverse sequencing the quality obtained was not enough to extract more information.

Figure 2: Electrophoresis in 1% agarose gel using SYBR™ Green (ThermoFisher Scientific®). a: Amplification result (PCR) of M. perniciosa cDNA. (M): 1kb molecular weight marker plus DNA Ladder Invitrogen (ThermoFisher Scientific®); (1,2 e3): Amplification products using the primer pair Endo_464. (M): 1kb molecular weight marker plus DNA Ladder Invitrogen (ThermoFisher Scientific®); b: PCR product purification



3.5 PROTEIN SEQUENCE ANALYSIS

Bioinformatics programs that perform *in silico* predictions for a given sequence are important to obtain hypothetical information about proteins that have not yet been characterized (Silveira, 2006). In addition, it is widely used to solve issues related to nucleotide and protein sequence analysis, similarity surveys, gene and genome annotation, genome sequencing and assembly, as well as in the areas of comparative genomics, molecular phylogeny, and molecular modeling (Otto, *et al*., 2007).

Thus, an alignment was made through the online tool Clustal (Sievers, 2018), between the nucleotide sequence obtained from sequencing (Seq_464) and the genome sequence (Endo_464), in order to identify the degree of identity between them.

When aligning the nucleotide sequences, the presence of non-corresponding regions (gaps) was observed at the beginning and end of the alignment, justified by the fact that the Seq_464 was smaller in terms of the amount of amino acid residues compared to the Endo_464 sequence.

The presence of gaps, equivalent to 4 missing nucleotides, was observed within the sequence of the Endo_464 genome. It is known that the presence of these *gaps*, depending on the affected region, can have molecular consequences and, in coding regions, these events can cause changes in the reading frame of the protein and make it non-functional.

The translation of Seq_464 and Endo_464 f sequences via the ExPASy bioinformatics resource portal (Gasteiger, *et al.* 2003), resulted in a polypeptide of 205 and 253 amino acid residues, respectively (Chart 1).

A new alignment was made between the amino acid residues of the two sequences, to find identical regions present in their primary structures, presenting 56.74% of identity, it was found that the region corresponding to the gap in the *genome resulted in alterations in the reading frame (Chart 2), since in the region prior to* the gap *the sequences presented 100% identity.*

Table 1: Sequences of the GH-45 protein of M. perniciosa; Seq_464: Result of post-sequencing translation of cDNA with 205 amino acid residues; Endo_464: Sequence obtained from the genome of M. perniciosa with 253 amino acid residues

Seq_464
RQTTGVTTRYWDCCKPSCAWTGKASVSAPVLTCDAGGNTLTDPDVKSGCDGGTAFTCTN
MSPFAVDDSLAYGFAAVNIAGSNEAGWCCACYELTFTDGPVAGKKMVVQATNTGGDLGG
NHFDILIPGGGVGIFTQGCPAQFGSWNGGAQYGGVSSAAECSNLPAAVQEGCRFRFDWMG
GADNPGVTFQEVSCPAQITSVSGCSRQ
Endo_464
MLTKAVLLSLITAAAAQTTGVTTRYWDCCKPSCAWTGKASVSAPVLTCDAGGNTLTDPDV
KSGCDGGTAFTCTNMSPFAVDDSLAYGFAAVNIAGSNEAGWCCACYELTFTDGPVAGKK
MVVQATNTGGDLGGNHFDILIPGGGVSSLKDVPLSSVAGTGVHNMEVYPALRNAATYQQ
PSKKVADSDSTGWAVLTTLVSPSRKYRARRRLQVFRDARDNEGNVFTFIDTPTNIQSSAVE
NTCDPRDGAYEELA

Table 2: Alignment between the amino acid residues of the Endo_464 and Seq_464 sequences, the genome sequence of M. perniciosa and the sequence derived from the sequencing, respectively; Highlight in red for the region corresponding to the gap found in the nucleotide sequence.

```
Endo_464      MLTKAVLLSLITAAAAQTTGVTTRYWDCCKPSCAWTGKASVSAPVLTCDAGGNTLTDPDV  60
Seq_464       ---------------RQTTGVTTRYWDCCKPSCAWTGKASVSAPVLTCDAGGNTLTDPDV  45
                             ********************************************

Endo_464      KSGCDGGTAFTCTNMSPFAVDDSLAYGFAAVNIAGSNEAGWCCACYELTFTDGPVAGKKM  120
Seq_464       KSGCDGGTAFTCTNMSPFAVDDSLAYGFAAVNIAGSNEAGWCCACYELTFTDGPVAGKKM  105
              ************************************************************

Endo_464      VVQATNTGGDLGGNHFDILIPGGGVSSLKDVPLSSVAGTGVHNMEVYPALRNAATYQQPS  180
Seq_464       VVQATNTGGDLGGNHFDILIPGGGVGIFTQGCPAQFGSW--NGGAQYGGVSSAAECSN--  161
              *************************. :.:   :....   :.   * .: .**  .:

Endo_464      KKVADSDSTGWAVLTTLVSPSRKYRARRRLQVFRDARDNEGNVFTFIDTPTNIQSSAVEN  240
Seq_464       -------------LPAAVQEGCRF----RF-DWMGGADNPGVTFQEVSCPAQITSVS---  200
                          * : *. . ::    *:  : .. ** * .*  :. *::* * :

Endo_464      TCDPRDGAYEELA        253
Seq_464       GCSRQ--------        205
               *. :
```

* Identity of all amino acid residues aligned in the indicated column.
: Similarity of amino acid residues aligned in the indicated column
. Low similarity of amino acid residues aligned in the indicated column
- There is no similarity between the amino acid residues aligned in the indicated column

To investigate possible amplification or sequencing errors, the obtained sequence (Seq_464) was submitted to a comparative analysis for similarity with complete sequences of Glycosyl Hydrolases of family 45 (GH45) deposited in GenBank, using BLASTp to investigate possible amplification or sequencing errors. The same was done for genome sequencing (Endo_464).

The alignment of the Seq_464 sequence with the database showed 100% identity with sequence EEB93859.1, noted as a hypothetical partial protein of *Moniliophthora perniciosa* with a coverage percentage of 94%. The result of the alignment of the sequence (Endo_464) with the database also showed 100% identity with the sequence EEB93859.1, but its coverage percentage was 47%, as it was a partial sequence in the database.

The amino acid residues of the three sequences (Endo 464, Seq_464 and EEB93859.1), were then aligned (Chart 3) using the online tool Clustal (Sievers, 2018).

The residues of the two sequences (Endo 464 and Seq_464), aligned separately with sequence EEB93859.1 (Gen_938), showed high identity indices corresponding to 94.65%. On the other hand, the alignment between the sequences (Gen_464 and Endo_464) presented 52.56%, a result similar to that found in the alignment between the Endo 464 and Seq_464 sequences.

Chart 3: Comparative analysis of the identity of amino acid residues between sequences EEB93859.1, Seq_464 and Endo_464, sequence found in the database, cDNA sequencing sequence and sequence obtained from the genome, respectively. Highlight in green for the region corresponding to the signature of the active site; yellow highlight for aspartic acid (D), which participates in the catalytic mechanism; Red highlight for the possible gap region.

```
Endo_464        MLTKAVLLSLITAAAAQTTGVTTRYWDCCKPSCAWTGKASVSAPVLTCDAGGNTLTDPDV 60
EEB93859.1      ------------------------DCCKPSCAWTGKASVSAPVLTCDAGGNTLTDPDV 34
Seq_464         ---------------RQTTGVTTRYWDCCKPSCAWTGKASVSAPVLTCDAGGNTLTDPDV 45
                                        **********************************

Endo_464        KSGCDGGTAFTCTNMSPFAVDDSLAYGFAAVNIAGSNEAGWCCACYELTFTDGPVAGKKM 120
EEB93859.1      KSGCDGGTAFTCTNMSPFAVDDSLAYGFAAVNIAGSNEAGWCCACYELTFTDGPVAGKKM 94
Seq_464         KSGCDGGTAFTCTNMSPFAVDDSLAYGFAAVNIAGSNEAGWCCACYELTFTDGPVAGKKM 105
                ************************************************************

Endo_464        VVQATNTGGDLGGNHFDILIPGGGVSSLKDVPLSSVAGTGVHNMEVYPALRNAATYQQPS 180
EEB93859.1      VVQATNTGGDLGGNHFDILIPGGGVGIFTQGCPAQFGSW--NGGAQYGGVSSAAECSN-- 150
Seq_464         VVQATNTGGDLGGNHFDILIPGGGVGIFTQGCPAQFGSW--NGGAQYGGVSSAAECSN-- 161
                **************************. :.:   :....   :.   * .: .**  .:

Endo_464        KKVADSDSTGWAVLTTLVSPSRKYRARRRLQVFRDARDNEGNVFTFIDTPTNIQSSAVEN 240
EEB93859.1      -------------LPAAVQEGCRF----RF-DWMGGADNPGVTFQEVSCPAQITSVS--- 189
Seq_464         -------------LPAAVQEGCRF----RF-DWMGGADNPGVTFQEVSCPAQITSVS--- 200
                           * : *. . ::    *:  : .. ** * .*  :. *::* * :
```

Information regarding the Seq_464 and Endo 464 sequences on the conserved domains and regions (active site), their portions and residues involved in the probable catalysis of proteins and identification of signaling peptides, were obtained through the following tools: ProSite Software (Sigrist, *et al.* 2012), Smart Database (Letunic, 2018), Pfam (El-Gebali, S., et al. 2019*), and InterPro (*Mitchell, et al. 2019). Both sequences showed the region used as the signature pattern of the active site of these enzymes, characteristic of the protein domain of the 45 glycosyl hydrolases family (Valadares, *et al.*, 2016), located at the N-terminal end, containing 12 amino acid residues (TTRYWDCCKPS), including aspartic acid (D) which acts as a nucleophile in the catalytic mechanism.

For the sequencing of the cDNA (Seq_464), the presence of a sequence corresponding to the signal peptide was not detected, this may be associated with incomplete sequencing in that region, as the percentage of sequence recovered corresponded to 87%.

For the genome sequence (Endo_464), the presence of the signal peptide in the N-terminal region, corresponding to residues 1 to 16, was verified, which was removed for further analysis. The signal peptide corresponds to a region that does not compose the functional structure of proteins, its absence has no implications for sequence analysis.

To assess the structural impact of this gap between genome sequence (Endo_464) and cDNA sequencing (Seq_464), 3D structure prediction was made for probable proteins, although the region corresponding to the gap is not directly involved with the amino acid residues corresponding to the active site region.

3.6 STRUCTURAL ANALYSIS BY COMPARATIVE MODELING AND MOLD SELECTION AND IDENTIFICATION

Understanding the biochemical function of a protein, as well as the characteristics of interactions, it is important to know its three-dimensional structure. However, there are even fewer experimentally resolved protein structures compared to known protein sequences (Bordoli, *et al*., 2009).

Based on this assumption, a search was carried out to find proteins that could serve as a structural model for the prediction of the 3D structure for the sequences under study (Seq_464, Endo_464 and Gen_938).

The global alignment for homology modeling identified as the best structure the PDB protein 5GLY.1.A, which corresponds to a 1,4-β-endoglucanase isolated from the fungus Thielavia terrestris belonging to the Glycosyl hydrolases family 45 (Gao, *et al*., 2017), showing a resolution of 1.58 Å, with 63.38% of sequential identity in relation to the template for the sequence Seq_464, and resolution of 1.6 Å with 63.35% sequential identity for sequence Gen_938.

Sequences that have a sequential identity above 50% are considered significant for the construction of molecular models (Lemuchi, *et al*., 2013; Verli, 2014).

For the Endo_464 sequence, the global alignment for homology modeling showed that the best identity and coverage indices in the alignment with the target sequences was the PDB protein 6MVI.1.A, with 48.80% sequential identity and resolution of 1.9 Å.

Once the structural template was chosen, a multiple alignment was made between the primary structures (Seq_464, Gen_938, Endo_464, PDB 5GLY.1.A and 6MVI.1.A) in order to evaluate the degree of conservation among the residues of the structures, especially those defined as the signature pattern of the protein family under study. It was found that in all primary structures was present the conserved region described as the signature pattern for the GH45 family, both with aspartic acid in identical positions.

From the alignment, it was verified that in the region of the possible gap, only the sequence Endo_464 presented different residues, presenting a serine residue (S), where for the other sequences it corresponds to a glycine (G), resulting in divergence in the alignment of that region.

In addition to the sequences selected for analysis, a profile similar to the Seq_464 and the molds was observed for other fungal cellulases found by the search for similarity in the explosion.

3.7 CONSTRUCTION OF THE 3D MODEL

The technique of constructing rigid body models is based on the evaluation of the conservation of structures between homologous proteins or with a significant degree of identity, providing the construction of a good quality model (Lemuchi, *et al*., 2013).

For the model under study, the GMQE and QMEAN values for the Seq_464 (cDNA sequence) and Gen_938 (Genbank sequence) sequences were 0.83 and -0.58 and 0.85 and -0.64, respectively. For the Endo_464 sequence (genome sequence), these values were 0.65 for GMQE and -4.67 for QMEAN.

The models were also analyzed in relation to stereochemical quality by the Ramachandran plot (Ramachandran, 1968), which considers the percentage of residues within the energy-favorable regions. The models generated for the Seq_464 (cDNA sequence) and Gen_938 (GenBank sequence) sequences showed 94.97% and 95.81% of the residues in highly favorable regions, respectively (Figures 3A and 3B).

A model is considered satisfactory when it has at least 90% of the waste within the energy-friendly regions (Laskowski, *et al.,* 1994; Yang, *et al.,* 2018). These results suggest values considered favorable for a global analysis of the generated model.

For the model generated from the sequence Endo_464, based on the Ramachandran graph, 78.9% of the residuals are in highly favorable regions and 17.5% in allowed regions (Figure 4), this percentage value does not provide good quality for the overall analysis of the model, considering that there was a lower identity between it and the mold, one can attribute low quality to the modeling in this region from this mold.

Figure 3: Ramachandran graph generated by the Prochek platform through the PDBsum online tool, showing the residues that are in highly favored regions (region in red). a: graph generated for model Seq_464; b: graph generated for model Gen_938

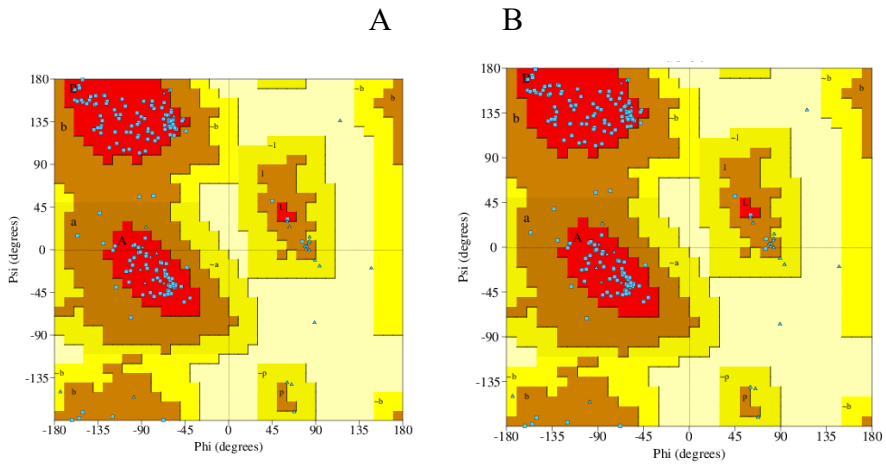
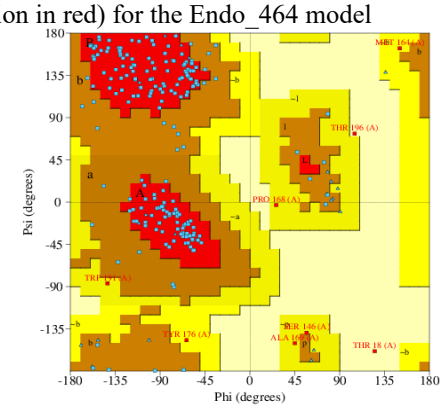


Figure 4: Ramachandran graph generated by the Prochek platform through the PDBsum online tool, showing the residues that are in highly favored regions (region in red) for the Endo_464 model



For Seq_464 (cDNA sequence) and Gen_938 the values were considered satisfactory for the prediction of the 3D structure (Figures 5 and 6). For the Endo_464 sequence (genome sequence) these values indicate a low quality model, the 3D structure is represented in figure 7.

Figures 5 and 6: Model of the hypothetical protein built based on the PDB 5GLY model generated by the PyMol tool depicted in the cartoon. Figure 5: Model Seq_464 (cDNA sequencing); 6: Model Gen_938 (sequence found in GenBank). Orange highlighted region corresponding to the signature of the catalytic site; highlighted in green represents the beginning of the region that showed a difference in the sequence Endo_464.

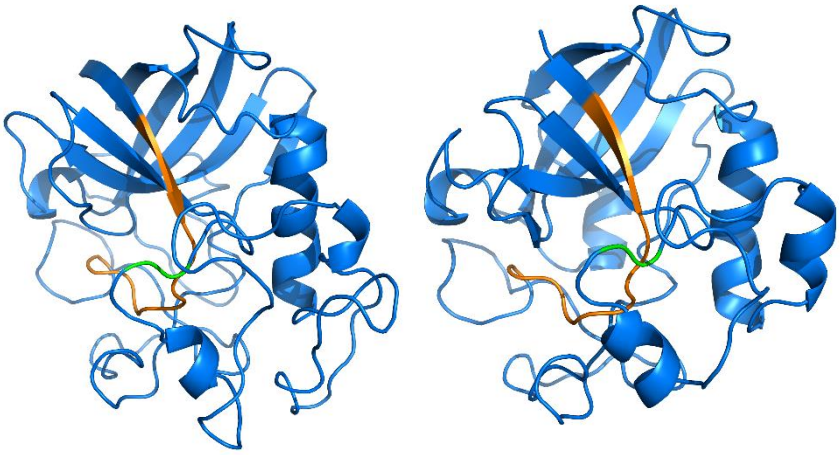
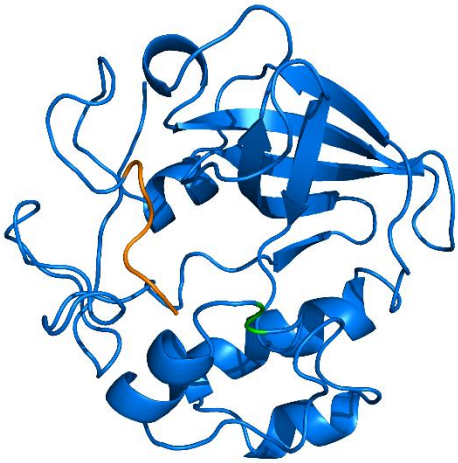


Figure 7: Model of the hypothetical protein Endo_464 (genome sequence), built on the basis of the PDB 6MVI.1.A model generated by the PyMol tool, represented in the highlights of the Orange cartoon for the region of the catalytic site; highlighted in green represents the beginning of the region that showed a difference in the sequence Seq_464



The generated models showed a folding pattern similar to the classic active endoglucanases of the Glisosil Hidrolase 45 family, especially the Seq_464 and EEB93859.1 models, presenting a beta barrel with six β strips, previously found for fungi, such as Thielavia *terrestris* TtCel45A (Gao, et al. 2017) and *Thielavia arenaria* XZ7 TaCel45 (Yang, *et al*., *2018*).

The enzymes of the GH45 family fold into two regions, one is a six-stranded β barrel and the other consists of loop regions. Between the two regions is the substrate binding channel, which is an open slit that runs through the surface of the protein (Gao, *et al*., 2017).

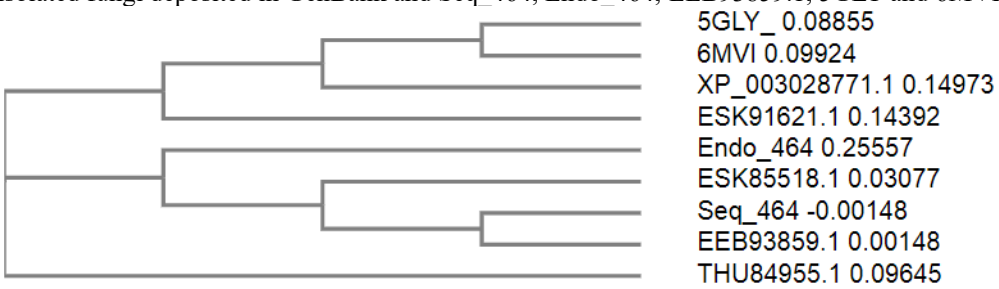
To verify the divergences in the secondary structures of the gap between the available genome sequence (Endo_464), the alignment generated by SWISS-MODEL was analyzed, which assigns the secondary structure to the amino acid residues based on the information from the 3D coordinates, it was found that the Seq_464 (cDNA sequence) and Gen_938 (Genbank sequence) models, They had secondary structures similar to their respective molds, with β and α propellers in identical positions.

In the alignment of Endo_464 (genome sequence) with its template, divergences were observed in the formation of secondary structure elements in the region after the gap, possibly due to the fact that in this region the target protein does not have a sequence corresponding to the template.

To confirm that Seq_464 (cDNA sequence) is functionally more related to other endoglucanases than Endo_464 (genome), a multiple alignment was made between the Seq_464 and Endo_464 sequences, Gen_938 including the 5GLY and 6MVI templates that have a certain three-dimensional structure and other fungal endoglucanases found in the GenBank database (XP_003028771; ESK91621.1; THU84955.1; and ESK85518.1). From the alignment, a phylogenetic tree was generated, where it is observed that all sequences are related, however the sequence Seq_464 presents a closer relationship with one of the groups formed than the Endo_464 (Figure 8).

Based on these analyses, it was possible to verify that only the sequence for Endo_464 presents a residue discrepancy in the region corresponding to the gap, starting at the serine residue (S), while for the others it corresponds to a glycine residue (Gao, *et al*. 2017).

Figure 8: Phylogenetic tree generated by the Clustal tool, resulting from the multiple alignment between endoglucanase sequences of isolated fungi deposited in GenBank and Seq_464, Endo_464, EEB93859.1, 5GLY and 6MVI.



Based on this information, it can be inferred that the cDNA sequence (seq_464) is more related to other sequences deposited in the bank functionally described as endoglucanases than to the genome sequence (Endo_464). It is, therefore, more suitable for investigating possible biotechnological applications.

4 CONCLUSIONS

The detection by cDNA analysis of the expression related to a hypothetical protein of the GH45 family of the fungus Moniliophthora perniciosa involved in the degradation of cellulose using cellulolytic residues as substrate, shows that the fungal culture conditions applied in the present study are favorable for the induction of cellulases and also suggests that the protocols applied were effective to detect the expression of the enzyme of interest. achieving success in RNA extraction and fragment amplification by RT-PCR.

The cDNA sequence obtained through sequencing presents significant conserved regions, which characterizes it as belonging to the Glycosyl Hydrolase family 45. The determination of the 3D structure by comparative modeling presented a high quality model, presenting a GH45 with classic structural characteristics of endoglucanases.

The analysis of the structural difference resulting from the gap in disagreement with the available genome sequence, compared with other active endoglucanase structures, confirms that the sequence obtained by cDNA (seq_464), is more related to functionally active structures, suggesting that the genome sequence (Endo_464) may be a mutated variant or even a sequencing/curation error in the genome.

The Seq_464 sequence obtained from the cDNA can be indicated as a candidate for heterologous expression, purification and characterization of the recombinant protein based on the structural analysis of the sequence of the endoglucanase protein of M. perniciosa analyzed in this work.

THANKS

We thank everyone who contributed to the development of this work.

REFERENCES

Alvim, F. C. (2009) Analysis of the influence of carbon sources on the pathogenicity of Moniliophthora perniciosa pathogenicity in Theobroma cacao. PhD thesis, State University of Campinas, São Paulo.

Barbosa, C. S., et al. (2018) Genome sequence and effectorome of *Moniliophthora perniciosa* and Moniliophthora roreri subpopulations. BMC genomics. 19, 509.

Basso, T. S. (2015) Ultrastructural aspects related to endoplasmic reticulum morphology, differentiation and stress in Moniliophthora perniciosa. PhD thesis, State University of Santa Cruz, Ilhéus, Bahia.

Bhat, M. K. (2000) Cellulases and related enzymes in biotechnology. Biotechnology advances. 18, 355-383.

Biasini, M. E. A. et al. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic acids research, 42, W252-W258.

Bordoli, L. et al. (2009) Protein structure homology modeling using SWISS-MODEL workspace. Nature protocols. 4, 1-13.

Carneiro, R. T. C. O. et al. (2017) *Trametes villosa* lignin peroxidase (TvLiP): Genetic and Molecular Characterization. J. Microbiol. Biotechnol. 27, 179-188.

Castro, A. M; Perreira Jr, N. (2010) Produção, propriedades e aplicações de celulases na hidrólise de resíduos agroindustriais. Química Nova. 33, 181-188.

Castro, A. M; Perreira Jr, N. (2010) Production, properties and applications of cellulases in the hydrolysis of agro-industrial waste. Química Nova. 33, 181-188.

Demain, A. L.; Vaishnav, P. (2009) Production of recombinant proteins by microbes and higher organisms. Biotechnology advances. 27, 297-306.

El-Gebali, S., et al. (2019) The Pfam protein families database in 2019. Nucleic acids research. 47, D427-D432.

Esposito, E.; Azevedo, J.L. (2004). Fungos: uma introdução à biologia, bioquímica e biotecnologia. Caxias do Sul: Editora da Universidade de Caxias do Sul (EDUCS). 510 p. ISBN 85-7061-244-3.

Gao, J., et al. (2017) Characterization and crystal structure of a thermostable glycoside hydrolase family 45 1, 4-β-endoglucanase from *Thielavia terrestris*. Enzyme and microbial technology. 99, 32-37.

Gasteiger, E. et al. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic acids research. 31, 3784-3788.

Guedes, C. E. S. et al. (2014) Enzyme activity of cellulases produced by *Moniliophtora perniciosa*, the causing factor of "witch–broom" in cacao plants. MAGISTRA. 26, 545-553.

Inforsato, F. J.; Porto, A. L. M. (2016) Atividade enzimática de celulases pelo método dns de fungos Isolados de sementes em germinação. Revista Brasileira de Energias Renováveis. 5, 444-465.

Jones, J. D. G; Dangl, J. L. (2006) The plant immune system. nature. 444, 323-329.

Jorgensen, H.; Kristensen, J. B. Felby, C. (2007) Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. Biofuels, Bioproducts and Biorefining. 1, 119-134.

Kamoun, S. (2007) Groovy times: filamentous pathogen effectors revealed. Current opinion in plant biology. 10, 358-365.

Kleman-leyer, K. M., et al. (1996) The cellulases endoglucanase I and cellobiohydrolase II of *Trichoderma reesei* act synergistically to solubilize native cotton cellulose but not to decrease its molecular size. Appl. Environ. Microbiol. 62, 2883-2887.

Laskowski, R. A. MacArthur, M. W. and Thornton, J. M. (1994) Knowledge-based validation of protein structure coordinates derived by X-ray crystallography and NMR spectroscopy. Current Opinion in Structural Biology. 4, 731-737.

Lemuchi, M. O. et al. (2013) Use of Comparative Modeling in Structural Determination of Yersinias Phytase. BBR - Biochemistry and Biotechnology Reports. 2, 25-30.

Letunic, I.; Bork, P. (2018) 20 years of the SMART protein domain annotation resource. Nucleic acids research. 46, D493-D496.

Meinhardt, L. W. et al. (2008) *Moniliophthora perniciosa*, the causal agent of witches' broom disease of cacao: what's new from this old foe? Molecular plant pathology. 9, 577-588.

Mhlongo, S.I. et al. (2015) Lignocellulosic hydrolysate inhibitors selectively inhibit/deactivate cellulase performance. Enzyme and microbial technology. 81, 16-22.

Mitchell, A. L. et al. (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic acids research. 47, D351-D360.

Mueller, O. Lightfoot, S. and Schroeder A. (2004) RNA integrity number (RIN)–standardization of RNA quality control. Agilent application note. 1, 1-8.

Muyal, J. P. et al. (2009) Systematic comparison of RNA extraction techniques from frozen and fresh lung tissues: checkpoint towards gene expression studies. Diagnostic pathology. 4, 9.

Nimlos, M. R. et al. (2012) Binding preferences, surface attachment, diffusivity, and orientation of a family 1 carbohydrate-binding module on cellulose. Journal of Biological Chemistry. 287, 20603-20612.

Otto, T. D., et al. (2007) A plataforma PDTIS de bioinformática: da seqüência à função. Revista Eletrônica de Comunicação, Informação e Inovação em Saúde. 1, 288-296.

Ramachandran, G. N. Sasisekharan, V. (1968) Conformation of polypeptides and proteins. In: Advances in protein chemistry. Academic Press. 23, 283-437..

Rodríguez-Zúñiga, U.F. et al. (2011) Produção de celulases por Aspergillus niger por fermentação em estado sólido. Pesquisa Agropecuária Brasileira. 46, 912-919.

Scarpari, L. M., Meinhardt, L. W., Mazzafera, P., Pomella, A. W. V., Schiavinato, M. A., Cascardo, J. C. M., & Pereira, G. A. G. (2005) Biochemical changes during the development of witches' broom: the most important disease of cocoa in Brazil caused by *Crinipellis perniciosa*. Journal of Experimental Botany. 56, 865-877.

Sievers, F.; Higgins, D. G.(2018) Clustal Omega for making accurate alignments of many protein sequences. Protein Science. 27, 135-145.

Sigrist, et al. (2012) New and continuing developments at PROSITE. Nucleic acids research. 41, D344-D347,

Silveira, T. N. (2006) Expressão da proteína recombinante RasGEF1b: um novo fator de troca de nucleotídeos guanina associado à proteína Ras induzido pelo *Trypanosoma cruzi*. Masters dissertation, Fundação Oswaldo Cruz.

Souza, W. C. O. D. et. al. (2015) Behavior in vitro of *Chalara paradoxa*, causal agent of pineapple black rot in different growth conditions. Revista Brasileira de Fruticultura. 37, 845-851.

Srivastava, N. et al. (2018) Applications of fungal cellulases in biofuel production: advances and limitations. Renewable and Sustainable Energy Reviews. 82, 2379-2386.

Tomme, P. et al. (1988) Studies of the cellulolytic system of *Trichoderma reesei* QM 9414: analysis of domain function in two cellobiohydrolases by limited proteolysis. European journal of biochemistry. 170, 575-581.

Valadares, A. et al. (2016) Exploring glycoside hydrolases and accessory proteins from wood decay fungi to enhance sugarcane bagasse saccharification. Biotechnology for biofuels. 9, 1-12.

Verli, H. (2014) Bioinformática da Biologia à Flexibilidade Molecular. (1.ed) São Paulo, Brazil. pp. 1–291.

Woodward, J. Wiseman A. (1982) Fungal and other β-D-glucosidases, their properties and applications. Enzyme Microb Technol. 4, 73–93.

Yang, H., et al. (2018) Impact of disulfide bonds on the folding and refolding capability of a novel thermostable GH45 cellulase. Applied microbiology and biotechnology. 102, 9183-9192.

Zeeshan, N. et al. (2018) Heterologous expression and enhanced production of β-1, 4-glucanase of Bacillus halodurans C-125 in Escherichia coli. Electronic Journal of Biotechnology. 34, 29-36.