Plant Physiology and Biochemistry 98 (2016) 162–170

Contents lists available at ScienceDirect

Plant Physiology and Biochemistry

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

Research article

Characterization of the legumains encoded by the genome of *Theobroma cacao* L

Juliano Oliveira Santana [a], Laís Freire [a], Aurizangela Oliveira de Sousa [a], Virgínia Lúcia Fontes Soares [a], Karina Peres Gramacho [b], Carlos Priminho Pirovani [a, *]

[a] Biotechnology and Genetics Center, State University of Santa Cruz, 45662-900 Ilhéus, BA, Brazil
[b] Cocoa Research Center, CEPLAC/CEPEC, P.O. Box 7, 45600-970 Itabuna, BA, Brazil

ARTICLE INFO

Article history:
Received 20 July 2015
Received in revised form
29 October 2015
Accepted 16 November 2015
Available online 2 December 2015

Keywords:
Cysteine-proteinase
Cocoa
Modeling
qPCR
Gene expression
Witches' broom

ABSTRACT

Legumains are cysteine proteases related to plant development, protein degradation, programmed cell death, and defense against pathogens. In this study, we have identified and characterized three legumains encoded by *Theobroma cacao* genome through *in silico* analyses, three-dimensional modeling, genetic expression pattern in different tissues and as a response to the inoculation of *Moniliophthora perniciosa* fungus. The three proteins were named *Tc*LEG3, *Tc*LEG6, and *Tc*LEG9. Histidine and cysteine residue which are part of the catalytic site were conserved among the proteins, and they remained parallel in the *loop* region in the 3D modeling. Three-dimensional modeling showed that the propeptide, which is located in the terminal C region of legumains blocks the catalytic cleft. Comparing dendrogram data with the relative expression analysis, indicated that *Tc*LEG3 is related to the seed legumain group, *Tc*LEG6 is related with the group of embryogenesis activities, and protein *Tc*LEG9, with processes regarding the vegetative group. Furthermore, the expression analyses proposes a significant role for the three legumains during the development of *Theobroma cacao* and in its interaction with *M. perniciosa*.

1. Introduction

Legumains are cysteine proteases in the C13 family (EC 3.4.22.34) and are found in plant tissues (Santos-Silva et al., 2012), parasites (Klinkert et al., 1989), and mammals (Chan et al., 2009). These proteins are also called Vacuolar Processing Enzymes (VPEs), due to their ability to recognize and cleave asparagine (Asn) or aspartic acid (Asp) residue in polypeptides (Hara-Nishimura et al., 1995; Müntz and Shutov, 2002).

VPEs are synthesized as inactive prolegumains, and they are transferred to the vacuoles or to the cell wall. Once they reach those compartments under acid pH conditions, prolegumains undergo activation through self-cleavage in the propeptides that are located in the amino and carboxy-terminal. Then they become active legumains in order to play their biological role in plant tissues (Dall and Brandstetter, 2012; Li et al., 2003). In plants, those proteins were initially isolated from mature castor beans (*Ricinus communis*) and from soybean cotyledons (*Glycine max*) (Hara-Nishimura et al., 1991; Scott et al., 1992), and they have conserved the Cys and His catalytic residues which are preceded by a block of four hydrophobic amino acids (Chen et al., 1998).

In plants, VPEs are involved in several physiological processes, and they are classified in three groups, according to the tissues where they are expressed the most. VPEs in from the vegetative group are involved in programmed cell death (PCD) regulation (Hatsugai et al., 2004), senescence (Donnison et al., 2007), and pathogen response (Rojo et al., 2004). VPEs from seeds are related to the processing and mobilizing of reserve proteins during seed germination (Nakaune et al., 2005; Radchuk et al., 2011). VPEs in the embryogenesis group seem to play a role in the formation process of the external tegument of seeds (Nakaune et al., 2005).

In *Arabidopsis thaliana*, three legumain genes were found, and they are referred to as αVPE, βVPE, and γVPE (Kinoshita et al., 1995a, 1995b). Later, on the same plant, a fourth protein was reported through the observation of its expression in young developing seeds, called δVPE (Gruis et al., 2002). Studies regarding the expression of VPE genes showed that αVPE and γVPE are inserted in the vegetative group, showing higher expression in roots, senescent leaves, and in PCD tissues (Kinoshita et al., 1999). βVPE and δVPE were cataloged in seed VPEs; the first of which was mainly

* Corresponding author.
E-mail address: pirovani@uesc.br (C.P. Pirovani).

http://dx.doi.org/10.1016/j.plaphy.2015.11.010

expressed in dry seeds, embryonic axis, and cotyledon, and δVPE was expressed in developing seeds (Hara-Nishimura et al., 2005; Nakaune et al., 2005).

Cocoa tree (*Theobroma cacao* L.) is highly important economically as source of almonds, raw material for chocolate and sweets, motivating several biotechnological studies which led to the complete sequencing of its genome (Argout et al., 2011). Fungal diseases are responsible for significant losses in the production of almonds, especially witches' broom, caused by the basidiomycete fungus *Moniliophthora perniciosa* (Aime and Phillips-Mora, 2005). This fungus is hemibiotrophic and its life cycle is divided into two phases, a biotrophic and a necrotrophic phase, causing histological, physiological, morphological alterations, and death of infected tissues in plants (Ceita et al., 2007; Orchard et al., 1994; Scarpari et al., 2005).

In this study, we have identified and characterized the three VPEs (*Tc*LEG3, *Tc*LEG6 e *Tc*LEG9) that are coded by the *Theobroma cacao* genome through *in silico* analyses, three-dimensional modeling, genetic expression patterns in different tissues, and propose a significant role of VPEs during plant development and in response against the *M. perniciosa* pathogen, which causes witch's broom in the cocoa tree.

2. Materials and methods

2.1. Primary in silico analysis

The primary amino acid sequences that were predicted for the VPEs which are coded by the *Theobroma cacao* genome were obtained from the CocoaGenDB database (http://cocoagendb.cirad.fr/). The three identified VPEs were submitted to different bioinformatics analyses: sequence alignment – ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/); signal peptide prediction – SignalP 4.1 Sever (http://www.cbs.dtu.dk/services/SignalP/) (Bendtsen et al., 2004); tracking of functional domains – Pfam (http://pfam.xfam.org/search/sequence), and analysis on MEROPS (http://merops.sanger.ac.uk/).

The phylogenetic dendrogram was built with the help of MEGA 6 software (Kumar et al., 2004), using Neighbor-Joining method and the confidence of the branching order was verified by making 1000 bootstrap replicates. The dendrogram was built with 31 sequences of homologous proteins that were distributed in eleven plant species, as follows: *Nicotiana tabacum* (tobacco), *Solanum lycopersicum* (tomato), *Oryza sativa* (rice), *Vitis vinifera* (grape), *R. communis* (castor oil plant), *Triticum aestivum* (wheat), *Hordeum vulgare* (barley), *Cucumis sativus* (cucumber), *Sesamum indicum* (sesame), *A. thaliana*, and *Theobroma cacao* (cocoa). The searches for homologous sequences were conducted in NCBI database (National Center for Biotechnology Information - http://www.ncbi.nlm.nih.gov/BLAST), and the sequences identified with their access numbers. One of the cystatins (*Tc*Cys1) of cocoa, which is a competitive inhibitor of cysteine protease, described by Pirovani et al. (2010), was used as an outlier.

2.2. Three-dimensional modeling

Homology modeling was used to obtain the active and inactive three-dimensional structures from the three *T. cacao* VPEs. The searches for template structures as resolved through x-ray crystallography or nuclear magnetic resonance (NMR) were carried out by the PDB database (http://www.rcsb.org/pdb/home/home.do).

Modeller 9.12 software (Eswar et al., 2007) was used in the three-dimensional modeling of the structures. The scripts for processing were generated by the Notepad++ software (Version 6.5.1), with one ".ali" extension file (Dynamics AX Label Index File) and five ".py" extension files (Python). The ".ali" contains information regarding the target protein sequence, whereas the scripts in the ".py" format have modeling commands and three-dimensional structures of template proteins selected in a ".pdb" format. The generated protein structures were refined and validated by softwares (Charmm, Anolea, and Procheck) which analyze possible structural errors in specific regions and stereochemical parameters (Johnson et al., 1994) (Figs. S1–S3).

2.3. Tissues and fruits analyzed

Plants of the CCN51 variety were used to analyze the genetic expression in different tissues (root, tegument, seed, leaf and stalk). They were grown under natural lighting, drip irrigation, and room temperature. The seed cotyledons of the variety "Parazinho" were germinated in a growth chamber at a 27 ± 2 °C temperature, with photoperiod of 16 h of light and 8 h of darkness, with 36 μmol m^{-2} s^{-1} irradiance. Samples were collected at ten different germination stages, as follows: zero hour (quiescent seed), 24 h, 48 h, 72 h, day 4, day 6, day 8, day 12, day 20, and day 30, after the germination. Fruits of the Parazinho variety were collected in four different development stages in order to analyze the expression of VPEs in the seed development phase (Fig. S4). Fruits were measured with the aid of a caliper ruler, and weighed through precision scales. After collection tissues were immediately frozen in liquid nitrogen, lyophilized, and stored at −80 °C, until DNA extraction.

2.4. Germination conditions of infected tissues

Two hundred seeds of the CCN51 variety were planted and kept in the greenhouse under natural light, spray irrigation, and room temperature conditions. After 20 growth days, seedlings were taken to a moist chamber, making up for a total 100 plants infected by *M. perniciosa* and 100 control plants inoculated with 0.3% agar-water. Inoculation was conducted through the spraying of 5.0×10^5 basidiospores/mL of *M. perniciosa* in a 0.3% agar-water suspension (Surujdeo-Maharaj et al., 2003). The seedlings, after the inoculation, were stored at 23 °C for 48 h, under relative humidity >97%, and controlled by an automated nebulization system. Afterwards, seedlings returned to the greenhouse. Biological sample collections (meristems) were conducted on different days, as follows: 0 h (right after the inoculation), 24 h, 48 h, 72 h, 15 days and 45 days after the inoculation. The collected samples were immersed in liquid nitrogen and lyophilized prior to extraction.

2.5. Criteria for primer design

The oligonucleotides that were used for the qPCR (Table S1) were designed with the v.3.0 Primer Express® software (Applied Biosystems®). In order to generate the design of primers for the VPEs (*Tc*LEG3, *Tc*LEG6, and *Tc*LEG9), the following criteria were considered: CG percentage, size between 18 and 25 bp, melting temperature Tm (58–60 °C), and amplicon size (50–150 pb). Actin (ACT), malate dehydrogenase (MDH), and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) genes were selected as endogenous controls, since they presented homogeneous expression in the analyzed tissues (Pinheiro et al., 2011).

The relative expression levels of genes *Tc*Leg3, *Tc*Leg6, and *Tc*Leg9 were calculated according to the delta Ct (ΔCt) method, and analyzed in two biological replicates (different tissues), five biological replicates (developing seed), ten biological replicates (infected tissues). All tissues were analyzed in experimental quintuplicates. After Ct values were obtained, the average value of the Ct of each target gene was normalized with the Ct value from the endogenous control that was established by the NormFinder