



## Phylogeny and molecular dating of the cerato-platanin-encoding genes

Hanying Yu<sup>1</sup> and Lin Li<sup>2</sup>

<sup>1</sup>Department of Petroleum Engineering, Northeast Petroleum University, Daqing, China.

<sup>2</sup>Shenbei New District Department of Human Resources and Social Security, Shenyang, China.

### Abstract

The cerato-platanin family consists of proteins that can induce immune responses, cause necrosis, change chemotaxis and locomotion and may be related to the growth and development of various fungi. In this work, we analyzed the phylogenetic relationships among genes encoding members of the cerato-platanin family and computed the divergence times of the genes and corresponding fungi. The results showed that cerato-platanin-encoding genes could be classified into 10 groups but did not cluster according to fungal classes or their functions. The genes transferred horizontally and showed duplication. Molecular dating and adaptive evolution analyses indicated that the cerato-platanin gene originated with the appearance of saprophytes and that the gene was under positive selection. This finding suggests that cerato-platanin-encoding genes evolved with the development of fungal parasitic characteristics.

**Keywords:** adaptive evolution, cerato-platanin gene, phylogeny, relaxed clock.

Received: October 17, 2013; Accepted: February 20, 2014.

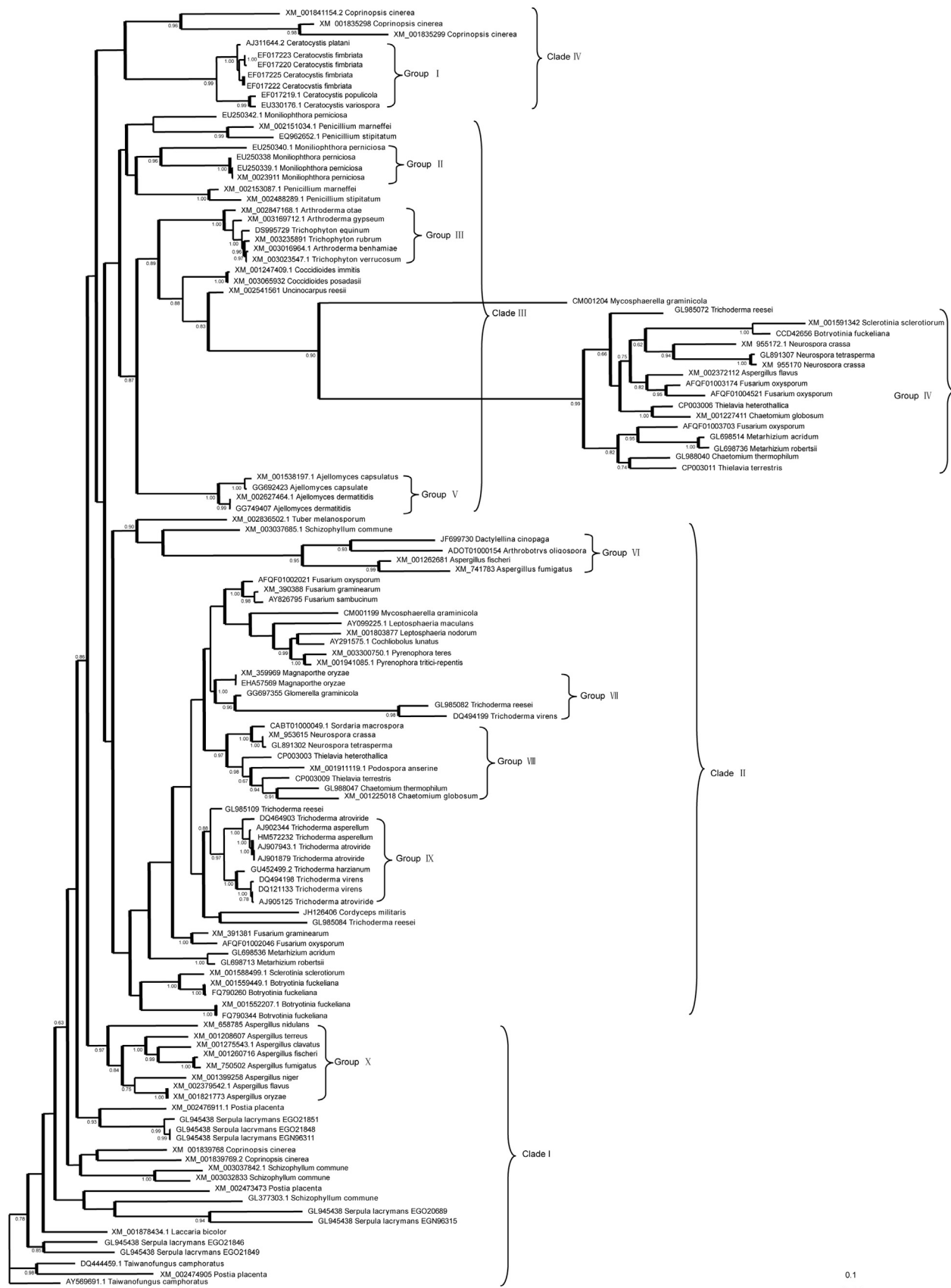
Hydrophobins, also known as small-secreted proteins, are widespread in fungi and have been implicated in the morphology of the cell wall and in fungus-host plant interactions (Veit *et al.*, 2001). The cerato-platanin (CP) protein family, initially described by Pazzagli *et al.* (1999), contains low molecular mass proteins that share high sequence similarity and four conserved cysteine residues involved in the formation of intramolecular disulfide bonds. Cerato-platanins self-assemble to form amyloid-like aggregates but do not form insoluble fibrils. The resulting aggregates interact with hydrophobic surfaces and enhance the hypersensitive response of *Platanus* (Pazzagli *et al.*, 2009). Cerato-platanin from *Ceratocystis fimbriata* f. sp. *platanii* shares several characteristics with hydrophobins, such as high hydrophobicity, an N-terminus sequence that is highly homologous to cerato-ulmin, the sequence Cys-Ser-Asn-22 and an ability to accumulate on the fungal cell surface. However, this cerato-platanin does not have the eight-cysteine pattern found in hydrophobins (Pazzagli *et al.*, 1999). Despite lacking the typical eight-cysteine pattern, proteins in the cerato-platanin family can be grouped with the small cysteine-rich proteins (avirulence gene products, elicitors and hydrophobins) that play an important role in host specificity, recognition and adhesion of some symbiotic fungi (Templeton *et al.*, 1994; Martijn, 2005). Cerato-platanin homologs have been cloned and sequenced from

different species of ascomycete and basidiomycete fungi. Some proteins of this family are produced during infection by pathogenic fungi.

In this paper, we describe a phylogenetic analysis of the genes coding for members of the cerato-platanin family. We calculated the divergence times between the genes and evaluated the possibility of adaptive evolution among genes encoding for infection-, immunity- and biocontrol-related proteins.

The phylogenetic analysis indicated that the sequences clustered into four major clades (Figure 1). Based on the support provided by posterior probabilities, the cerato-platanin family genes may be split into 10 groups (Figure 1). Some sequences could not be confidently assigned to a specific family and were therefore designated “orphan”. The major clades were not supported by the posterior probabilities although almost all the interior nodes and branches were supported, generally because of a loss of many genes. This finding was consistent with the results of Chen *et al.* (2013). The proteins did not cluster according to classes that corresponded to fungi or their functions, which was consistent with the phylogeny of some proteins in the cerato-platanin family (Seidl *et al.*, 2006; Yang *et al.*, 2009; Zapparoli *et al.*, 2009).

The sequences from Arthrodermataceae and Orbiliaceae, which belong to different orders, clustered into one clade, whereas Chaetomiaceae, Clavicipitaceae, Fomitopsidaceae, Onygenaceae, Pleosporaceae, Trichocomaceae and Sordariaceae clustered into different clades. Only the sequences of the genera *Ajellomyces*, *Arthroderma*,



**Figure 1** - Phylogenetic relationship of genes encoding proteins in the cerato-platanin family. The nodes corresponding to adaptive evolution are shown as bold lines. Posterior probabilities > 0.6 are shown. The sequences encoding putative proteins with glycosylation sites are indicated in bold.

*Ceratocystis*, *Coccidioides* and *Pyrenophora* clustered together. The genes of other genera such as *Aspergillus*, *Fusarium*, *Penicillium*, *Trichoderma*, etc., clustered into different clades, indicating that there was gene transfer during evolution of the family. The classification of the species is shown in supplemental Table S1.

The fungi belonging to clade I were from Polyporales, Agaricales and Boletales of Basidiomycota and *Aspergillus* spp. from Eurotiales of Ascomycota, the sequences of which possessed an insertion encoding for (L/V/S)(S/L)(S/P/A/L)(A/S/L/V)(L/A/V). Clade II included the sequences of the fungi *Podospora anserina* and *Schizophyllum commune* from Basidiomycota and those of Capnodiales, Eurotiales, Glomerellales, Hypocreales, Magnaporthales, Orbiliaceae, Pezizales, Pleosporales and Sordariales from Ascomycota. The sequences of *Dactylellina cionopaga*, *Arthrotrichum oligospora*, *Aspergillus fischeri* (XM\_001262681) and *A. fumigatus* (XM\_741783), which were clustered into a clade, possessed an insertion coding for (S/P/A)(N/G)(V/T). The sequences of *S. commune* (XM\_003037685.1) and *Tuber melanosporum* had an insertion coding for VSQGH and PAPAA, respectively. The putative protein sequences of *D. cionopaga*, *A. oligospora*, *A. fumigatus*, *S. commune* and *T. melanosporum* had glycosylation sites. The sequences of *Trichoderma virens* (GL985082) and *T. reesei* (DQ494199), which did not cluster into a clade with other genes coding the cerato-platanin-like proteins of *Trichoderma*, had the insertion coding for LF near the start codon and KLHAYDF before the stop codon and a deletion of the (A/V)(A/Q)A fragment-encoding sequence. Apart from the insertion and deletion, the similarity between GL985082 and DQ494199 and other sequences in the same clade was clear.

The protein encoded by another sequence (GenBank accession no. GL985084) from *Trichoderma*, which did not cluster with the main clade of the genera and was more homologous with that of *Cordyceps militaris*, had an STAS insertion before the stop codon. The sequences of *Fusarium* belonging to clade II clustered into two clades: one of these consisted of AFQF01002021, XM\_390388 and AY826795 while the other consisted of AFQF01002046 and XM\_391381. All the other sequences of cerato-platanin protein-encoding genes were from the genera located in clade III.

The fungi containing the genes in clade III belonged to Agaricales, Eurotiales, Helotiales, Hypocreales and Sordariales. All the sequences of *Penicillium* spp. (XM\_002151034.1, XM\_002153087.1, EQ962652.1 and XM\_002488289.1) had glycosylation sites. Compared to XM\_002151034.1 and EQ962652.1, XM\_002153087.1 and XM\_002488289.1 showed the deletion of an F codon. The sequences with GenBank accession nos. EU250338, EU250339.1 and XM\_0023911 from the basidiomycete *Moniliophthora perniciosa* in clade III had an insertion

coding for SAHTT, while the sequence EU250340.1 had an insertion coding for TFQGV. Compared with the other sequences, the sequence EU250342.1 of *M. perniciosa*, which clustered with that of *Penicillium* spp., had a codon insertion for G.

Clade IV contained homologs from *Coprinopsis cinerea*, *C. fimbriata*, *C. populicola* and *C. variospora*. The sequences of *Ceratocystis* spp. had an insertion coding for (V/I)P, while those from *C. cinerea* (XM\_001835298 and XM\_001835299) had insertions coding for GGGP and GAGGQ, respectively. In contrast to the homologs in clade I, the sequences XM\_001841154.2, XM\_001835298 and XM\_001835299 of *C. cinerea* had insertions coding for PLV, QQR and NNT, respectively. Based on the constructed phylogenetic relationship, the cerato-platanin-like protein-encoding genes of Basidiomycetes clustered with the homologs from Ascomycetes, which had some short insertions. The deletions and insertions accounted for the differences among the sequences and their conservation.

Group IV included the cerato-platanin genes from several genera, a finding indicative of gene transfer. Analysis of the sequences and alignments showed that additional conservative, high molecular mass fragments existed in these sequences. The hydrophobicity of these putative proteins was computed with the program ProScale using window 9 in the Kyte and Doolittle scale. These additional DNA fragments coded for the amino acids that formed continuous hydrophobic and hydrophilic regions (Table S2) and were located at the surface of the proteins or buried within the protein structure (Table S3). These short protein sequences may be involved in mediating fungus-host relationships or protein aggregation.

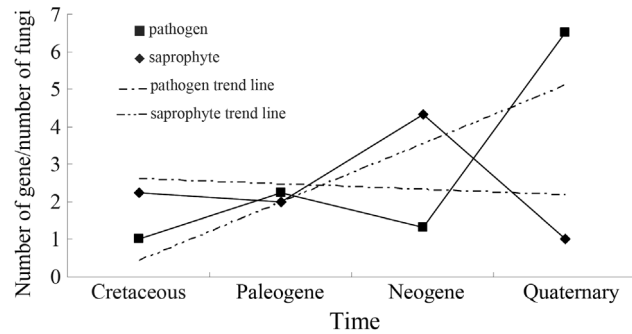
Proteins of the cerato-platanin family self-assemble *in vitro* and this ability plays a role in growth and development and in the parasitic ability of fungi (Pazzagli *et al.*, 2009). The fact that these sequences appeared in different fungal genomes and had higher similarity than other regions implied that the genes might be copies of the cerato-platanin family with specialized functions. Vargas *et al.* (2008) proposed that a subdivision of the cerato-platanin protein family existed and was defined by the presence of a glycosylation site and the ability of the proteins to aggregate. Putative glycosylating proteins were retrieved by searching against the PROSITE database (data not shown). The results indicated that the cerato-platanin-encoding genes that possessed glycosylation sites clustered into limited subclades, but did not form a single clade. In addition to conservative oligosaccharide-binding sites identified in some members of this protein family (de Oliveira *et al.*, 2011), our results indicate that there are also divergent (non-conserved) sites in several of the proteins analyzed.

Several cerato-platanin proteins have functions that include roles in growth and development, parasitism, cell wall morphogenesis, immune response and chemotaxis. Based on the phylogenetic tree obtained here, the se-

quences from the genera *Ceratocystis*, *Trichoderma* and *Leptosphaeria*, as well as those from *A. oligospora* and *M. cionopagum*, grouped into a clade and most of the sequences of *Aspergillus* and *Moniliophthora* clustered together, indicating that these genes most likely possess similar functions in fungi. However, the fungi did not cluster with similar parasites in the phylogenetic tree of cerato-platanin family genes.

So far, no study has provided conclusive evidence for the evolutionary dating of the cerato-platanin protein family, although other protein families have been analyzed (Bryant *et al.*, 2009). Using the coding sequence and the topological restriction of the Bayesian-generated tree, the divergence times of this protein family were calculated. The divergence times of the cerato-platanin protein genes analyzed here are shown in Table S4. The phylogenetic relationship of fungi possessing known sequences of proteins in the cerato-platanin family was analyzed with an internal transcribed spacer (ITS). The results are shown in supplemental Figure S1. ITS sequences are good markers for closely related species. The sequences can also be used for phylogenetic analyses of relationships at the order level (Liggenstoffer *et al.*, 2010), although the molecular dating analysis on a single gene (ITS), the only gene available from all of the selected fungi, was flawed. The phylogenetic relationship presented by ITS indicated that the relation of *Trichoderma* spp. and *Aspergillus* spp. was consistent with that presented by the cerato-platanin family.

Fossils of *Aspergillus* and nematophagous fungi with high similarity to the modern anamorphic genera *Arthrobotrys*, *Dactylaria* and *Monacrosporium* were used to calibrate the calculations of evolutionary time (Schmidt *et al.*, 2008). These fossils represented the earliest origin of the genera and their parasitism. Calculations for the divergence events of the corresponding fungi were calibrated with the same two fossils in the Ascomycetes (Figure S2) and Basidiomycetes (Figure S3) trees. The results of the divergence time analysis indicated that the earliest sequences in clades I, II, III and IV originated 124.08, 107.37, 90.15 and 117.44 mya, respectively. The fossil ages of the pathogen hosts were compared with the divergence times for the fungi and their genes (Table S4). The results of the dating analyses for the cerato-platanin-like genes and fungi indicated that the evolutionary sequence of the fungi was saprophyte, plant pathogen, animal pathogen and mycoparasite (Figure 2). Fungi originated after the host and the genes originated after the fungi had evolved for approximately 25 million years. The appearance of these genes may be related to the presence of saprophytes that developed into parasitic fungi (Figure 2). These findings indicate that functional divergence of the genes occurred after divergence of the fungi. The latter most likely evolved adaptively, with fungal development occurring in parallel with host and parasite characteristics.



**Figure 2** - Distribution of the cerato-platanin-encoding sequences showing the evolution of saprophytic and parasitic fungi in relation to their animal, plant, animal and plant, or fungal hosts.

The divergence times and origination times calculated here for Ascomycetes and Basidiomycetes differed from the fossil record and previously reported times. This divergence most likely reflects the inclusion of parasitism in the fossils and species used for the analysis. Genes with early origination times, *e.g.*, XM\_001275543.1 (21.7786 mya), XM\_001262681 (28.8235 mya) and XM\_002372112 (25.7494 mya), were present in fungi that diverged later, such as *A. clavatus* (15.7816 mya), *A. fischeri* (5.1217 mya), *A. flavus* (4.669 mya), suggesting that they were transferred horizontally to the fungi. However, the sequences of EU250339.1 (0.2782 mya), XM\_0023911 (0.2782 mya) and EU250338 (0.3032 mya) of *M. perniciosa* (73.1294 mya) appeared to be homologous with EU250340.1 (42.1462 mya), implying that the genes arose by gene duplication.

The cerato-platanin-encoding genes had one positive and 106 negative selection sites ( $p < 0.1$ ). The positive site was located in the putative oligosaccharide binding region. The cumulative synonymous substitutions in branches with sequences from saprophytes, pathogens and genes of known functions were fewer than non-synonymous substitutions, indicating that these sequences were under positive selection (Table S5). That the genes demonstrating positive selection originated during the Mesozoic era, in which the oldest fossil of an animal pathogen appeared, suggested that the cerato-platanin family evolved with fungi and that some copies could have evolved toward functions in parasitism and development. However, the positive and negative selection of nodes and branches did not occur at any specific time period or age.

The genes encoding proteins of the cerato-platanin family are important, widespread functional genes in fungi. The putative proteins, some of which did not contain signal peptides, *e.g.*, XM\_0023\_79542.1 and XM\_741783, were monomers, dimers, or self-aggregates in their native states, and occupied different cell locations and varied in pI. Characterization of the sequences showed that they varied in length, in the number of insertions and deletions, and in their divergence times. Until now, there has been no evidence to confirm that these genes formed subgroups con-

taining pseudo-genes. The range of functions attributed to these proteins confirmed the physiological importance of the cerato-platanin family. The phylogenetic analysis of the sequences coding for cerato-platanin family proteins did not completely adhere to the expected functional or fungal class. The oldest cerato-platanin gene diverged later than the fungi and showed adaptive evolution associated with the appearance and development of fungi. Additional functional analyses and characterization of the putative members of the cerato-platanin family should provide insights into the ancestral states of the corresponding genes.

## Acknowledgments

This work was sponsored by the Postdoctoral Science Foundation of China (grant no. 20070410155).

## References

- Bryant MK, Schardl CL, Hesse U and Scott B (2009) Evolution of a subtilisin-like protease gene family in the grass endophytic fungus *Epichloë festucae*. *BMC Evol Biol* 9:168.
- Chen H, Kovalchuk A, Keriö S and Asiegbu FO (2013) Distribution and bioinformatic analysis of the cerato-platanin protein family in *Dikarya*. *Mycologia* 105:1479-1488.
- de Oliveira AL, Gallo M, Pazzagli L, Benedetti CE, Cappugi G, Scala A, Pantera B, Spisni A, Pertinhez TA and Cicero DO (2011) The structure of the elicitor cerato-platanin (CP), the first member of the cp fungal protein family, reveals a double  $\psi\beta$ -barrel fold and carbohydrate binding. *J Biol Chem* 286:17560-17568.
- Liggenstoffer AS, Youssef NH, Couger MB and Elshahed MS (2010) Phylogenetic diversity and community structure of anaerobic gut fungi (phylum Neocallimastigomycota) in ruminant and non-ruminant herbivores. *ISME J* 4:1225-1235.
- Martijn R (2005) Small proteins of plant-pathogenic fungi secreted during host colonization. *FEMS Microbiol Lett* 253:19-27.
- Pazzagli L, Cappugi G, Manao G, Camici G, Santini A and Scala A (1999) Purification of cerato-platanin, a new phytotoxic protein from *Ceratocystis fimbriata* f. sp. platani. *J Biol Chem* 274:24959-24964.
- Pazzagli L, Zoppi C, Carresi L, Tiribilli B, Sbrana F, Schiff S, Pertinhez TA, Scala A and Cappugi G (2009) Characterization of ordered aggregates of cerato-platanin and their involvement in fungus-host interactions. *Biochim Biophys Acta* 1790:1334-1344.
- Seidl V, Marchetti M, Schandl R, Allmaier G and Kubicek CP (2006) Epl1, the major secreted protein of *Hypocrea atroviridis* on glucose, is a member of a strongly conserved protein family comprising plant defense response elicitors. *FEBS J* 273:4346-4359.
- Schmidt AR, Dorfelt H and Perrichot V (2008) *Palaeoanellus dimorphus* gen. et sp. nov. (Deuteromycotina): A cretaceous predatory fungus. *Am J Bot* 95:1328-1334.
- Templeton MD, Rikkerink A and Beaver RE (1994) Small cysteine-rich proteins and recognition in fungal-plant interactions. *Mol Plant-Microb Interact* 7:320-325.
- Vargas WA, Djonovic S, Sukno SA and Kenerley CM (2008) Dimerization controls the activity of fungal elicitors that trigger systemic resistance in plants. *J Biol Chem* 283:19804-19815.
- Veit S, Worle JM, Nurnberger T, Koch W and Seitz HU (2001) A novel protein elicitor (PaNie) from *Pythium aphanidermatum* induces multiple defense responses in carrot, *Arabidopsis*, and tobacco. *Plant Physiol* 127:832-841.
- Yang YY, Zhang HJ, Li GJ, Li W, Wang XE and Song FM (2009) Ectopic expression of MgSM1, a cerato-platanin family protein from *Magnaporthe grisea*, confers broad-spectrum disease resistance in *Arabidopsis*. *Plant Biotechnol J* 7:763-777.
- Zaparoli G, Cabrera OG, Medrano FJ, Tiburcio R, Lacerda G and Pereira GG (2009) Identification of a second family of genes in *M. perniciosa*, the causal agent of witches' broom disease in cacao, encoding necrosis-inducing proteins similar to cerato-platanins. *Mycol Res* 113:61-72.

## Supplementary Material

The following online material is available for this article:

Table S1 - The species examined in this work and their classification.

Table S2 - Hydrophobicity of the putative amino acids from the sequences in group IV.

Table S3 - The secondary structure and surface-interior prediction of the putative amino acids of the sequences in group IV.

Table S4 - The divergence times of the cerato-platanin protein genes, corresponding fungi and the fungal host.

Table S5 - Cumulative substitutions in branches with positive selection.

Figure S1 - Phylogenetic relationships and divergence times of fungi with known genes in the cerato-platanin family.

Figure S2 - Phylogenetic relationship and divergence times of genes in the cerato-platanin family of Ascomycetes.

Figure S3 - Phylogenetic relationship and divergence times of genes in the cerato-platanin family of Basidiomycetes.

This material is available as part of the online article from <http://www.scielo.br/gmb>.

Associate Editor: Célia Maria de Almeida Soares

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.