BMC
Genomics

RESEARCH ARTICLE   **Open Access**

The unique architecture and function of cellulose-interacting proteins in oomycetes revealed by genomic and structural analyses

Mathieu Larroque[1,2†], Roland Barriot[3,4†], Arnaud Bottin[1,2], Annick Barre[1,2,5], Pierre Rougé[1,2,5], Bernard Dumas[1,2] and Elodie Gaulin[1,2*]

Abstract

Background: Oomycetes are fungal-like microorganisms evolutionary distinct from true fungi, belonging to the Stramenopile lineage and comprising major plant pathogens. Both oomycetes and fungi express proteins able to interact with cellulose, a major component of plant and oomycete cell walls, through the presence of carbohydrate-binding module belonging to the family 1 (CBM1). Fungal CBM1-containing proteins were implicated in cellulose degradation whereas in oomycetes, the Cellulose Binding Elicitor Lectin (CBEL), a well-characterized CBM1-protein from *Phytophthora parasitica,* was implicated in cell wall integrity, adhesion to cellulosic substrates and induction of plant immunity.

Results: To extend our knowledge on CBM1-containing proteins in oomycetes, we have conducted a comprehensive analysis on 60 fungi and 7 oomycetes genomes leading to the identification of 518 CBM1-containing proteins. In plant-interacting microorganisms, the larger number of CBM1-protein coding genes is expressed by necrotroph and hemibiotrophic pathogens, whereas a strong reduction of these genes is observed in symbionts and biotrophs. In fungi, more than 70% of CBM1-containing proteins correspond to enzymatic proteins in which CBM1 is associated with a catalytic unit involved in cellulose degradation. In oomycetes more than 90% of proteins are similar to CBEL in which CBM1 is associated with a non-catalytic PAN/Apple domain, known to interact with specific carbohydrates or proteins. Distinct Stramenopile genomes like diatoms and brown algae are devoid of CBM1 coding genes. A CBM1-PAN/Apple association 3D structural modeling was built allowing the identification of amino acid residues interacting with cellulose and suggesting the putative interaction of the PAN/Apple domain with another type of glucan. By Surface Plasmon Resonance experiments, we showed that CBEL binds to glycoproteins through galactose or N-acetyl-galactosamine motifs.

Conclusions: This study provides insight into the evolution and biological roles of CBM1-containing proteins from oomycetes. We show that while CBM1s from fungi and oomycetes are similar, they team up with different protein domains, either in proteins implicated in the degradation of plant cell wall components in the case of fungi or in proteins involved in adhesion to polysaccharidic substrates in the case of oomycetes. This work highlighted the unique role and evolution of CBM1 proteins in oomycete among the Stramenopile lineage.

Keywords: Cellulose, Oomycete, Lectin, Immunity, Plant, Adhesion, Fungi

---

* Correspondence: gaulin@lrsv.ups-tlse.fr
†Equal contributors
[1]Université de Toulouse, UPS, Laboratoire de Recherche en Sciences Végétales, 24 chemin de Borde Rouge, BP42617, Auzeville, Castanet-Tolosan F-31326, France
[2]CNRS, Laboratoire de Recherche en Sciences Végétales, 24 chemin de Borde Rouge, BP42617, Auzeville, Castanet-Tolosan F-31326, France
Full list of author information is available at the end of the article

Background

Carbohydrate-binding modules (CBM) are protein domains that recognize and bind to oligo- and polysaccharide ligands. Within the Carbohydrate-Active Enzyme Database (CAZy; CAZyme database, www.cazy.org, [1]), CBMs are divided into 64 families, based on amino acid sequence similarity, and members of each family display common ligand specificity. Regarding reported specificities, characterized CBMs have been found to bind to crystalline or non-crystalline cellulose, chitin, β-1,3-glucans, β-1,3-1,4-mixed linkage glucans, xylan, mannan, galactan or starch, while others behave like lectins, binding to a variety of cell-surface glycans [2].

As protein domains, one or more CBMs are generally associated with other protein domains, typically glycosyl hydrolase (GH) modules, and can be localized at either the N or C-termini of proteins, although proteins formed exclusively of single CBMs have been described [3,4]. Many CBMs have been biochemically characterized and structural data are now available providing insight into structure/function relations (for review see [2]). Since it is known that many CBMs are able to bind to polysaccharides, it is thought that when these are attached to catalytic domains their presence ensures intimate contact with the substrate and thus potentiated catalysis [5-7]. Moreover, it has been postulated that some CBMs might possess the ability to locally destruct polysaccharide structure (e.g. lower crystallinity in cellulose), thus improving enzyme accessibility [8].

Family 1 CBMs are small modules composed of 32 to 36 amino acids and are known as a "fungal CBM family", because they were first detected in fungal cellulases and are exclusively produced by eukaryotes. The first characterized CBM1 was the cellobiohydrolase I from *Trichoderma reesei* [9]. Since then, numerous CBM1 proteins from various fungi have been reported [10-12]. The role of CBM1 in cellulases has been studied by separating the catalytic domain from its CBM, thus facilitating the study of the activity of the isolated catalytic domain on one hand and, on the other hand, the binding ability of the CBM. Data acquired in this way has indicated that CBM1 binds strongly to crystalline cellulose and that its presence is required for full activity of the enzyme against the insoluble polysaccharide [6,13,14]. A structural study of CBM1 from *T. reesei* cellobiohydrolase I have shown that overall architecture forms a wedge shape that is formed by irregular antiparallel triple-stranded β-sheet, which is stabilized by 2 disulfide bridges [15]. A flat face of the wedge bears three aligned aromatic residues (Y5, Y31 and Y32) that, along with polar residues (Q7 and N29), appear to form the cellulose binding interface [16,17]. This is corroborated by the fact that the removal of any of these residues reduces the ability of the enzyme to degrade crystalline cellulose

[17]. Nevertheless, the role of CBM1 at the molecular level is not fully characterized.

Interestingly, CBM1-containing proteins have also been identified in fungal-like organisms called oomycetes [18]. Like fungi, oomycetes are ubiquitous in marine, freshwater and terrestrial environments [19]. They have similar modes of nutrition and ecological roles to true fungi and form tip-growing branching hyphae. Oomycetes were initially classified within the kingdom of Fungi, but molecular phylogenetic studies have now firmly established the distinct taxonomic positions of true fungi and oomycetes. Oomycetes belong to the Kingdom Stramenipila, which includes diatoms, chromophyte algae, and other heterokont protists [20-22]. Numerous oomycete species are plant pathogens, such as the causal agent of potato blight *Phytophthora infestans* or the sudden oak death pathogen *Phytophthora ramorum*. Features characterizing oomycetes are usually based on biochemical studies focused on *Phytophthora sp.* and particularly, the presence of cellulose rather than chitin in their cell wall. However the presence of either chitin or chitosaccharides was observed in the Saprolegniale oomycetes *Saprolegnia monoica* and *Aphanomyces euteiches*, where these compounds play an essential structural role [23,24].

The first oomycetal CBM1- protein described is the cellulose-binding elicitor lectin (CBEL) from *Phytophthora parasitica*, the causal agent of tobacco black shank disease [25]. This non-enzymatic cell wall glycoprotein harbors two CBM1s associated to two PAN/Apple modules known to interact with polysaccharides or proteins [26,27]. Knockdown *P. parasitica*-CBEL transformants are affected in cell wall polysaccharide deposition and adhesion to cellulosic substrates, including plant cell walls [28]. In addition to structural and adhesive roles, CBEL also induces plant defense responses, such as the production of reactive oxygen species, cytosolic calcium variation, expression of PR-proteins, and necrosis in several plant species [28,29]. The mutation of a recombinant CBEL has revealed that functional CBM1 is required for the full elicitor activity of CBEL [30]. Moreover, it has been suggested that interaction of CBM1 module with the plant cellulose microfibrils is perceived by plant cells as a warning signal [30,31]. Similar results have been obtained with a fungal CBM1 from *T. reesei* suggesting that plants are able to perceive oomycetal as well as well fungal CBM1s [32].

The discovery of CBM1-containing proteins in oomycetes has raised the question of their origin in a lineage distantly related to fungi. It has been recently suggested that some genes encoding proteins involved in the breakdown of plant cell wall components have been acquired by oomycetes from fungi through horizontal gene transfer [33]. However, CBM1-containing proteins were not detected in this analysis.

To better understand the origin, evolution and biological role of CBM1-containing proteins in oomycetes, we have performed data mining of fungal and oomycetal genomes and compared the protein organizations of different CBM1-containing proteins. In this way, we have revealed that oomycete-unique association of CBM1 with PAN/Apple domains.
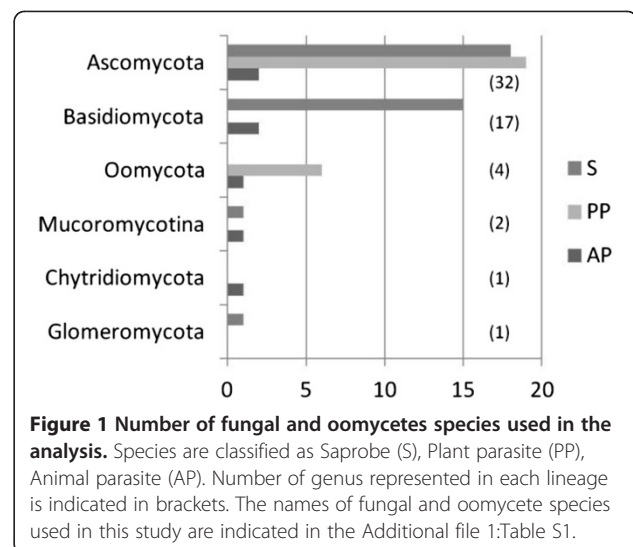
Moreover, using CBEL from *P. parasitica*, which was shown to be *a bona fide* cellulose-binding protein, we propose a model structure of an oomycetal CBM1 and a role for the PAN/Apple domain in binding of the protein to additional carbohydrates. Accordingly, we present experimental evidence for a galactose or N-acetylgalactosamine-specific lectin activity associated with CBEL. Taken together, the results suggest that oomycetal CBM1-containing proteins have an ancient origin in the oomycete lineage and are involved in specific roles including adhesion to self and non-self components rather than in substrate degradation.

Results

Establishment of a comprehensive repertoire of CBM1-containing proteins in fungi and oomycetes

CBM1-containing proteins were collected from the CAZy database [1] and curated for Stramenopiles sequences which were further retrieved by mining predicted proteome on dedicated databases (Additional file 1: Table S1). Seven oomycete genomes were mined for CBM1 (IPR000254) containing sequences. The oomycete species studied are characterized by different lifestyles, namely the obligate biotrophic species *Hyaloperonospora arabidopsidis* and *Albugo laibachii*, four hemibiotrophic *Phytophthora* species (*P. infestans, P. sojae, P. ramorum, P. capsici*) and the necrotroph *Pythium ultimum*. Sequences from various fungal databases, absent in CAZy (i,e *Chaetomium globosum*), were added. Overall a set of 60 fully sequenced fungi representing the major evolutionary lineages within the fungal kingdom were included in the analysis (Figure 1). Sequences were collected to build a representative set spanning the diversity of fungi and oomycetes. In total 518 sequences were selected for further analyses.

The survey of CBM1 revealed the presence of at least one CBM1-protein in 43 of the 67 (77%) completed fungal/oomycete genome sequences examined. The total numbers of CBM1-containing proteins per sequenced fungal/oomycete species classified by life style (saprotroph, necrotroph, hemibiotroph, symbiont and animal pathogen) is shown in Figure 2. *Chaetomium globosum* as well as two other saprobes, the white rot fungus *Phanerochaete chrysosporium* and the dung fungus *Podospora anserina* encode the highest number of predicted CBM1-containing protein coding genes among sequenced fungal genomes. A closer look shows that most fungi that harbor necrotrophic or hemibiotrophic infection strategy (e.g.

Figure 1 Number of fungal and oomycetes species used in the analysis. Species are classified as Saprobe (S), Plant parasite (PP), Animal parasite (AP). Number of genus represented in each lineage is indicated in brackets. The names of fungal and oomycete species used in this study are indicated in the Additional file 1:Table S1.

Verticillium dahliae, Magnaporthe grisea), have the highest number of CBM1-protein coding genes per genome (>10 up to 28 CBM1 copy number). Interestingly, none or a single CBM1-protein gene is detected in biotrophic plant pathogen (e.g. *Ustilago maydis, Puccinia graminis, Melampsora populina*) and symbionts (e.g. *Glomus intraradices, Laccaria bicolor, Tuber melanosporum*). Thus, the expansion of CBM1-encoding genes in phytopathogenic fungi is correlated with hemibiotrophic or necrotrophic lifestyles. However, genes encoding CBM1-containing proteins are not restricted to fungal plant pathogens. Indeed, the same number of genes encoding CBM1-containing proteins is present in *Aspergillus fumigatus* and *A. aculeatus*, the former being a human pathogen and the latter a plant pathogen. Likewise, the opportunist human pathogen *Rhizopus oryzea* also displays genes encoding CBM1-containing proteins while genomes of other animal pathogens, including the amphibian pathogen *Batrachochytrium dendrobatidis*, lack detectable genes encoding cellulose-binding proteins.

Rather like true fungi, necrotrophic and hemibiotrophic oomycetes contain more genes encoding CBM1-containing proteins (up to 12) than biotroph species (*A. laibachi, H. arabidopsidis*). To complete this study, the genomes of *Arabidopsis thaliana* and *Oryza sativa*, and those of the closet phylogenic cousins of oomycetes, the diatoms *Thalassiosira pseudonana, Phaeodactylum tricornutum* and the brown algae *Ectocarpus siliculosus*, which also belong to the heterokont lineage, were screened for CBM1-encoding sequences. However, no CBM1-encoding sequences were detected in these organisms.

Fungal and oomycetal CBM1-containing proteins display distinct domain associations

The domain architecture of CBM1-containing proteins was investigated using the boundaries defined in Pfam/
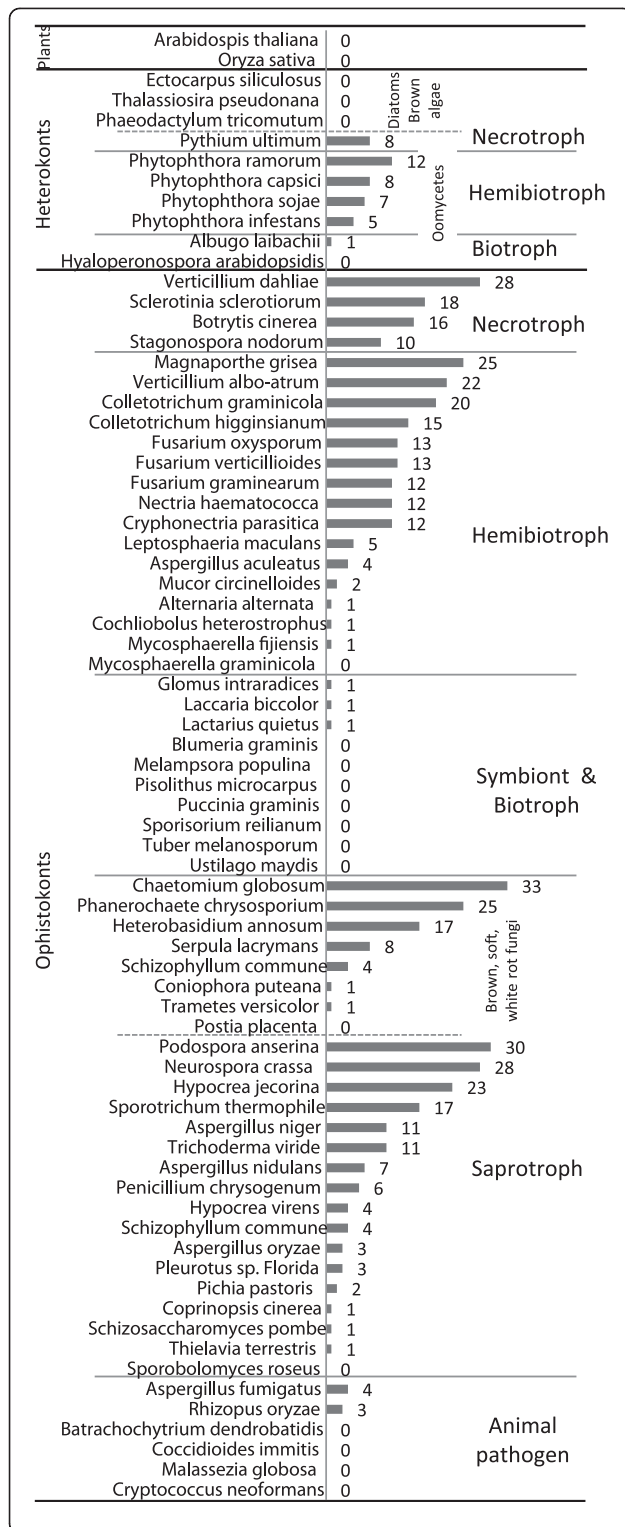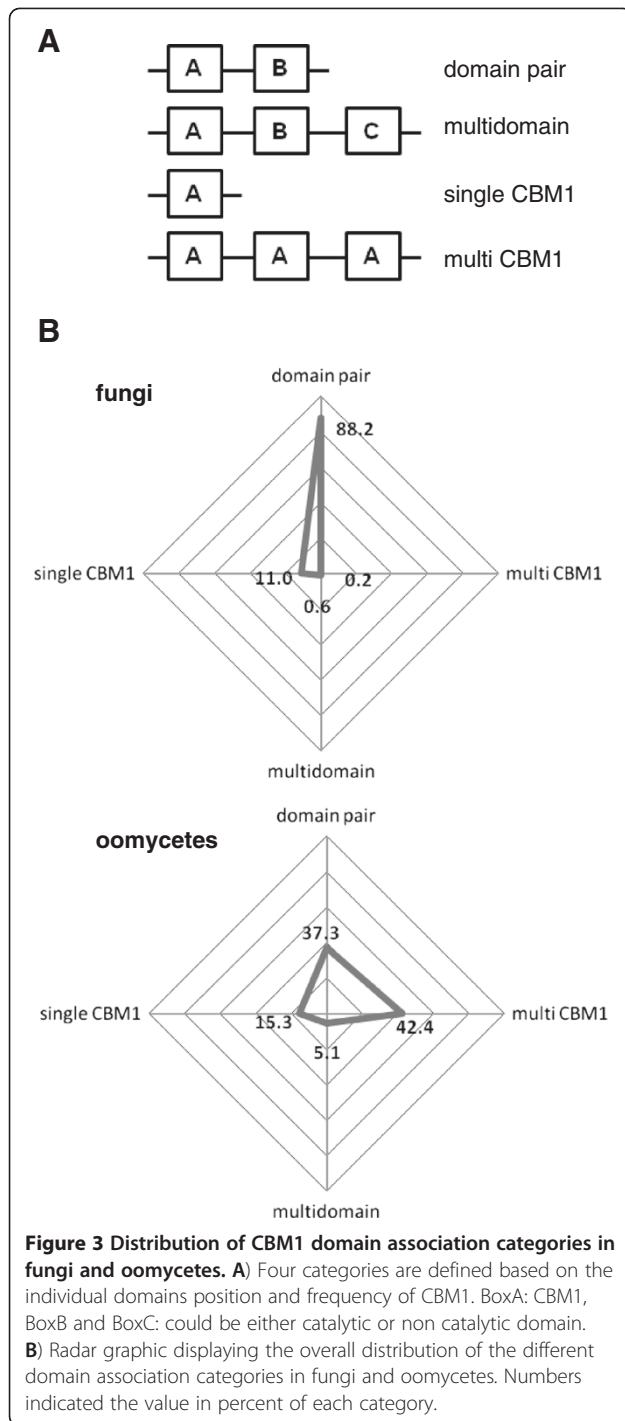
Figure 2 Distribution of CBM1-containing proteins in fungi and oomycetes. Proteomes were screened for the presence of CBM1 domain (IPR00254) thanks to the InterProScan software. The number of putative CBM1-containing proteins per species is reported on the graphic. Microorganisms are classified based on their life style as indicated. Genomes of diatoms (*Phaeodactylum tricomutum, Thalassiosira pseudonana*), brown algae (*Ectocarpus siliculosus*) and plants (*A. thaliana, Oryza sativa*) are included in the analysis.

Smart annotations. Domain architectures were assigned to 4 categories: i) "domain pair" which are sequences that contain a CBM1 and one other domain, ii) "multi-domain", which are sequences that contain one CBM1 and two other domains, iii) "single CBM1", which are sequences that are composed of a single isolated CBM1 and iv) "multi-CBM1", which are sequences that display more than one CBM1 (Figure 3A). Among eukaryotes, fungal proteomes display more than 50% of proteins can be classed as 'single domain' [34]. Regarding CBM1 associations, the 'single domain' category is somewhat under-represented and pertains to roughly ~15% of the fungal and oomycetal CBM1-containing proteins (Figure 3B). In contrast, 88% and 37% of fungal and oomycetal CBM1-containing proteins respectively possess 'domain pair' architecture. Moreover, the 'multi CBM1' architecture is more widespread in oomycetes (42%) than in fungi (0.2%).

In fungi, 89% of the 'domain pair' category are proteins that contain a CBM1 appended to a catalytic domain belonging to the GH superfamily and 11% correspond to a CBM1 appended to non-catalytic modules (chitin-binding module, carbohydrate-binding module family 4, fucose lectin, unknown (DUF) or wide functions (BNR/Asp-box repeats, see Additional file 2: Table S2 for IPR number of corresponding domain). In sharp contrast, in oomycetes 90% of the proteins in the 'domain pair' category are characterized by a CBM1 associated with a non-catalytic domain, corresponding to the PAN/Apple module, which are known to display protein-carbohydrate or protein-protein functions. Using SignalP 3.0 [35], it was found that 91 % of the proteins are predicted to contain a signal peptide, indicating that they are probably secreted. Overall, the large-scale systematic analysis of domain architecture reported here clearly revealed the distinct architecture of fungal and oomycetal CBM1-containing proteins.

Taking our study further, we compared the architecture distribution between the different fully-sequenced organisms. The abundance of domain associations per genome was computed and hierarchical clustering was used to assemble similar domain associations. An overview of the heat map (Figure 4, Additional file 2: Table S2) reveals a distinct distribution of domain combination in fungal and oomycetal proteins. A large proportion of

Figure 3 Distribution of CBM1 domain association categories in fungi and oomycetes. A) Four categories are defined based on the individual domains position and frequency of CBM1. BoxA: CBM1, BoxB and BoxC: could be either catalytic or non catalytic domain. **B**) Radar graphic displaying the overall distribution of the different domain association categories in fungi and oomycetes. Numbers indicated the value in percent of each category.

originally described as endo-1,4-β-D-glucanases displaying weak activity, but recent studies have shown that GH61 members are in fact copper-dependent oxidases, that promote lignocellulosic substrate degradation [36,37]. Within this main group, one can observe a subgroup exclusively constituted by fungal plant pathogens that cause plant wilting symptoms (*Verticillium sp.* and *Fusarium sp.*), which are provoked by the colonization and proliferation of the microorganism in the xylem vessels of the host. In this small cluster, proteins predicted to degrade pectin (polysaccharide lyase families PL1 and PL3 appended to a C-ter CBM1) are also found.

Oomycetes species are clustered in one, unique group presenting two main characteristics. First, catalytic domains classified in the CAZyme database are not detected, except those of the GH5 family. Second, most of CBM1-containing proteins are organized as multiple CBM1 modules, which are either standalone proteins or are appended to PAN/Apple domains. Therefore, to investigate whether the absence of GH domains in oomycete CBM1-containing proteins reflects a general scarcity of glycosyl-hydrolases genes in their genome, the number of GH genes per genome was calculated. As shown in Table 1, oomycete genomes contain a large set of glycosyl-hydrolases genes (>100 gene models/genome), except *A. laibachii* (<50). Thus the rarity in oomycetes of GH.CBM1 associations, coupled to an enrichment of CBM1.PAN/Apple associations, supports the hypothesis that specific functions can be attributed to CBM1-containing proteins from these microorganisms and bears witness to a distinct evolutionary history compared to fungi.

Sequence and structural analysis of oomycetal CBM1-containing proteins

In order to identify residues that are conserved in both fungal and oomycetal CBM1 domains, the CBM1 amino acid sequences of the 518 proteins were extracted and visually compared using WebLogo [38]. As shown in Figure 5, cysteine residues known to participate in the folding of fungal CBM1s [18] appeared at conserved positions in the oomycete sequences, though an additional cysteine residue was also detected (C^{24}). The conserved doublet of glycine residues found in the N-termini of fungal CBM1s is not conserved in oomycetes but the aromatic amino residues ([WY][3], [WY][39], Y^{40}) and the polar residue glutamine (Q^{42}) that are known binding determinants in fungal CBM1s [4,6,17] were frequently identified in the oomycete sequences. Nevertheless, tryptophan and tyrosine residues in these motifs were often replaced by phenylalanine, suggesting greater diversity among the cellulose-binding determinants in oomycetes than in fungi.
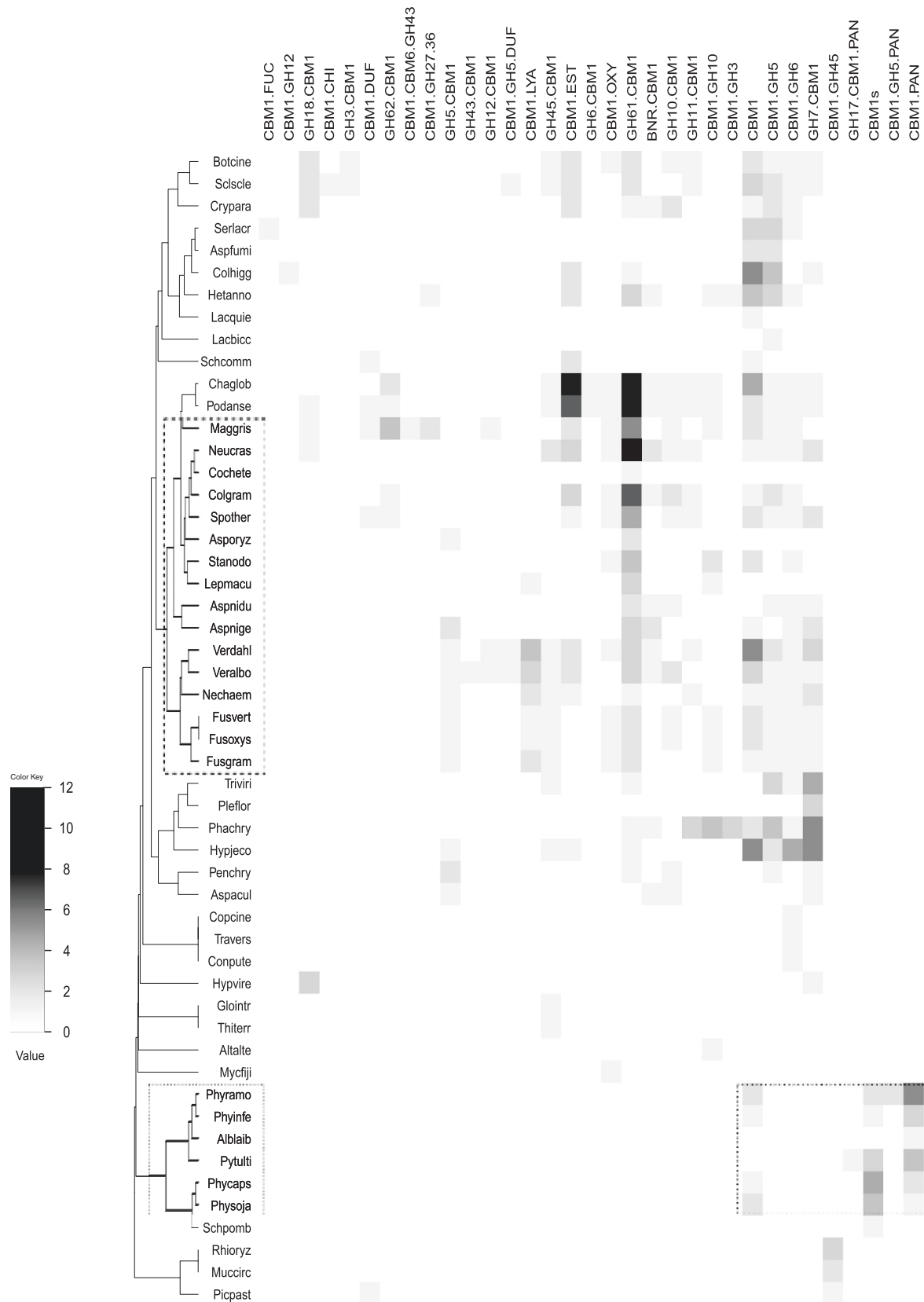
fungal proteins, encompassing 70% of fungal plant pathogens, is characterized by CBM1s associated with catalytic domains involved in cellulose degradation. This includes fungal cellobiohydrolases (GH6.CBM1, GH7.CBM1) and β-glucosidases, exemplified by the CBM1.GH3 association. Interestingly, a large number of fungal CBM1-containing proteins are characterized by the GH61.CBM1 association. The GH61 family was

Figure 4 (See legend on next page.)

(See figure on previous page.)
Figure 4 Hierarchical clustering based on CBM1 domain combinations in fungi and oomycete genomes. Rows correspond to species and columns to architecture of CBM1-containing proteins. Grey intensity on the heat map represents the number of proteins found for a domain combination in a microorganism. The oomycete cluster is indicated as a black line and fungal pathogens as dashed box. Alias used for the species names are indicated in Additional file 1: Table S1.

To detect structural specificities of oomycete domains, a modeling approach was undertaken using the best characterized oomycetal cellulose-interacting protein, CBEL. This protein is composed of a single polypeptide chain of 266 amino acids which starts with a 20 residue long signal peptide and then is built from the two symmetric N- and C-terminal parts, each resulting from the association of a CBM1 and a PAN/Apple domain (CBM1-1:21–54 + PAN1: 55–133; CBM1-2: 158–190 + PAN2: 191–268) separated by a hinge region (134–157) that is rich in proline and threonine residues. The rather high percentage of identity (~37%) and similarity (~60%) between the *T. reseei* cellobiohydrolase I CBM1 and the CBM1s of CBEL (Figure 6A) meant that the CBM1s exhibited very similar HCA plots (data not shown), thus allowing the accurate prediction of the three β-sheet strands in both CBEL_CBM1-1 and CBEL_CBM1-2. In addition the four cysteine residues forming two disulphide bridges in TrCBM1 are strictly conserved in the CBEL_CBM1-1 as shown by the alignment of the domains (Figure 6A). Accordingly, the predicted three-dimensional models of CBEL_CBM1-1 and CBEL_CBM1-2, built from the NMR-coordinates of TrCBM1 (RCSB PDB code 1CBH), exhibited a very similar fold (Figure 6B). They consist of three anti-parallel strands of β-sheet (β1, β2 and β3) interconnected by loops. Two disulphide bonds (C^8-C^{24} and C^{18}-C^{34} in CBM1, C^8-C^{23} and C^{17}-C^{33} in CBM1-2) contribute to the stability of folding, and especially to the stability of the stands (β1 and β3 and loop (loop connecting strands (β2 to β3) which contains the exposed aromatic residues involved in cellulose binding. Both CBM1s contain two exposed aromatic residues F^5 and Y^{31} in CBM1-1 and Y^5 and F^{30} in CBM1-2 that are homologous to Y^5 and Y^{32} of TrCBM1 and are thus suitability positioned to stack onto cellulose chains (Figure 6B and C). An additional Q^{34} residue is also known to participate in the binding of TrCBM1 to cellopentaose and cellohexose [39]. This residue is conserved in CBEL (Q^{33} in CBEL_CBM1-1, Q^{32} in CBEL_CBM1-2), suggesting that it could participate in binding to crystalline cellulose (Figure 6C).

Although the PAN/Apple domains of CBEL share low identity (~20%) and similarity (~36%) with the PAN/Apple domain of the human hepatocyte growth factor (HGF), their HCA plots suggested a very similar structural organization (data not shown). All PAN/Apple domains consist of three and four strands of β-sheets separated by a short α-helical segment. The four cysteine residues linked by two disulphide bridges occurring in

the HGF PAN/Apple domain are conserved in both CBEL_PAN/Apple domains. The three-dimensional models of CBEL_PAN1 and CBEL_PAN2 domains built from the NMR-coordinates of the N-terminal human HGF PAN domain (RCSB PDB code 2HGF) consist of a central five-stranded antiparallel β-sheet associated to a short α-helical segment and to two loops containing a short strand of β-sheet (Figure 7A). In the HGF-PAN/Apple domain, the α-helix is linked to two antiparallel strands of β-sheets by an extended loop to form a hairpin-loop region stabilized by two disulphide bridges. A similar harpin-region stabilized by the disulphide bridges C^8-C^{24} and C^{18}-C^{34} (CBM1-2) or C^8-C^{23} and C^{17}-C^{33} (CBM1-2) also occurs in CBEL. However, the hairpin-loop region is shortened by the deletion of four amino acid residues corresponding to the ^{80}LPFT^{83} motif of the HGF-PAN/Apple domain. The carbohydrate-binding sites of plant lectins usually consist of shallow depressions resulting from the confluence of exposed loops containing hydrophilic and acidic (Asp or Glu) residues that anchor a sugar residue through a network of hydrogen bonds [40]. In addition, an aromatic residue (*i.e.* F or Y) completes the interaction by stacking against the pyranose ring of the sugar. The two exposed loops ^{18}Asn-Val-Asp-**Phe**-Arg-Gly-Asp-Asp^{35} of PAN1 or ^{18}Asp-Lys-Asp-**Tyr**-Arg-Gly-Asn-Asp^{25} of PAN2, and ^{66}Ser-Gly-Thr-Gly-Thr-Arg-Thr^{72} of PAN1 or ^{66}Ser-Ala-Ala-Gly-Thr-Ala-Thr^{72} of PAN2, fulfill these structural requirements and thereby could act as the carbohydrate-binding sites responsible for the hemagglutinating activity of CBEL [18,41].

Six amino acid residues (NYYQCL for CBM1-1/PAN1, SFYQCI for CBM1-2/PAN2) that form the transition between the CBM1s and PAN domains in CBEL correspond to a strand of β-sheet (YQCL and YQCI in CBEL_CBM1-1 and CBEL_CBM1-2 respectively). The first strand of β-sheet (LDLPA in CBEL-PAN1 and IQPPA in CBEL-PAN2) of the CBEL-PAN/Apple domains also contains the ultimate leucine or isoleucine residue of the β-strand of the CBEL-CBM1s. This suggests that a continuous strand of β-sheet (YQCLDLPA and YQCIQPPA) interconnects the CBM1 and PAN domains in the CBEL molecule. Using this information, a tentative three-dimensional model was built for the CBM1-1/ PAN1 N-terminal part of CBEL (Figure 7B). Although speculative, this model readily fulfills geometric constraints (>70% of residues occur in the allowed areas of the Ramachandran plot and >90% in the generously

Table 1 Occurrence of CBM1 in glycosyl-hydrolases in fully sequenced oomycete genomes and selected fungal plant pathogens

		Number of glycosyl-hydrolases predicted genes		
		Total	with CBM1	(%)
Oomycetes	*A. laibachii*	44	0	none
	P. infestans	157	0	none
	P. ramorum	114	2	1.8
	P. sojae	125	0	none
	P. ultimum	85	1	1.2
Fungi	*B. cinerea*	118	11	9.3
	F. oxysporum	368	9	2.4
	M. grisea	232	19	8.2
	V. dahliae	281	14	5.0

Data were collected from *A. laibachii* [69], *Phytophthora sp* [70], *P. ultimum* [71], *B. cinerea* [72], *F. oxysporum, V. dahliae* and *M. grisea* [73].

allowed area) and reasonably accounts for both the associated cellulose-binding and carbohydrate-binding properties previously reported for CBEL [18,41]. A very similar model was obtained for the CBM1-2/PAN2 C-terminal part of CBEL (result not shown).

Surface Plasmon resonance analysis of CBEL lectin activity
In initial studies, the lectin activity of CBEL was demonstrated by a hemagglutination test [18]. To get further insight into CBEL lectin activity and confirm the presence of the carbohydrate binding site inferred from molecular modelling, Surface Plasmon Resonance (SPR) technology was used. Since the detection of molecular interactions using this technology is highly dependent on the

molecular weight of the analytes [42], rather that attempting the direct measurement of the interaction of CBEL with various low molecular weight single sugars, we decided to check whether single sugars could interfere with the binding of large glycoproteins to CBEL. Screening of a collection of various pure glycoproteins allowed to detect the association of CBEL with human lactotransferrin and with a melon hydroxyproline-rich glycoprotein [43,44], both glycoproteins sharing the fact that their glycan moiety contain galactose residues. Subsequent assays showed that the respective protein complexes were strongly destabilized by *N*-acetylgalactosamine (Figure 8). Galactose also showed some activity in this assay, though less pronounced. This data confirms the lectin activity of CBEL and suggests that it is involved in binding to polysaccharides or glycoproteins through the interaction with N-acetylgalactosamine or galactose residues.

Discussion

In this work, a comprehensive set of fungal and oomycetal CBM1-containing proteins was described on the basis of CBM1 detection, and the domain organization of proteins was predicted. CBM1-containing protein genes were found in the vast majority of the analyzed genomes, and their number was clearly related to the lifestyle of the microorganisms. Fungal and oomycetal saprotrophs, necrotrophs and hemibiotrophs express the largest number of CBM1-encoding genes whereas very few, if any, are found in biotrophs. This result is correlated with a dramatic reduction of proteins interacting with plant cell wall components in fungal or oomycete biotrophs [45,46]. This situation could be explained by
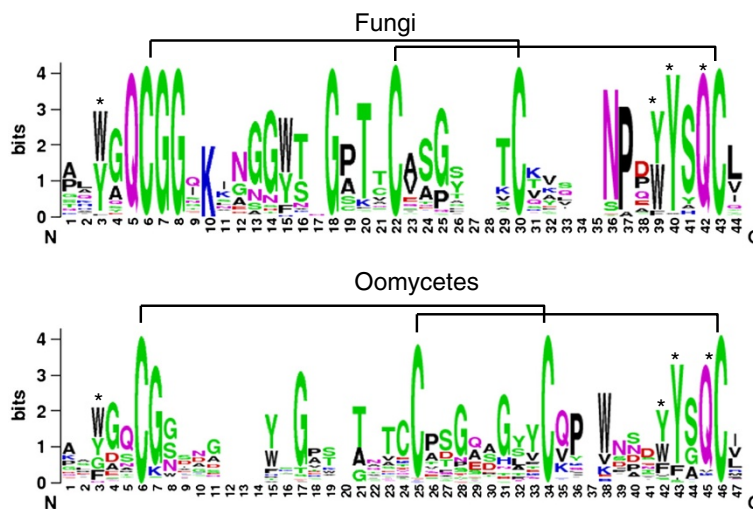

Figure 5 Weblogos built from fungal and oomycetal CBM1 sequences. The height of the residues represents their conservation among the peptide sequences. Cysteine residues engaged in disulphide bonds are materialized as black lines. The glutamine and aromatic residues that are known in fungi to be important for interaction with crystalline cellulose are shown by asterisks. Similarly these residues susceptible in oomycetes to interact with the carbohydrate substrate are depicted with asterisks.
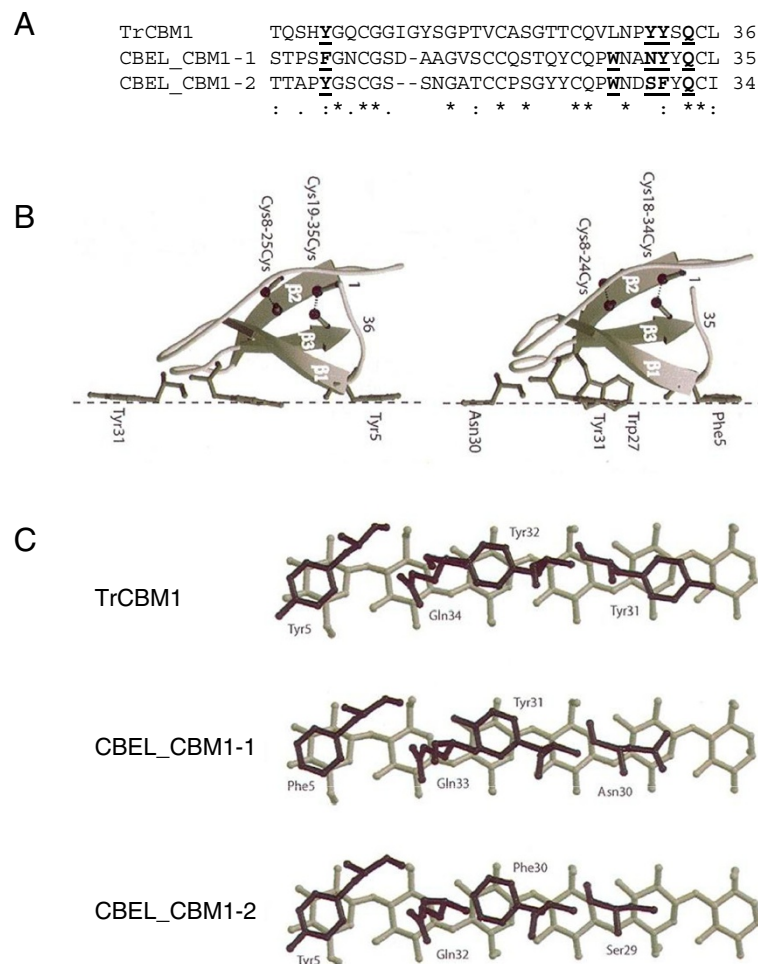
```
A        TrCBM1      TQSHYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL 36
         CBEL_CBM1-1 STPSFGNCGSD-AAGVSCCQSTQYCQPWNANYYQCL 35
         CBEL_CBM1-2 TTAPYGSCGS--SNGATCCPSGYYCQPWNDSFYQCI 34
                     : . :*.**.    * : * *   ** * : **:
```


Figure 6 Modeling of the CBM1s from the CBEL protein. A) Multiple sequence alignment of CBEL_CBM1-1, CBEL_CBM1-2 with the CBM1 of cellobiohydrolase I from *Trichoderma reesei* (TrCBM1, Uniprot/KB accession number: P62694). Amino residues putatively important for binding to crystalline cellulose are in bold and underlined. **B**) Comparison of the lateral faces of the ribbon diagrams of TrCBM1 (left) and CBEL_CBM1-1 (right). The strands of the β-sheet are numbered β1, β2 and β3 from the N- to the C-terminus. The aromatic residues involved in the binding to cellulose are displayed in grey ball-and-sticks. The Asn (replaced by Ser in CBEL_CBM1-2) and Trp residues susceptible to interact with cellulose are similarly displayed in grey ball-and-sticks. The four Cys residues linked to disulphide bonds (dashed lines) are displayed in black ball-and-sticks and numbered. **C**). Precise view of the binding model of CBM1s from TrCBM1 and CBEL. Only the aromatic residues (Phe and Tyr) stacking on the pyranose ring of the glucose units and close hydrophilic residues (Gln, Asn and Ser) susceptible to create hydrogen bonds with the hydroxyls of the glucose units, are displayed in black ball-and-sticks. Only six β1,4-linked glucose units are drawn in grey ball-and-sticks. Cartoons are drawn with Molscript [67], and rendered with Raster3D [68].

the induction of plant defense by plant cell wall-interacting proteins, which could be detrimental to the biotrophic lifestyle. In the case of oomycetes, this interpretation could be extended to non-enzymatic proteins such as CBEL, which has been shown to be a potent elicitor of plant defenses [18,29,30]. While necrotroph and hemibiotroph oomycetes encode 5 to 12 cellulose-interacting proteins, only one was found in the biotroph *Albugo laibachii* and no unambiguous CBM1-encoding genes were detected in the *Hyaloperonospora arabidopsidis* genome.

We further observed that oomycetes express a unique combination of domains corresponding to an association of a CBM1 with a PAN/Apple module. This combination does not appear in other eukaryote including diatoms and brown algae. So far, no enzymatic activity has been associated with CBM1-PAN/Apple proteins, with the best characterized protein of this type being the *P. parasitica* CBEL. It has been shown that CBEL plays a major role in *Phytophthora* cell wall integrity, probably through interactions with the cellulose component of the oomycetal cell wall [28]. Thus, oomycete CBM1-containing proteins play multiple roles that target both the plant and the oomycete cell walls. However, this situation might be unique in the Stramenopile lineage, since no CBM1-containing proteins were detected in
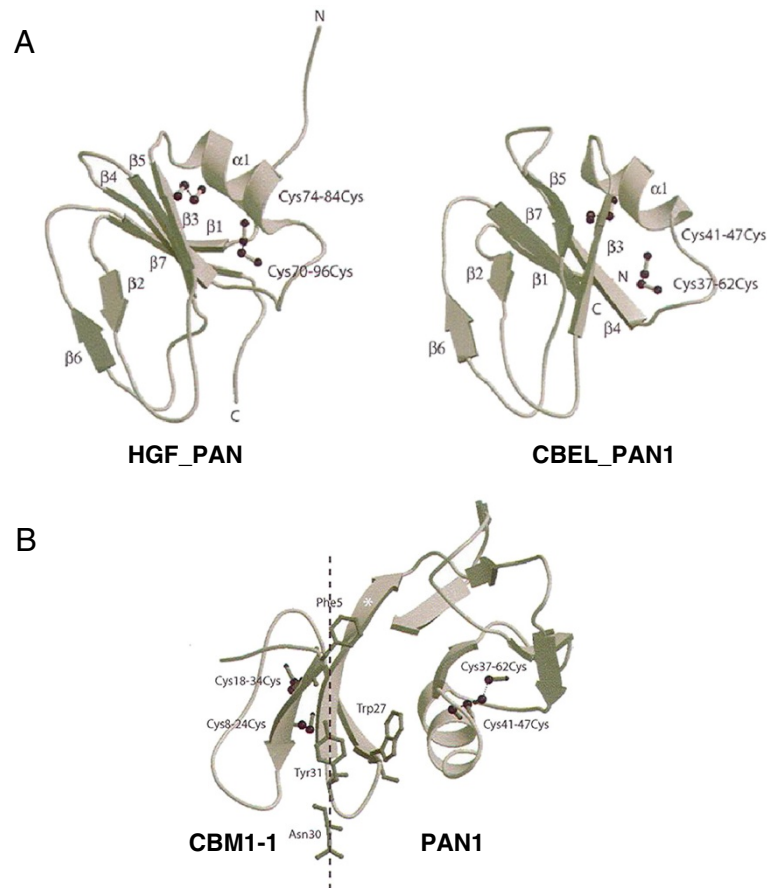
Figure 7 Stereo ribbon representation of PAN and CBM1-PAN1 from CBEL bound to cellulose. A) Comparison of the ribbon diagrams of domains HGF_PAN (left) and CBEL_PAN1-1 (right). The α-helix (α1) and strands of b-sheets (β1-β2) are numbered from the N- to the C- terminus. The cysteine residues linked by disulphide bonds (dashed lines) are displayed in black ball-and-sticks and numbered. **B**) The cysteine residues linked by disulphide bonds (dashed lines) are displayed in black ball-and-sticks and numbered. The extended strand of β-sheet interconnecting the two domains is indicated by an asterisk. The aromatic residues of CBM1 interacting with cellulose are in grey ball-and-sticks and labelled. The figures are drawn using Molscript [67], and Raster3D [68].

other stramenopiles with a cellulosic wall, such as *Ectocarpus silicolosus* [47]. Strikingly, while oomycetes potentially express a large set of glycosyl-hydrolases, only 2 genes were found to be associated with a CBM1, whereas in fungi CBM1 is almost exclusively associated with an enzymatic domain. This result probably bears witness to the distinct evolutionary history of CBM1 and plant cell wall degrading enzymes in fungi and oomycetes. This could be related to the specific functions of CBM1-containing proteins, which in oomycetes are involved in cell wall organization, whereas in fungi, CBM1-containing proteins only target plant components, since fungal cell walls do not contain cellulose.

While CBM1s from oomycetes share similarity with fungal CBM1s, they also display specific features, as revealed by sequence alignment and structural modeling of CBM1s from CBEL. This protein has been shown to bind crystalline cellulose and mutation of its CBM1s has revealed their role in cellulose binding [30]. The two

CBM1s of CBEL, which both exhibit two well-exposed, spatially-aligned aromatic residues (F^5 and Y^{31} of CBM1-1, Y^5 and F^{30} of CBM1-2), readily account for the cellulose-binding properties displayed by both native and recombinant proteins [30]. These aromatic residues are homologous to the aromatic residues (Y^5, Y^{31}, Y^{32}) that are cellulose-binding determinants in the C-terminal CBM1 of the *T. reesei* cellobiohydrolase I [16,17], and are thus suspected to fulfill a similar function. Indeed, mutational analysis of these residues revealed their important role in the binding of CBEL to cellulose and cell wall components [30]. An additional Q^{34} residue (conserved as Q^{31} and Q^{32} in the CBM1-1 and CBM1-2 of CBEL, respectively) also participates in the binding of the TrCBM1 to cellopentaose and cellohexaose [16]. This residue is conserved in CBEL_CBM1-1 (Q^{34}) and CBEL_CBM1-2 (Q^{33}), suggesting that it could also participate in conferring cellulose-binding ability to CBEL. Moreover, residue N^{30} of CBEL_CBM1-1 and S^{29} of

CBEL_CBM1-2 are well exposed and spacially-aligned with the aromatic binding determinants, and thus could also participate in cellulose binding. This is also the case for W^{27} of CBEL_CBM1-1 and W^{26} of CBEL_CBM1-2, which is also located in the vicinity of the aromatic residues.

The occurrence of two PAN/Apple domains in CBEL offers an interesting example of the widespread distribution of a structural motif among proteins of very distinct origins [26,27,48,49]. Although the exact function of the PAN/Apple motif still remains questionable, its involvement in protein-protein and protein-carbohydrate interactions has been postulated from studies performed on plasminogen [50], prekallikrein [51] and on the human hepatocyte growth factor [26]. Interestingly, a PAN/Apple surface protein from the apicomplexan parasite *Toxoplasma gondii* has recently been shown to bind chondroitin sulfate, a sulfated glycosaminoglycan found attached to proteins as part of surface proteoglycans in animal cells [52]. One main constituent of the chondroitin polymer is N-acetylgalactosamine [53]. Accordingly SPR analysis suggests that CBEL is a *N*-acetylgalactosamine or galactose- binding lectin, two sugars that are frequently found as components of glycoprotein glycans. Therefore, it can be hypothesized that this lectin activity is mediated by the PAN/Apple domain.

The tentative three-dimensional model proposed for the molecular organization of the CBM1-PAN/Apple association suggests that there is no steric hindrance between the two domains, thus allowing CBEL to simultaneously display cellulose-binding and lectin activities.

It has been shown that CBEL is localized both in the inner and outer cell wall layers of *P. parasitica* [25], a dual localization which is coherent with its molecular organization and its multiple functions [30]. Interaction of CBEL with a complex glycan, eventually bound to a cell wall polypeptide, could help its proper targeting and molecular docking to endogenous or exogenous cellulose microfibrils, and hence play roles both in cell wall scaffolding and in adhesion to exogenous cellulose. A lectin-based study has shown the presence of galactose and *N*-acetylgalactosamine residues at the cell surface of the oomycete *P. parasitica* [54]. Likewise, a proteomic analysis has shown that mucins, which are known to be highly glycosylated galactose- and *N*-acetylgalactosamine-containing proteins [55], form part of the *P. infestans* cell wall proteome [56]. This strongly suggests the presence of endogenous ligands for the anchoring of CBEL to the oomycete cell wall through its lectin activity. The fact that CBEL is formed from two repeated CBM1-PAN/Apple associations further reinforces the potential of CBEL as a versatile adhesin. In this context, the binding of CBEL to plant HRGPs could be of biological importance and should be further investigated.

The unique symmetric organization of CBEL provokes the question of the molecular evolution of this type of molecule. CBEL probably results from the duplication and fusion events of an ancestral gene, which itself results from the previous fusion of two genes encoding CBM1 and PAN/Apple domains respectively. So far, this particular combination of domains has only been found in oomycetes belonging to the Peronosporale lineage, since in
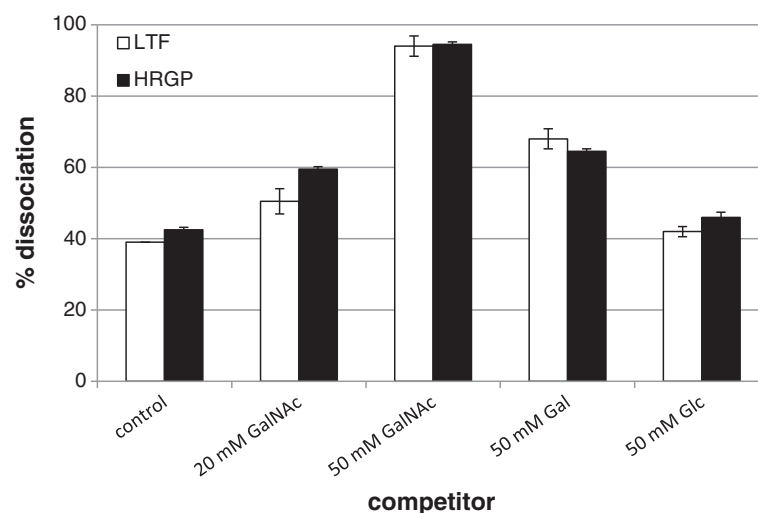

Figure 8 Surface Plasmon Resonance analysis of CBEL-glycoproteins interactions. Human lactotransferin (LTF) or melon hydroxyprolin-rich glycoprotein (HRGP) were passed over the CBEL chip surface at concentrations of 25 or 100 µg/ml respectively. Then, either HBS-Ep buffer alone (control) or monosaccharide solutions (GalNAc, Gal, Glc) were injected during the dissociation phase, 50 or 94 seconds after the end of LTF or HRGP injection respectively. The percentage of protein complex dissociation was measured 385 seconds after the beginning of buffer or monosaccharide injection. Data are the means of the values recorded on the two channels of the Biacore X flow cell ± SD.

distantly related oomycetes, such as *A. euteiches* or *S. parasitica*, CBM1s are associated with another type of interacting domain [57]. Further analysis of oomycete genomes will probably help to clarify the origin of CBM1-containing proteins. The recent identification of CBM1-encoding genes in the basal oomycete *Eurychasma dicksonii* (Gachon and Kim, personal communication), suggests an ancient origin of these proteins, probably related to their specific role in the oomycete cell wall and during interaction with their host.

Conclusions

Using a genome mining approach to analyse fungal and oomycetal genomes, this study aimed at elucidating the origin and function of CBM1-containing proteins from oomycetes. Accordingly, we revealed a unique combination of domains in these organisms in which CBM1s are linked to PAN/Apple domains. This finding in combination with 3D structural modelling and Surface Plasmon Resonance analysis indicate that while CBM1s from fungi and oomycetes are similar, they are actually associated with different protein domains that confer quite different functions: while fungal CBM1s are combined to plant cell wall degradation enzymatic domains, those of oomycetes are associated with domains involved in adhesion to endogenous or exogenous ligands.

Methods

Data collection

Protein sequences were collected from the CAZy database [1], and then curated for oomycetes sequences. Oomycetes proteomes (*Phytophthora infestans, P. sojae, P. ramorum, P. capsici, Albugo laibachii, Hyaloperonospora parasitica*) were downloaded from the Broad Institute or the JGI, and submitted to an InterPro analysis to detect CBM1-containing proteins [58]. The recently sequenced fungal genomes, not yet referenced in the CAZYdatabase, were added after screening their proteome with the InterPro software to identify CBM1-containing proteins (e.g. *Chaetomium globosum*). Only sequences with E-values below 10^{-6} for PFAM or SMART domains corresponding to IPR000254 – Cellulose binding domain, fungal – were kept to minimize false positives. For a list of completely sequenced organisms that were used in this analysis see Additional file 1: Table S1.

Domains architecture determination

All CBM1-containing proteins were submitted to a local InterProScan to identify other domains in the peptide sequences from the SMART and PFAM databases. Domains identified by both the SMART and PFAM databases were merged after checking their compatibility

(location and InterPro identifiers equivalence) by adjusting domain boundaries to the largest domain. Overall, no domain inconsistencies were found between the two databases.

Generation of protein architecture heatmaps

Previously determined protein domain architectures were summarized as follows to obtain more general architecture classes. When multiple CBM1 domains were found in the protein, the architecture is denoted CBM1s. Domains appearance on the sequence was reordered alphabetically except for glycosyl-hydrolase domains *i.e.*, both CBM1-LYA architecture (2 proteins) and LYA-CBM1 architecture (2 proteins) were set to CBM1-LYA whereas CBM1-GH5 architecture and GH5-CBM1 architecture were kept distinct. Multiple occurrences of BNR – bacterial neuramidase repeat – were simplified to only one occurrence *i.e.*, BNR-BNR-BNR-BNR-BNR-CBM1 was mapped to the BNR-CBM1 class. This was motivated by the fact that the number of BNR found in proteins varies from 5 to 8 occurrences and this domain is found in only 5 CBM1-containing proteins. After this generalization, there are 35 distinct classes of protein domain organization. These classes were used to build a contingency table of species and architecture classes. The rows correspond to species profiles: this provides for each species the number of proteins found exhibiting each domain organization. The columns correspond to protein architectures providing the number of proteins found in each species exhibiting such a domain organization. The species and architecture profiles were grouped by hierarchical clustering (average linkage using Euclidean distance) and the resulting classifications were used to draw a heat map in which cell intensities reflect the number of proteins found having a given architecture (column) in a given species (row).

Molecular modeling of CBM1 and PAN domains from CBEL

The HCA (Hydrophobic Cluster Analysis [59]) was performed to delineate the conserved secondary structural features (strand of β-sheet and stretches of a-helix) along the amino acid sequence of CBEL by comparison with the CBM1 of the cellobiohydrolase I from *Trichoderma reesei* [15] and the PAN domain of the human hepatocyte growth factor (HGF,[60]) used as models. Multiple amino acid sequence alignments were carried out with CLUSTAL-W [61]. HCA plots were generated using the program drawhca of L. Canard (http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py?form=HCA). Molecular modeling of the CBM1s and PANs of CBEL was carried out on a Silicon Graphics O2 R10000 workstation using the program InsightII, Homolgy and Discover3 (Accelerys, San Diego CA, USA). The atomic coordinates of the

CBM1 of the cellobiohydrolase I of *T. reesei* (RCSB Protein Data Bank code 1CBH) and the PAN domain of the human HGF (RCSB Protein Data Bank code 2HGF), were used to build the three-dimensional models of the homologous CBEL domains. Steric conflicts were corrected during the model building procedure using the rotamer library [62] and the search algorithm implemented in the Homology program [63] to maintain proper side-chain orientation. An energy minimization of the final models was carried out by 100 cycles of steepest descent using the cvff (consistent valence force field) forcefield of Discover and keeping the cysteine residues involved in disulphide bridges. The program Turbo-Frodo was run to draw the Ramachandran plot and to perform the superposition of the models. PROCHECK [64] was used to assess the geometric quality of the three-dimensional models. Molecular cartoons were drawn with PyMOL (W.L. DeLano, http://pymol.sourceforge.net).

Surface Plasmon resonance analysis

SPR analysis was conducted on a Biacore X device (GE Healthcare, Saclay, France) set at a flow rate of 5μL/min using the CBEL glycoprotein purified from *P. parasitica* mycelium [18]. CBEL fixation on the sensor chip was achieved by hexadecyl-3-methylammonium bromide (CTAB) micelle-mediated immobilization under the following conditions: the sensor chip surface was first equilibrated in a 10 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES) pH 7.4 buffer containing 1 mM CTAB. The sensor chip surface was then washed with 5 μl of 10 mM NaOH, and the carboxymethylated dextran sensor surface was activated by 35 μl of a mixture (1v/1v) of 100 mM *N*-hydroxy-succinimide and 400 mM *N*-ethyl-*N'*-(3-diethylaminopropyl) carbodimide. This activation was followed by injection of 80 μl of a solution of CBEL (12.5 μM) dissolved in 10 mM Hepes pH 7.4, 1 mM CTAB. Remaining ester groups were blocked by 35 μl of 100 mM ethanolamine chlorhydrate pH 8.5, before injection of 1 μl of 10 mM NaOH. Solutions of various pure standard glycoproteins dissolved in 10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% vol/vol surfactant P20 (HBS-Ep buffer, Biacore) were then injected into the flow cell and passed over the CBEL surface during 1 min, and their interaction was followed in real time at different analyte concentrations. The tested glycoproteins were fetuin and asialofetuine (Sigma-Aldrich), human lactotransferrin (a kind gift of Dr H Debray, Université des Sciences et Technologies, Lille, France), and melon Hydroxyprolin-Rich GlycoProtein (HRGP; [43]).

Fetuin is a heavily *N*- and *O*-glycosylated protein where the most abundant *N*-linked glycans are triantennary species, and where both *N*- and *O*-linked glycans are sialylated on galactose terminal residues [65]. Human lactotransferrin glycans are of the sialyl *N*-acetyllactosaminic type and are fucosylated on *N*-acetylglucosamine residues [43]. HRGP is a *O*-glycosylated cell wall protein containing arabinose and oligoarabinoside side chains linked to hydroxyproline residues, and galactose units linked to serine residues [43,66]. Specificity of the CBEL-glycoprotein interactions was checked by measuring the level of protein complex dissociation in presence or absence of various monosaccharides. Complex dissociation was calculated at a fixed time point after injection during 5 minutes, at the beginning of the dissociation phase, of either HBS-Ep buffer or glucose, galactose or *N*-acetylgalactosamine (GalNAc) solutions. Data were analysed using the BIAviewer 3.1 software (Biacore AB, Uppsala, Sweden).

Additional files

Additional file 1: Table S1. List of fungal and oomycete species used in the current study. The phylogenetic classification of the species and the source of the data are indicated in the table.

Additional file 2: Table S2. Alias used for domain organization of CBM1-containing proteins. The IPR number of individual domain appended to CBM1 is indicated, and the corresponding alias used to perform the heat-map of the distribution of architectures among fully sequenced fungi and oomycetes.

Abbreviations
CBEL: Cellulose-binding elicitor lectin; CBM1: Carbohydrate-binding module, family 1.

Competing interests
The authors are not aware of any financial, affiliations, funding that might be perceived as affecting the objectivity of this manuscript.

Authors' contributions
ML, RB, EG performed the computational analysis of the data. RB, BD, EG analyzed the computational results. ML and RB contributed equally to this work. AB conceived and performed the SPR analysis. A. Barre and PR carried out the 3D modeling of the CBEL protein. EG conceived the study and coordinated the manuscript. EG drafted and submitted the manuscript. All authors read and approved the final manuscript.

Author details
[1]Université de Toulouse, UPS, Laboratoire de Recherche en Sciences Végétales, 24 chemin de Borde Rouge, BP42617, Auzeville, Castanet-Tolosan F-31326, France. [2]CNRS, Laboratoire de Recherche en Sciences Végétales, 24 chemin de Borde Rouge, BP42617, Auzeville, Castanet-Tolosan F-31326, France. [3]Université de Toulouse, UPS, Laboratoire de Microbiologie et

Génétique Moléculaire, Toulouse F-31000, France. [4]Centre National de la Recherche Scientifique; LMGM, Toulouse F-31000, France. [5]Present address: Université de Toulouse, UPS, Laboratoire PHARMA-DEV IRD UMR 152, 35 Chemin des Maraîchers, Toulouse 31400, France.

References

1. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: The carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* 2009, **37**(Database issue):D233–D238.
2. Guillén D, Sanchez S, Rodriguez-Sanoja R: Carbohydrate-binding domains: multiplicity of biological roles. *Applied Microbiol Biotech* 2010, **85**:1241–1249.
3. Shoseyov O, Shani Z, Levy I: Carbohydrate binding modules: biochemical properties and novel applications. *Microbiol Mol Biol Rev* 2006, **70**:283–295.
4. Hashimoto H: Recent structural studies of carbohydrate-binding modules. *Cell Mol Life Sci* 2006, **63**(24):2954–2967.
5. Bolam DN, Ciruela A, McQueen-Mason S, Simpson P, Williamson MP, Rixon JE, Boraston A, Hazlewood GP, Gilbert HJ: *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem J* 1998, **331**(Pt 3):775–781.
6. Carrard G, Koivula A, Soderlund H, Beguin P: Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. *Proc Natl Acad Sci USA* 2000, **97**(19):10342–10347.
7. Hervé C, Rogowski A, Blake AW, Marcus SE, Gilbert HJ, Knox JP: Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proc Natl Acad Sci USA* 2010, **107**(34):15293–15298.
8. Din N, Gilkes N, Tekant B, Miller R, Warren R, Kilburn D: Non-hydrophobic disruption of celluose fibers by the binding domain of bacterial cellulase. *Bio/Technology* 1991, **9**:1096–1099.
9. Bhikhabhai R, Johansson G, Petterson G: Cellobiohydrolase from *Trichoderma reesei*. Internal homology and prediction of secondary structure. *Int J Pep Prot Res* 1985, **25**:368–374.
10. Couturier M, Haon M, Coutinho PM, Henrissat B, Lesage-Meessen L, Berrin JG: *Podospora anserina* hemicellulases potentiate the *Trichoderma reesei* secretome for saccharification of lignocellulosic biomass. *Appl Environ Microbiol* 2011, **77**(1):237–246.
11. Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, Aerts A, Asiegbu FO, Baker SE, Barry K, Bendiksby M, *et al*: The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. *Science* 2011, **333**(6043):762–765.
12. Pham TA, Berrin JG, Record E, To KA, Sigoillot JC: Hydrolysis of softwood by *Aspergillus* mannanase: role of a carbohydrate-binding module. *J Biotechnol* 2010, **148**(4):163–170.
13. Carrard G, Linder M: Widely different off rates of two closely related cellulose-binding domains from *Trichoderma reesei*. *Eur J Biochem* 1999, **262**(3):637–643.
14. Gilkes NR, Jervis E, Henrissat B, Tekant B, Miller RC Jr, Warren RA, Kilburn DG: The adsorption of a bacterial cellulase and its two isolated domains to crystalline cellulose. *J Biol Chem* 1992, **267**(10):6743–6749.
15. Kraulis PJ, Clore G, Nilges T, Jones G, Petterson J, Knowles A, Gronenborn P: Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*: a study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *J Biochem* 1989, **28**:7241–7257.
16. Beckham GT, Matthews JF, Bomble YJ, Bu L, Adney WS, Himmel ME, Nimlos MR, Crowley MF: Identification of amino acids responsible for processivity in a Family 1 carbohydrate-binding module from a fungal cellulase. *J Phys Chem B* 2010, **114**(3):1447–1453.
17. Linder M, Mattinen ML, Kontteli M, Lindeberg G, Stahlberg J, Drakenberg T, Reinikainen T, Pettersson G, Annila A: Identification of functionally important amino acids in the cellulose- binding domain of *Trichoderma reesei* cellobiohydrolase I. *Protein Sci* 1995, **4**(6):1056–1064.
18. Mateos FV, Rickauer M, Esquerré-Tugayé MT: Cloning and characterization of a cDNA encoding an elicitor of *Phytophthora parasitica* var. *nicotianae* that shows cellulose-binding and lectin-like activities. *Mol Plant Microbe Interact* 1997, **10**(9):1045–1053.
19. Dick M: *Straminipilous fungi*. Kluwer: Academic Publisher; 2001.
20. Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF: A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 2000, **290**(5493):972–977.
21. Beakes GW, Glockling SL, Sekimoto S: The evolutionary phylogeny of the oomycete "fungi". *Protoplasma* 2012, **249**:3–19.
22. Cavalier-Smith T: A revised six-kingdom system of life. *Biol Rev Camb Philos Soc* 1998, **73**(3):203–266.
23. Badreddine I, Lafitte C, Heux L, Skandalis N, Spanou Z, Martinez Y, Esquerré-Tugayé M, Bulone V, Dumas B, Bottin A: Cell wall chitosaccharides are essential components and exposed patterns of the phytopathogenic oomycete *Aphanomyces euteiches*. *Eukaryot Cell* 2008, **7**(11):1980–1993.
24. Guerriero G, Avino M, Zhou Q, Fugelstad J, Clergeot PH, Bulone V: Chitin synthases from *Saprolegnia* are involved in tip growth and represent a potential target for anti-oomycete drugs. *PLoS Pathog* 2010, **6**(8):e1001070.
25. Séjalon-Delmas N, Mateos FV, Bottin A, Rickauer M, Dargent R, Esquerré-Tugayé MT: Purification, elicitor activity, and cell wall localization of a glycoprotein from *Phytophthora parasitica* var. *nicotianae*, a fungal pathogen of tobacco. *Phytopathology* 1997, **87**(9):899–909.
26. Tordai H, Banyai L, Patthy L: The PAN module: the N-terminal domains of plasminogen and hepatocyte growth factor are homologous with the apple domains of the prekallikrein family and with a novel domain found in numerous nematode proteins. *FEBS Lett* 1999, **461**(1–2):63–67.
27. Zhou H, Casas-Finet JR, Heath Coats R, Kaufman JD, Stahl SJ, Wingfield PT, Rubin JS, Bottaro DP, Byrd RA: Identification and dynamics of a heparin-binding site in hepatocyte growth factor. *Biochemistry* 1999, **38**(45):14793–14802.
28. Gaulin E, Jauneau A, Villalba F, Rickauer M, Esquerre-Tugaye M, Bottin A: The CBEL glycoprotein of *Phytophthora parasitica var. nicotianae* is involved in cell wall deposition and adhesion to cellulosic substrates. *J Cell Sci* 2002, **115**(23):4565–4575.
29. Khatib M, Lafitte C, Esquerré-Tugayé MT, Bottin A, Rickauer M: The CBEL elicitor of *Phytophthora parasitica* var. *nicotianae* activates defence in *Arabidopsis thaliana* via three different signalling pathways. *New Phytol* 2004, **162**:501–510.
30. Gaulin E, Drame N, Lafitte C, Torto-Alalibo T, Martinez Y, Ameline-Torregrosa C, Khatib M, Mazarguil H, Villalba-Mateos F, Kamoun S, *et al*: Cellulose binding domains of a *Phytophthora* cell wall protein are novel pathogen-associated molecular patterns. *Plant Cell* 2006, **18**(7):1766–1777.
31. Dumas B, Bottin A, Gaulin E, Esquerre-Tugaye MT: Cellulose-binding domains: cellulose associated-defensive sensing partners? *Trends Plant Sci* 2008, **13**(4):160–164.
32. Brotman Y, Briff E, Viterbo A, Chet I: Role of swollenin, an expansin-like protein from *Trichoderma*, in plant root colonization. *Plant Physiol* 2008, **147**(2):779–789.
33. Richards TA, Soanes DM, Jones MDM, Vasieva O, Leonard G, Paszkiewicz K, Foster PG, Hall N, Talbot NJ: Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. *Proc Natl Acad Sci USA* 2011, **108**(37):15258–15263.
34. Wang M, Caetano-Anolles G: Global phylogeny determined by the combination of protein domains in proteomes. *Mol Biol Evol* 2006, **23**(12):2444–2454.
35. Bendtsen JD, Nielsen H, von Heijne G, Brunak S: Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004, **340**(4):783–795.
36. Harris PV, Welner D, McFarland KC, Re E, Navarro Poulsen JC, Brown K, Salbo R, Ding H, Vlasenko E, Merino S, *et al*: Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry* 2010, **49**(15):3305–3316.
37. Quinlan RJ, Sweeney MD, Lo Leggio L, Otten H, Poulsen JC, Johansen KS, Krogh KB, Jørgensen CI, Tovborg M, Anthonsen A, *et al*: Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci USA* 2011, **108**(37):15079–15084.
38. Crooks GE, Hon G, Chandonia JM, Brenner SE: WebLogo: a sequence logo generator. *Genome Res* 2004, **14**(6):1188–1190.
39. Mattinen ML, Linder M, Drakenberg T, Annila A: Solution structure of the cellulose-binding domain of endoglucanase I from *Trichoderma reesei* and its interaction with cello-oligosaccharides. *Eur J Biochem* 1998, **256**(2):279–286.

40. Peumans WJ, Van Damme EJ, Barre A, Rouge P: **Classification of plant lectins in families of structurally and evolutionary related proteins.** *Adv Exp Med Biol* 2001, **491**:27–54.

41. Larroque M, Ramirez D, Lafitte C, Borderies G, Dumas B, Gaulin E: **Expression and purification of a biologically active *Phytophthora parasitica* cellulose binding elicitor lectin in *Pichia pastoris*.** *Prot Expr Purif* 2011, **80**(2):217–223.

42. Cunningham S, Gerlach JQ, Kane M, Joshi L: **Glyco-biosensors: Recent advances and applications for the detection of free and bound carbohydrates.** *Analyst* 2010, **135**(10):2471–2480.

43. Spik G, Strecker G, Fournet B, Bouquelet S, Montreuil J, Dorland L, van Halbeek H, Vliegenthart JF: **Primary structure of the glycans from human lactotransferrin.** *Eur J Biochem* 1982, **121**(2):413–419.

44. Mazau D, Rumeau D, Esquerretugaye MT: **Two different famlies of hydroxyproline-rich glycoproteins in melon callus - Biochemical and immunochemical studies.** *Plant Physiol* 1988, **86**(2):540–546.

45. Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, Kemen E, Thines M, Ah-Fong A, Anderson R, Badejoko W, *et al*: **Signatures of adaptation to obligate Biotrophy in the *Hyaloperonospora arabidopsidis* genome.** *Science* 2010, **330**(6010):1549–1551.

46. Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stuber K, Ver Loren Van Themaat E, Brown JK, Butcher SA, Gurr SJ, *et al*: **Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism.** *Science* 2010, **330**(6010):1543–1546.

47. Gurvan M, Tonon T, Scornet D, Cock MJ, Kloareg B: **The cell wall polysaccharide metabolism of the brown alga *Ectocarpus siliculosus*. Insights into the evolution of extracellular matrix polysaccharides in Eukaryotes.** *New Phytol* 2010, **188**(1):82–97.

48. Naithani S, Chookajorn T, Ripoll DR, Nasrallah JB: **Structural modules for receptor dimerization in the S-locus receptor kinase extracellular domain.** *Proc Natl Acad Sci USA* 2007, **104**(29):12211–12216.

49. Brown PJ, Gill AC, Nugent PG, McVey JH, Tomley FM: **Domains of invasion organelle proteins from apicomplexan parasites are homologous with the Apple domains of blood coagulation factor XI and plasma pre-kallikrein and are members of the PAN module superfamily.** *FEBS Lett* 2001, **497**(1):31–38.

50. Banyai L, Patthy L: **Importance of intramolecular interactions in the control of the fibrin affinity and activation of human plasminogen.** *J Biol Chem* 1984, **259**(10):6466–6471.

51. Herwald H, Renne T, Meijers JC, Chung DW, Page JD, Colman RW, Muller-Esterl W: **Mapping of the discontinuous kininogen binding site of prekallikrein. A distal binding segment is located in the heavy chain domain A4.** *J Biol Chem* 1996, **271**(22):13061–13067.

52. Gong H, Kobayashi K, Sugi T, Takemae H, Kurokawa H, Horimoto T, Akashi H, Kato K: **A novel PAN/apple domain-containing protein from *Toxoplasma gondii*: characterization and receptor identification.** *PLoS One* 2012, **7**(1):e30169.

53. Esko JD, Kimata K, Lindahl U, Esko JD, Kimata K, Lindahl U, *et al*: **Proteoglycans and sulfated glycosaminoglycans**. In *Essentials of Glycobiology*. 2nd edition. Edited by Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009:211–252.

54. Carzaniga R, Bowyer P, O'Connell RJ: **Production of extracellular matrices during development of infection structures by the downy mildew *Peronospora parasitica*.** *New Phytol* 2001, **149**(1):83–93.

55. Brockhausen I, Schachter H, Stanley P: **O-GalNAc Glycans**. In *Essentials of Glycobiology*. 2nd edition. Edited by Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009:183–194.

56. Grenville-Briggs LJ, Avrova AO, Hay RJ, Bruce CR, Whisson SC, Van West P: **Identification of appressorial and mycelial cell wall proteins and a survey of the membrane proteome of *Phytophthora infestans*.** *Fungal Biol*, **114**(9):702–723.

57. Gaulin E, Madoui MA, Bottin A, Jacquet C, Mathé C, Couloux A, Wincker P, Dumas B: **Transcriptome of *Aphanomyces euteiches*: new oomycete putative pathogenicity factors and metabolic pathways.** *PLoS One* 2008, **3**(3):e1723.

58. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, *et al*: **InterPro: the integrative protein signature database.** *Nucleic Acids Res* 2009, **37**(Database issue):D211–D215.

59. Gaboriaud C, Bissery V, Benchetrit T, Mornon JP: **Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences.** *FEBS Lett* 1987, **224**(1):149–155.

60. Zhou H, Mazzulla MJ, Kaufman JD, Stahl SJ, Wingfield PT, Rubin JS, Bottaro DP, Byrd RA: **The solution structure of the N-terminal domain of hepatocyte growth factor reveals a potential heparin-binding site.** *Structure* 1998, **6**(1):109–116.

61. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**(22):4673–4680.

62. Ponder JW, Richards FM: **Tertiary templates for proteins - Use of packing criteria in the enumeration of allowed sequences for different structural classes.** *J Mol Biol* 1987, **193**(4):775–791.

63. Mas MT, Smith KC, Yarmush DL, Aisaka K, Fine RM: **Modeling the anti-CEA antibody combining site by homology and conformational search.** *Proteins* 1992, **14**(4):483–498.

64. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM: **AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.** *J Biomol NMR* 1996, **8**(4):477–486.

65. Townsend RR, Hardy MR, Cumming DA, Carver JP, Bendiak B: **Separation of branched sialylated oligosaccharides using high-pH anion-exchange chromatography with pulsed amperometric detection.** *Anal Biochem* 1989, **182**(1):1–8.

66. Campargue C, Lafitte C, Esquerré-Tugayé MT, Mazau D: **Analysis of hydroxyproline and hydroxyproline-arabinosides of plant origin by high-performance anion-exchange chromatography-pulsed amperometric detection.** *Anal Biochem* 1998, **257**(1):20–25.

67. Esnouf RM: **Further additions to MolScript version 1.4, including reading and contouring of electron-density maps.** *Acta Crystallogr D Biol Crystallogr* 1999, **55**(Pt 4):938–940.

68. Merritt EA, Murphy ME: **Raster3D Version 2.0. A program for photorealistic molecular graphics.** *Acta Crystallogr D Biol Crystallogr* 1994, **50**(Pt 6):869–873.

69. Kemen E, Gardiner A, Schultz-Larsen T, Kemen AC, Balmuth AL, Robert-Seilaniantz A, Bailey K, Holub E, Studholme DJ, Maclean D, *et al*: **Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*.** *PLoS Biol* 2011, **9**(7):e1001094.

70. Haas B, Kamoun S, Zody MC, Jiang RH, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, *et al*: **Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*.** *Nature* 2009, **461**(7262):393–398.

71. Lévesque CA, Brouwer H, Cano L, Hamilton JP, Holt C, Huitema E, Raffaele S, Robideau GP, Thines M, Win J, *et al*: **Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire.** *Genome Biol* 2010, **11**(7):R73.

72. Amselem J, Cuomo CA, van Kan JA, Viaud M, Benito EP, Couloux A, Coutinho PM, de Vries RP, Dyer PS, Fillinger S, *et al*: **Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*.** *PLoS Genet* 2011, **7**(8):e1002230.

73. Klosterman SJ, Subbarao KV, Kang S, Veronese P, Gold SE, Thomma BP, Chen Z, Henrissat B, Lee YH, Park J, *et al*: **Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens.** *PLoS Pathog* 2011, **7**(7):e1002137.