Molecular insight into cellulose degradation by the phototrophic green alga Scenedesmus.

Julieta Barchiesi[1], María B. Velazquez[1], María Busi[1], Diego F. Gomez-Casati[1], and Chitralekha Nag-Dasgupta[2]

[1]Centro de Estudios Fotosinteticos y Bioquimicos
[2]Research Cell Lucknow Amity University Uttar Pradesh India

September 1, 2022

Abstract

Lignocellulose is the most abundant natural biopolymer on earth and a potential raw material for the production of fuels and chemicals. However, only some organisms such as bacteria and fungi produce the necessary enzymes to metabolize it. In this work we detected the presence of extracellular cellulases in the genome of five species of *Scenedesmus.* These microalgae grow in both, freshwater and saltwater regions as well as in soils, displaying highly flexible metabolic properties. The comparison of sequences of the different cellulases with hydrolytic enzymes from other organisms by means of multi-sequence alignments and phylogenetic trees showed that these enzymes belong to the families of glycosyl hydrolases 1, 5, 9 and 10. In addition, most of these presented a greater similarity of sequence with enzymes from invertebrates, fungi, bacteria and other microalgae than with cellulases from plants; and the 3D modeling data obtained showed that both the main structures of the modeled proteins and the main amino acid residues implicated in catalysis and substrate binding are well conserved in *Scenedesmus* enzymes. We propose that these cellulase-producing phototrophic microorganisms could act as catalysts for the hydrolysis of cellulosic biomass fueled by sunlight.

Molecular insight into cellulose degradation by the phototrophic green alga *Scenedesmus.*

María B. Velazquez[a], María V. Busi[a], Diego F. Gomez-Casati[a], Chitralekha Nag-Dasgupta[b]* and Julieta Barchiesi[a]*.

[a] Centro de Estudios Fotosintéticos y Bioquímicos (CEFOBI-CONICET), Universidad Nacional de Rosario, Argentina.

[b] Research Cell, Lucknow, Amity University Uttar Pradesh, India.

Running title: *Cellulose degradation by Scenedesmus.*

Julieta Barchiesi*

Corresponding autor at: Centro de Estudios Fotosintéticos y Bioquímicos (CEFOBI-CONICET), Suipacha 570, Rosario 2000, Argentina.

E-mail address :*barchiesi@cefobi-conicet.gov.ar*

Chitralekha Nag Dasgupta*

Co-corresponding autor at: Research Cell, Lucknow, Amity University Uttar Pradesh, Sector 125, Noida, 201313, India.

E-mail address : cndasgupta@lko.amity.edu

Abstract

Lignocellulose is the most abundant natural biopolymer on earth and a potential raw material for the production of fuels and chemicals. However, only some organisms such as bacteria and fungi produce the necessary enzymes to metabolize it. In this work we detected the presence of extracellular cellulases in the genome of five species of*Scenedesmus* . These microalgae grow in both, freshwater and saltwater regions as well as in soils, displaying highly flexible metabolic properties. The comparison of sequences of the different cellulases with hydrolytic enzymes from other organisms by means of multi-sequence alignments and phylogenetic trees showed that these enzymes belong to the families of glycosyl hydrolases 1, 5, 9 and 10. In addition, most of these presented a greater similarity of sequence with enzymes from invertebrates, fungi, bacteria and other microalgae than with cellulases from plants; and the 3D modeling data obtained showed that both the main structures of the modeled proteins and the main amino acid residues implicated in catalysis and substrate binding are well conserved in *Scenedesmus* enzymes.

We propose that these cellulase-producing phototrophic microorganisms could act as catalysts for the hydrolysis of cellulosic biomass fueled by sunlight.

Keywords

Cellulases, Scenedesmus, Endoglucanases, β-glucosidases, exocellulases

1. Introduction

Generation of renewable energy resources and waste management are the major concern in twenty first century. Lignocellulosic agricultural and forest wastes are the promising feedstock for production of biofuel and value-added products due its high availability and low cost[1]. Nevertheless, no commercial process has still been reported for the enzymatic hydrolysis of cellulose. The main reason is the high cost of the required enzymes, their low specific activity, their susceptibility to inactivation and the difficulty to recycle them[2].

A group of naturally occurring cellulases are reported from heterotrophic microorganisms including bacteria and fungi[3]. These organisms secrete cellulases to utilize cellulose as a carbon source. Bioconversion processes involve the hydrolysis of cellulose to produce reducing sugars; further fermentation of the sugars to ethanol and other bioproducts [4]. Cellulases hydrolyze the β-1,4 glycosidic bonds of the glucose polymer by two different ways, endoglucanases cut random positions along the cellulose chain, and exoglucanases progressively act on the terminal ends of the polymer, releasing either glucose molecules, or cellobiose[3]. Finally, the cellobiose molecules produced are converted to glucose by intra- and extracellular β -glucosidases (EC 3.2.1.21), celludextrinases (EC 3.2.1.4), and cellodextrin phosphorylases (EC 2.4.1.49), depending upon the characteristic of each cellulolytic species [5]. Other than heterotrophs, cellulases belonging to glucoside hydrolase family (GH9) are also described from higher plants[6]. However, it has been reported that plant cellulases participate in the biosynthesis and/or remodeling of cellulose rather than in its degradation [6,7].

Algae are phototrophs, ubiquitous with versatile metabolic pathways, which have been well exploited to obtained multiple products through algal refinery [8,9]. However, the presence of cellulases and cellulolytic activity was poorly described in algae. In 1966, Dvořáková-Hladká et al. reported the presence of β-glucosidase activity in *S. obliquus* , which allows it to grow using cellobiose as a substrate [10]. In 1970 Burczyk and col. reported the presence of extracellular cellulases in *Scenedesmus obliquus*because cell walls accumulated in the medium as a result of mother cells autospore release were deprived of the cellulose layer present in daughter cells. [11]. In 2012, Blifernez-Klassen and col. observed that the photoheterotrophic microalgae *Chlamydomonas reinhardtii* , was also capable of degrading and assimilating exogenous cellulose [3]. This interesting finding led us to investigate the presence of cellulases in *S. quadricauda* . This organism is a freshwater, non-mobile green algae which usually forms colonies of four cells. It belongs to the same class of green algae (Chlorophyceae) as the genus *Chlamydomonas* . *S. quadricauda*has gained great importance due to its high capacity for effluent treatment, CO_2 capture and biofuel production as we showed in a previous work [9]. The *S. quadricauda*LWG002611 genome was sequenced and functional genes of different metabolic pathways were identified, such as those involved in the synthesis of triacyl glycerol (TAG) [9]. However, no evidence on cellulase secretion and cellulose utilization in this alga was found or reported, as well as genes that encode proteins with glycoside hydrolase activity have not yet been described.

In this work, we have identified different genes in the genome sequence of *S. quadricauda* LWG002611, belonging to GH1, GH5, GH9 and GH10 families. Furthermore, a comparative bioinformatic analysis was conducted in several available *Scenedesmaceae* algae genome (*Scenedesmus obliquus* EN0004 v1.0, *Scenedesmus obliquus*UTEX B 3031, *Scenedesmus obliquus* var. DOE0013 v1.0,*Scenedesmus* sp. NREL 46B-D3 v1.0, and *S. quadricauda*LWG002611) to identify multiple homologs of endoglucanase, β-glucosidase and exocellulase genes. Additionally, a phylogenetic analysis and a 3D protein modeling were achieved. Our results showed that the 3D structures of all the modeled domains obtained and the main catalytic amino acid residues implicated in cellulolytic activity are well conserved in the *Scenedesmus* analyzed enzymes.

These new findings open the opportunity to identify new cellulases from algae, as well as carry out their functional characterization to be used in biotechnological applications.

2. Methods

2.1. Sequence search, alignment and phylogenetic analysis

Cellulase sequences of *S. quadricauda* were identified from genome sequence data of our previous study [9] using protein folding homology analysis by Phyre2 [12] and Blast-N similarity study [13] with *Monoraphidium neglectum* taken as reference, and their details are included in table 1. Other analyzed sequences of *Scenedesmus* were taken from PhycoCosm [14] or NCBI {https://www.ncbi.nlm.nih.gov/} and their accession numbers are shown in table 2. Conserved domains, signal peptide, and GH-family assignment were identified with Prosite patterns [15], DeepLoc [16] and PredAlgo[17]. The sequences were aligned and processed with Clustal Omega [18] and visualized with ESPript 3.0[19]. To construct the phylogenetic trees, all the sequences were aligned with sequences from phylogenetically distant β-1,4-endoglucanases, β-glucosidases or exocellulases (respectively) from microalgae, fungi, plants, invertebrates and bacteria and processed with Gblock v0.91b before analyzing them in MEGA 6.06[20,21]. Enzymes signal peptides were not included in the phylogenetic analysis. The phylogenetic trees were built by Maximum Likelihood method in MEGA 6.06 version with the model and the restrictions suggested by the program. Phylogenies were determined by Bootstrap Analysis of 100 replicates.

2.2. Protein 3D modeling

The protein 3D models were generated with RaptorX Contact Prediction Server [22]. Superposition between each model and the templates was done using the align command in PyMOL 2.3.1 version {The PyMOL Molecular Graphics System}. The regions implicated in substrate binding and activity were manually annotated using the pattern sequences or 3D structure of cellulase templates available in Prosite and RCSB Pdb database (rcsb.org [23]).

3. Results and Discussion

3.1. Identification of genes encoding different cellulases in Scenedesmus **LWG002611**

In the present work, we performed an analysis of *Scendesmus quadricauda* LWG002611 draft genome sequence, and eleven endoglucanases, β-glucosidases and exoglucanases gene sequences have been detected by protein folding homology analysis by Phyre2 and similarity study with*Monoraphidium neglectum,* a closely related species which has been taken a reference sequence for draft genome analysis.

These sequences have shown 82.26-98.38% similarity with the reference sequences of *M. neglectum* (Table 1).

According to the amino acid sequence analysis by Pfam two GH9 (Scequ2611|3068 and Scequ2611|4665), one GH5 (Scequ2611|2009), three GH1 (Scequ2611|3544, Scequ2611|9833 and Scequ2611|10006), one GH10 (Scequ2611|547), and four undefined GH family short proteins (Scequ2611|8404, Scequ2611|9353, Scequ2611|13370 and Scequ2611|13657) were detected (Table 1, Supplementary file 1)

3.2. In silico **analysis of *Scenedesmaceae*cellulases**

We have used different bioinformatic tools to analyze the cellulases present in five different species of Scenedesmus:*Scenedesmus obliquus*EN0004 v1.0, *Scenedesmus obliquus* UTEX B 3031, *Scenedesmusobliquus* var. DOE0013 v1.0, *Scenedesmus sp* . NREL 46B-D3 v1.0, *Scenedesmus sp* PABB004 and *Scenedesmus quadricauda*LWG 002611. The presence of conserved motifs, the arrangement of diverse domains (catalytic domain, carbohydrate-binding module (CBM) and linker regions) and the phylogenetic relationship between the proteins were determined. Their amino acid sequences were compared with enzymes previously characterized from others taxonomic groups.

The cellulases sequences analyzed in this study were classified with KEGG ENZYME Database Entry by Phycocosm [14], into three groups: (i) endo-β-glucanases (EC 3.2.1.4), (ii) β-glucosidases (EC 3.2.1.21) and (iii) cellobiohydrolases or exoglucanases (EC 3.2.1.91) (Table 2).

In *S. quadricauda* LWG002611 sequences with a higher length than 300 amino acids and with higher similarity with previously studied cellulases are code by the genes: Scequ2611|2009, Scequ2611|9833 and Scequ2611|547 (percentage of identity: 98.32%, 93.93% and 82.26% respectively, table 1), which were chosen to be included in the posterior bioinformatic analysis

3.2.1. Ενδο-β-γλυςανασες

Among the selected species we found thirty genes that encode different endoglucanases. Their catalytic domains belong to the GH5 and GH9 family of CAZymes (Carbohydrate-Active Enzymes) [24] and seven of these proteins present CBM from families 1 and 2. While CBM2 were found only in GH5 endoglucanases, CBM1 were associated with some GH9 endoglucanases (Figure 1 and Table 2).

Our results showed that three GH9 endoglucanases, Sceob1|18668, SceobE4|579152, SceobDOE|757385, present Big_1 domains in their C-terminal. Big_1 is a bacterial immunoglobulin (Ig)-like domain, usually present in GH9 endoglucanases. The functions of these Ig-like modules are not clear; however, they are supposed to be involved in the catalytic efficiency or in the structural stability of GH9 endoglucanases [25,26]. On the other hand, KAF8065624.1 present a LPMO domain in its N-terminal region. LPMO is a lytic polysaccharide monooxygenase domain that usually act synergically with GH9 domains, increasing the cellulolytic enzyme activity[27].

Most of the GH9 endoglucanases analyzed are secreted or anchored to the cell membrane (with the catalytic domain localized to the outer surface of the plasma membrane). On the other hand, only three identified proteins GH5 endonucleases are predicted to be secreted, while the other four identified proteins would be cytoplasmic enzymes (Table 2).

The analysis performed with the Prosite Database showed three highly conserved regions present in GH9 endoglucanases which contains conserved residues important for their catalytic activity [5,28](Figure 2). The first region comprises the DAGD motif, where the first Asp (Asp54, Nasta 1KS8 numbering) is an active site residue. The second region contains a conserved RPHHR sequence, where the first His (His359, Nasta 1KS8 numbering) is also part of the active site of GH9 endoglucanases. Finally, Region III contains two Asp and Glu residues (Asp399 and Glu412, Nasta 1KS8 numbering) that would be involved in catalysis. Thus, all the proteins identified and analyzed in this study contain four acidic residues, D, H, D and E, in the mentioned regions, which would be part of the active site of GH9 endoglucanases and would be involved in catalysis, with the exception of ScsoPA4|KAF8062061.1 and SceobDOE|1035052 proteins, that lack the catalytic D from Region I and the H residue from Region II, respectively (Fig 2.A.).

S. quadricauda LWG002611 possess another two hypothetical GH9 endoglucanase proteins (Scequ2611|3068 and Scequ2611|4665) showing a high similarity with respect to the analyzed *Scenedesmus* endoglucanases (Scequ2611|3068 34-36% with KAF6265438.1 and KAF8066338.1, Scequ2611|4665 30% to KAF6264795.1) but they lack the essential Glu residue from Region III (Fig 2B). Further studies are needed to determine if Scequ2611|3068 and Scequ2611|4665 are catalytically inactive enzymes or if they have a different mechanism than traditional glucosyl hydrolases.

Cellulases usually present linkers with length from 6–14 residues long up to >100 residues [5,29]. From*Scenedesmus* endoglucanases analyzed, most of the CBM1 containing enzymes showed putative P/S-rich or poliQ linkers, mainly located between the GH9 and the CBM regions (Fig 2.A.). These proline or glutamine rich spacers constitute a rigid type of linkers which are thought to act as spacers to avoid non-native interactions between domains that may affect the correct folding of proteins[30]. In addition, the linkers would allow cellulases to push forward on the exterior of the polysaccharide with a caterpillar-like movement [5,31].

In contrast to what was described in GH9 endoglucanases from other algae, such as *Chlamydomonas* , *Volvox* and *Gonium*[5] CBM1 domains are located either at the C- or N-terminus of some the studied *Scenedesmus* GH9 endoglucanases. Notably, both, the N- and the C-terminus CBM1 analyzed, are cysteine-rich domains as previously described in *Chlamydomonas* .

A phylogenetic tree was also constructed (Fig. 3) using Gblock and the MEGA 6.06 software from the alignment of the amino acid sequences from *Scenedesmus* GH9 enzymes together with homologous sequences identified with Blast-P from invertebrates, fungi, plants and bacteria. The organization of the tree branches suggest that *Scenedesmus* GH9 endoglucanases are evolutionarily closer to termites, worms, sea urchins and bivalves GH9 cellulases (red branches of the tree) rather than the enzymes from higher plants, fungi and bacteria.

Figure 4 shows the 3D model of the Sceobl1|32711 GH9 and CBM1 domains constructed with RaptorX Contact Prediction. The model created presents a similar fold to that described for previously characterized GH9 endoglucanases [5,32] with a $(\alpha/\alpha)_6$-barrel fold. Besides, the catalytic amino acid residues are positioned in a similar spatial location when Sceobl1|32711 model was superposed with 1ks8 template (an endocellulase from the termite *Nasutitermes takasagoensis* 40.24% identical to Sceobl1|32711 (79% cover). On the other hand, its N-terminal CBM1 showed a high sequence identity (99.6%, cover 35%) with the cellulose-binding domain of endoglucanase I from*Trichoderma reesei* (PDB entry: 4BMF) and its model present a good spatial conservation when both models where superposed.

Regarding GH5 endoglucanases, they present the consensus pattern [LIV]-[LIVMFYWGA](2)-[DNEQG]-[LIVMGST]-{SENR}-N-E-[PV]- [RHDNSTLIVFY] [15]. The C-terminal Glu is an active site residue. The predicted catalytic residues, Glu168 and Glu309 (Pyrho 3W6M numbering) are strictly conserved in all the GH5 endonucleases analyzed (Figure 5).

The Figure 6 shows the GH5 endoglucanase phylogenetic tree. The tree branching organization suggest that most of the GH5 proteins analyzed are evolutionarily closer to those of other microalgae and higher plants. However, the group of enzymes containing a CBM2 are closer to fungal and bacterial endoglucanases, suggesting a microbial origin.

The homology model of the Sceobl1|14060 GH5 domain present the (α/β)8 TIM barrel fold classical of GH5 family [33](Figure 7). The superposition with Pyrho 3W6M PDB structure (an hyperthermophilic endocellulase from *Pyrococcus horikoshii*) showed a conservation of the catalytic amino acid spatial location.

Respects its N-terminal CBM2 domain, Sceobl1|14060 has a high sequence identity (97.3% identity, 68% cover) with 2RTT PDB structure (a chitin-binding domain of Chi18aC from *Streptomyces coelicolor*); moreover, both models showed a high structural and spatial conservation.

3.2.2. ***β-γλυςοσιδασες***

In the the analyzed genomes we found twenty-nine putative β-glycosidases, all of them belonging to the GH1 family of CAZymes (Figure 8). The most common enzymatic activities reported for glycoside hydrolases of this family are β-glucosidases and β-galactosidases.

It has been previously described that one of the highly conserved regions in GH1 sequences has a glutamic acid residue and is classified as GH1_1 [34]. This region between positions 388-392 (*Nanochloropsis* β-glucosidase GH1 numbering, PDB code: 5YJ7,) presents a conserved sequence (V/I)TENG. The Glu residue would participate in the cleavage of the glycosidic bond by acting as a nucleophile [34]. This catalytic nucleophile was first identified as Glu358 in a β-glucosidase from Agrobacterium[35].

In our work, the conserved region (V/I)TENG was found in all the GH1 sequences analyzed: (Figure 9.A) GH1_1. The extended region defined as consensus is: [LIVMFSTC]-[LIVFYS] [LIV]-[LIVMST]-E-N-G-[LIVMFAR]-[CSAGN]. All of the proteins chosen in this work possess in this region the Glu involved in catalysis followed by Asn and Gly, as can be seen in Figure 9. However, the four amino acids upstream Glu residue appear to differ in the proteins, with Ile-Trp-Ile-Thr being predominant.

As a second signature pattern, the conserved region GH1_2 (Figure 8.A) was chosen for our analysis. This region, defined as: F-x-[FYWM]-[GSTA]-x-[GSTA]-x-[GSTA](2)-[FYNH]-[NQ]-x-E-x-[GSTA], is located at the N-terminal of GH1 β-glucosidases, however, it may not be present in some proteins of this family. The alignment of Sceobl|9031 and Sceobl|10236 sequences showed that these proteins do not contain that region (Figure 9.A). On the other hand, the protein SceoblEN4 |617109, possesses only nine of the fifteen amino acids established as consensus, which is equivalent to 60 %, and therefore would be considered that this domain is present but with some variations in the amino acid sequence. Similarly, Sceobl1|35463 and Scequ2611|9833 proteins have only 46 and 40 % of the consensus sequence, respectively. Scesp1|1644545 only possesses the last four amino acids of this consensus sequence, while the other eleven, do not correspond to the established consensus. The sequence of KAF8060308.1 BGLU11 shows some particular characteristics in this region. Although it possesses 86 % sequence identity respect to the established consensus sequence, it presents an insertion of several amino acids downstream the consensus (34 residues), before x-E-x-[GSTA]. This protein also contains an additional domain in its N-terminal region, a protein disulfide isomerase domain (cl36828: ER_PDI_fam Superfamily,[36]), previously involved in protein folding [37].

Interestingly, only five of the proteins analyzed contains the first Phe residue, which is characteristic of the GH1_2 pattern. It is interesting to note that none *Scenedesmaceae* β- glycosidases analyzed in this study presents CBM nor linkers.

On the other hand, *S. quadricauda* LWG002611 have a hypothetical GH1 β-glucosidase protein, named Scequ2611|3544, with high similarity with other *Scenedesmus* endoglucanases (86% cover and 29.45 % identity with KAF8059426.1) but that lacks the region containing the catalytic Glu residue (Figure 9.B).

The phylogenetic tree performed shows four β-glucosidase subgroups: (i) the plant GH1 subgroup, (ii) the GH1 from algae, (iii) the GH1 from fungi, and (iv) the GH1 from bacteria (Figure 10). The Scesp1|1509300, SceobDOE|17466, SceoblEN4|575894, and SceobDOE|32074 proteins are grouped in a branch close to the bacteria enzyme, that proposes the possible acquisition of these genes by horizontal transfer. On the other hand, the Scequ2611|9833 protein was found within a large group of GH1 enzymes from algae. This result suggests the correct inference of its sequence and the possibility that orthologous genes are those that code for proteins found on a nearby branch.

There is at least one representative of each species corresponding to a genus in each branch of the group of proteins from algae. This result suggests that the different β-glucosidases present in the different species could fulfill the same function and that it would not be redundant within the genus.

The homology model of the Sceob1|9434 β-glucosidase constructed with RaptorX Contact Prediction is shown in Figure 11. The superposition with *Nannochloropsis* oceanica BGLN1 β-glucosidase crystal structure (PDB code: 5YJ7) shows a catalytic amino acid positional conservation in the central region. Also, it presents an overall structure of TIM barrel, and the ENG residues conserved, which suggest a reliable protein function assignment of this new protein in

Scenedemus quadricauda.

3.2.3. Exocellulases

Exoglucanases or cellobiohydrolases (CBH) (EC 3.2.1.74; 1,4-β-D-glucan-glucanhydrolase) catalyze the successive hydrolysis of residues from the reducing and non-reducing ends of the cellulose polysaccharide, releasing cellobiose molecules as main product of the reaction [38]. These enzymes account for 40 to 70% of the total component of the cellulase system, and are able to hydrolyze crystalline cellulose.

An excellent candidate for use as a bait to explore algal genomes is the GH10 from *Cellulomonas fimi* (Cex-P07986 Uniprot). High resolution crystal structures are available and there is a large literature on the kinetic characterization of the enzyme and the identification of amino acid residues important to the mechanism of catalysis [39]. Among all the Scenedesmaceae genomes studied we found at least eleven enzymes, all of them putative GH10 bifunctional cellulase/xylanase proteins (Figure 12A).

These enzymes are monomeric proteins with a molecular mass ranging from 50 to 65 kDa, although there are smaller variants (41.5 kDa) in some fungi, such as *Sclerotium rolfsii* [40]. The calculated molecular mass of five of the eleven algal proteins studied were higher (between 70 and 100kDa), while in the other six enzymes is within the expected range (table 2).

In general, most of the proteins containing GH10 domains show a structure that matches an eightfold α/ β-barrel with a profound channel in the center [41]. However, it has been proposed that the exocellulases from the GH10 family form a transient tunnel by the extension of some loop regions upon substrate binding[42].

It has been reported that GH10 enzymes present a double-displacement 'retaining' hydrolysis mechanism, where one catalytic residue acts a nucleophile and the other acts as a general acid/base[43]. The catalytic nucleophile in Cex is Glu274 and the putative acid/base catalyst is Glu168 [44]. As shown in figure 12B, both residues are full conserved in the algal protein sequences. However, they were replaced by an Asp and Ile residues in the Scequ2611|547 protein.

We also performed a phylogenetic tree for the GH10 exoglucanase (Figure 13) from the alignment of the amino acid sequences from*Scenedesmus* GH10 enzymes together with homologous sequences from invertebrates, fungi, plants and bacteria. The branches distribution suggests the GH10 analyzed proteins are evolutionarily closer to those of other microalgae and higher plants.

In addition, we built a sequence homology model of the Sceob1|4623 exocellulase using RaptorX Contact Prediction is shown in Figure 14. The superposition with *Cellulomonas fimi*exocellulase crystal structure (PBD code: 1EXG) shows a spatial location conservation of the algal protein residues Glu236 and Glu338 with respect to catalytic residues from the bacterial protein, also suggesting the conservation of the catalytic site.

Conclusion

The discovery of new microbial sources of cellulases is a crucial strategy to reduce costs of various industrial processes using such enzymes. Cellulases are produced by various microorganisms including bacteria, fungi and actinomycetes. Recently was reported that they are also produced by some animals like termites and crayfish without certainties about his role *in vivo* [45]. The search, isolation and identification of new cellulose degrading microorganisms from different environments are of crucial importance to get new cellulases with unique and distinctive characteristics.

Microalgae are considered a valuable source of new enzymes with biotechnological potential. However, the presence of cellulolytic enzymes is meagre studied form this photosynthetic microorganisms.

Different works published during the last decade report cellulolytic activity (either by experimental evidence or by bioinformatic analysis) in *C. reinhardtii* , *V. carteri* , *G. pectorale* and*A. protothecoides* but the genus Scenedesmus had not been analysed[3,5,46].

This is the first bioinformatic analysis of Scenedesmaceae cellulases reported. It comprises GH5 and GH9 β-1,4-endoglucanases, GH1 β-glucosidases and GH10 exoglucanases. Our results shows that GH9 endoglucanases analyzed are phylogenetically closer to invertebrates, termites and bivalve rather than higher plant, bacteria or fungi. On the other hand, most of GH1 β-glucosidases analyzed are evolutionarily closer to enzymes of other microalgae, however, four of them are grouped in a branch close to the bacteria enzymes, result that suggests the probable gaining of their genes by horizontal transfer. In contrast, GH5 and GH10 studied enzymes are evolutionarily closer to enzymes of other microalgae and higher plants.

Most of the analyzed enzymes present signal peptides for membrane anchoring or extracellular secretion. This result suggests the presence of extracellular cellulolytic machinery in Scenedesmaceae. Only some of the analyzed enzymes were found to have additional modules and linkers besides its GH domains, and particularly a few endoglucanases have CBM modules, from CBM1 and CBM2 families.

The combination of GH catalytic domains together with CBMs and, in some cases linkers, propose that these cellulases would present an enhanced cellulolytic activity.

The presence of this battery of enzymes in the photoheterotrophic algae*Scenedesmus* suggest that these organisms are perfectly prepared for use of cellulose as carbon source. This strategy would represent an advantage that would have allowed Scenedesmaceae to occupy many environments in nature.

The findings reported in this work explores just one family within the Chlorophyta taxon, but it increases the evidence in favor of the presence of conserved cellulolytic machinery in photoheterotrophic organisms and encourages to continue with the search for cellulases in other species of microalgae.

References

1 Saini, J. K., Saini, R. & Tewari, L. Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments. *3 Biotech***5** , 337-353, doi:10.1007/s13205-014-0246-5 (2015).

2 Lynd, L. R., Weimer, P. J., van Zyl, W. H. & Pretorius, I. S. Microbial cellulose utilization: fundamentals and biotechnology.*Microbiology and molecular biology reviews : MMBR* **66** , 506-577, table of contents (2002).

3 Blifernez-Klassen, O. *et al.* Cellulose degradation and assimilation by the unicellular phototrophic eukaryote Chlamydomonas reinhardtii. *Nature communications* **3** , 1214, doi:10.1038/ncomms2210 (2012).

4 Menon, V. & Rao, M. Trends in bioconversion of lignocellulose: Biofuels, platform chemicals & biorefinery concept. *Progress in Energy and Combustion Science* **38** , 522-550, doi:*https://doi.org/10.1016/j.pecs.2012.02.002*(2012).

5 Guerriero, G. *et al.* Novel Insights from Comparative In Silico Analysis of Green Microalgal Cellulases. *International journal of molecular sciences* **19** , doi:10.3390/ijms19061782 (2018).

6 Hayashi, T., Yoshida, K., Park, Y. W., Konishi, T. & Baba, K. Cellulose metabolism in plants. *Int Rev Cytol* **247** , 1-34, doi:10.1016/S0074-7696(05)47001-1 (2005).

7 Minic, Z. Physiological roles of plant glycoside hydrolases.*Planta* **227** , 723-740, doi:10.1007/s00425-007-0668-y (2008).

8 Dasgupta, C. N. *et al.* Dual uses of microalgal biomass: An integrative approach for biohydrogen and biodiesel production.*Applied Energy* **146** , 202-208, doi:*https://doi.org/10.1016/j.apenergy.2015.01.070*(2015).

9 Nag Dasgupta, C. *et al.* Draft genome sequence and detailed characterization of biofuel production by oleaginous microalga Scenedesmus quadricauda LWG002611. *Biotechnology for biofuels***11** , 308, doi:10.1186/s13068-018-1308-4 (2018).

10 Dvořáková-Hladká, J. Utilization of organic substrates during mixotrophic and heterotrophic cultivation of algae. *Biologia Plantarum* **8** , 354, doi:10.1007/bf02930672 (1966).

11 Burczyk, J., Grzybek, H., Banas, J. & Banas, E. Presence of cellulase in the algae Scenedesmus. *Experimental cell research***63** , 451-453 (1970).

12 Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10** , 845-858, doi:10.1038/nprot.2015.053 (2015).

13 Boratyn, G. M. *et al.* BLAST: a more efficient report with usability improvements. *Nucleic acids research* **41** , W29-33, doi:10.1093/nar/gkt282 (2013).

14 Grigoriev, I. V. *et al.* PhycoCosm, a comparative algal genomics resource. *Nucleic acids research* **49** , D1004-D1011, doi:10.1093/nar/gkaa898 (2021).

15 Sigrist, C. J. *et al.* PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief Bioinform***3** , 265-274, doi:10.1093/bib/3.3.265 (2002).

16 Thumuluri, V., Almagro Armenteros, J. J., Johansen, A. R., Nielsen, H. & Winther, O. DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic acids research* , doi:10.1093/nar/gkac278 (2022).

17 Tardif, M. *et al.* PredAlgo: a new subcellular localization prediction tool dedicated to green algae. *Molecular biology and evolution* **29** , 3625-3639, doi:10.1093/molbev/mss178 (2012).

18 Sievers, F. & Higgins, D. G. The Clustal Omega Multiple Alignment Package. *Methods Mol Biol* **2231** , 3-16, doi:10.1007/978-1-0716-1036-7_1 (2021).

19 Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research***42** , W320-324, doi:10.1093/nar/gku316 (2014).

20 Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution* **17** , 540-552, doi:10.1093/oxfordjournals.molbev.a026334 (2000).

21 Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0.*Molecular biology and evolution* **30** , 2725-2729, doi:10.1093/molbev/mst197 (2013).

22 Xu, J. Distance-based protein folding powered by deep learning.*Proceedings of the National Academy of Sciences of the United States of America* **116** , 16856-16865, doi:10.1073/pnas.1821309116 (2019).

23 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic acids research* **28** , 235-242, doi:10.1093/nar/28.1.235 (2000).

24 Drula, E. *et al.* The carbohydrate-active enzyme database: functions and literature. *Nucleic acids research* **50** , D571-D577, doi:10.1093/nar/gkab1045 %J Nucleic Acids Research (2021).

25 Nguyen, K. H. V. *et al.* Some characters of bacterial cellulases in goats' rumen elucidated by metagenomic DNA analysis and the role of fibronectin 3 module for endoglucanase function. *Anim Biosci* **34** , 867-879, doi:10.5713/ajas.20.0115 (2021).

26 Phitsuwan, P., Lee, S., San, T. & Ratanakhanokchai, K. CalkGH9T: A Glycoside Hydrolase Family 9 Enzyme from Clostridium alkalicellulosi.**11** , 1011 (2021).

27 Sabbadin, F. *et al.* An ancient family of lytic polysaccharide monooxygenases with roles in arthropod development and biomass digestion. *Nature communications* **9** , 756, doi:10.1038/s41467-018-03142-x (2018).

28 Tomme, P. *et al.* Identification of a histidyl residue in the active center of endoglucanase D from Clostridium thermocellum.*The Journal of biological chemistry* **266** , 10313-10318 (1991).

29 Sammond, D. W. *et al.* An iterative computational design approach to increase the thermal endurance of a mesophilic enzyme.*Biotechnology for biofuels* **11** , 189, doi:10.1186/s13068-018-1178-9 (2018).

30 George, R. A. & Heringa, J. An analysis of protein domain linkers: their classification and role in protein folding. *Protein engineering* **15** , 871-879, doi:10.1093/protein/15.11.871 (2002).

31 Receveur, V., Czjzek, M., Schulein, M., Panine, P. & Henrissat, B. Dimension, shape, and conformational flexibility of a two domain fungal cellulase in solution probed by small angle X-ray scattering. *The Journal of biological chemistry* **277** , 40887-40892, doi:10.1074/jbc.M205404200 (2002).

32 Foley, M. H. *et al.* A Cell-Surface GH9 Endo-Glucanase Coordinates with Surface Glycan-Binding Proteins to Mediate Xyloglucan Uptake in the Gut Symbiont Bacteroides ovatus. *Journal of molecular biology* **431** , 981-995, doi:10.1016/j.jmb.2019.01.008 (2019).

33 Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **3** , 853-859, doi:10.1016/S0969-2126(01)00220-9 (1995).

34 de Giuseppe, P. O. *et al.* Structural basis for glucose tolerance in GH1 beta-glucosidases. *Acta crystallographica. Section D, Biological crystallography* **70** , 1631-1639, doi:10.1107/S1399004714006920 (2014).

35 Withers, S. G. *et al.* Unequivocal demonstration of the involvement of a glutamate residue as a nucleophile in the mechanism of a retaining glycosidase. *Journal of the American Chemical Society***112** , 5887-5889, doi:10.1021/ja00171a043 (1990).

36 Lu, S. *et al.* CDD/SPARCLE: the conserved domain database in 2020. *Nucleic acids research* **48** , D265-D268, doi:10.1093/nar/gkz991 (2020).

37 Powers, S. L. & Robinson, A. S. PDI improves secretion of redox-inactive beta-glucosidase. *Biotechnology progress***23** , 364-369, doi:10.1021/bp060287p (2007).

38 Quiroz-Castañeda, R. E. & Folch-Mallol, J. L. Plant cell wall degrading and remodeling proteins: current perspectives %J Biotecnología Aplicada. **28** , 205-215 (2011).

39 Duedu, K. O. & French, C. E. Characterization of a Cellulomonas fimi exoglucanase/xylanase-endoglucanase gene fusion which improves microbial degradation of cellulosic biomass. *Enzyme and microbial technology* **93-94** , 113-121, doi:10.1016/j.enzmictec.2016.08.005 (2016).

40 Martin, M., Wayllace, N. Z., Valdez, H. A., Gomez-Casati, D. F. & Busi, M. V. Improving the glycosyltransferase activity of Agrobacterium tumefaciens glycogen synthase by fusion of N-terminal starch binding domains (SBDs). *Biochimie* **95** , 1865-1870, doi:10.1016/j.biochi.2013.06.009 (2013).

41 Gao, F., Jiang, Y., Zhou, G. H. & Han, Z. K. The effects of xylanase supplementation on growth, digestion, circulating hormone and metabolite levels, immunity and gut microflora in cockerels fed on wheat-based diets. *Br Poult Sci* **48** , 480-488, doi:10.1080/00071660701477320 (2007).

42 Schubot, F. D. *et al.* Structural basis for the exocellulase activity of the cellobiohydrolase CbhA from Clostridium thermocellum.*Biochemistry* **43** , 1163-1170, doi:10.1021/bi030202i (2004).

43 Gilkes, N. R. *et al.* Structural and functional relationships in two families of beta-1,4-glycanases. *European journal of biochemistry / FEBS* **202** , 367-377, doi:10.1111/j.1432-1033.1991.tb16384.x (1991).

44 MacLeod, A. M., Lindhorst, T., Withers, S. G. & Warren, R. A. The acid/base catalyst in the exoglucanase/xylanase from Cellulomonas fimi is glutamic acid 127: evidence from detailed kinetic studies of mutants.*Biochemistry* **33** , 6371-6376, doi:10.1021/bi00186a042 (1994).

45 Watanabe, H. & Tokuda, G. Animal cellulases. *Cellular and molecular life sciences : CMLS* **58** , 1167-1178, doi:10.1007/PL00000931 (2001).

46 Vogler, B. W. *et al.* Characterization of plant carbon substrate utilization by Auxenochlorella protothecoides. *Algal Research* **34** , 37-48, doi:*https://doi.org/10.1016/j.algal.2018.07.001*(2018).

Table 1: Predicted cellulases$^{\#}$ of *S. quadricauda* LWG 002611

S. No.	**Genes**	**Genes**	**Predicted mRNAs**	**Predicted proteins**	**Predicted proteins**	**Predicted**
	ID	**Length (bp)**	**Length (bp)**	**Length (aa)**	**Size (KD)**	**GH Fami**

S. No.	Genes	Genes	Predicted mRNAs	Predicted proteins	Predicted proteins	Predicte
1.	2009	2176	963	321	34.45	GH5
2.	3068	3243	1539	513	54.83	GH9
3.	8404	1472	303	101	11.19	inc
4.	9353	635	225	75	7.69	inc
5.	4665	3365	1332	443	48.97	GH9
6.	3544	4540	1272	424	48.28	GH1
8.	9833	5359	1173	390	43.37	GH1
9.	10006	1578	540	181	10.78	GH1
10.	13370	1613	291	96	10.78	inc
11.	13657	523	363	121	11.91	inc
12.	547	4033	2165	734	80.85	GH10

#Sequences are separately provided in Supplementary file; inc: Inconclusive result due to short sequence

Table 2

Organism	ID/PBD	Conserved domains	Cell Location	Size (aa), mol.wt	Taxonomic group
ENDOGLUCANASES					
Scenedesmus obliquus EN0004 v1.0	365111	CBM2, GH5	Cytoplasm	555, 59.86 kDa	Bacteria
Scenedesmus obliquus EN0004 v1.0	579152	GH9	Cytoplasm	716, 75.57 kDa	Green algae, Insect
Scenedesmus obliquus EN0004 v1.0	610731	5TM, GH9	Extracellular	891,91.43 kDa	Termites
Scenedesmus obliquus UTEX B 3031	199	1TM, GH9	Cell Membrane/ Extracelular	665, 70.87 kDa	Earthworms
Scenedesmus obliquus UTEX B 3031	8271	GH9	Extracellular	489, 53.02 kDa	Sea urchin (animal). Insect
Scenedesmus obliquus UTEX B 3031	14060	CBM2, GH5	Cytoplasm	551, 59.37 kDa	Bacteria
Scenedesmus obliquus UTEX B 3031	18668	CBM1, GH9	Extracellular	710, 74.88 kDa	Green algae, Insect
Scenedesmus obliquus UTEX B 3031	25222	GH9	Extracellular	861, 88.71 kDa	Green algae, Sea urchin
Scenedesmus obliquus UTEX B 3031	35062	GH9	Extracellular	487, 53.70 kDa	Termites, insects
Scenedesmus obliquus UTEX B 3031	19147	CBM1, GH9	Extracellular	730, 77.78 kDa	Thermophilic bacterium

Scenedesmus obliquus UTEX B 3031	32711	1TM, GH9	Extracellular	629, 67.22 kDa	Microalgae, Sea urchin, Insect
Scenedesmus obliquus var. DOE0013 v1.0	1035052	GH9	Extracellular	439, 47.68 kDa	Earthworms, Chordates
Scenedesmus obliquus var. DOE0013 v1.0	739074	GH9	Extracellular	479, 51.60 kDa	Green algae, Sea urchin
Scenedesmus obliquus var. DOE0013 v1.0	757385	GH9	Extracellular	733, 77.30 kDa	Green algae, Termites
Scenedesmus obliquus var. DOE0013 v1.0	776386	1TM, GH9	Cell Membrane (GH9 outside)	521, 57.08 kDa	Termites
Scenedesmus obliquus var. DOE0013 v1.0	826696	1TM, GH9	Cell Membrane (GH9 outside)	611, 65.76 kDa	Earthworms, Termites Green algae (una sola coccomyxa)
Scenedesmus obliquus var. DOE0013 v1.0	1002809	GH5	Extracellular	419, 45.87 kDa	Fungi, cellulolityc bacteria
Scenedesmus obliquus var. DOE0013 v1.0	1008808	CBM2 GH5	Cytoplasm	518, 55.90 kDa	Bacteria
Scenedesmus obliquus var. DOE0013 v1.0	761682	CBM1, GH9	Extracellular	682, 72.02 kDa	Green algae, Worm
Scenedesmus sp. NREL 46B-D3 v1.0	956336	GH9	Cytoplasm	511, 55.45 kDa	Worm, Termite
Scenedesmus sp. NREL 46B-D3 v1.0	1003724	GH5	Cytoplasm	424, 46.48 kDa	Archaea, Bacteria
Scenedesmus sp. NREL 46B-D3 v1.0	1274183	GH9	Extracellular	473, 50.60 kDa	Sea urchin Termites
Scenedesmus sp. NREL 46B-D3 v1.0	1508728	GH5	Extracellular	451, 49.56 kDa	Worms, termite
Scenedesmus sp. NREL 46B-D3 v1.0	1655835	GH9	Extracellular	360, 38.96 kDa	Amoeba, Bacteria, worms
Scenedesmus sp. NREL 46B-D3 v1.0	1693221	GH9	Extracellular	614, 66.93 kDa	Green algae, Termite
Scenedesmus PABB0004	KAF8061121.1 (celF)	CBM1, GH9	Cytoplasm	769, 80.61 kDa	Green algae, sea urchin
Scenedesmus PABB0004	KAF8066338.1 (celD)	LPMO, GH9	Cell membrane	672, 69.77 kDa	Green algae, Worm Anemones

Scenedesmus PABB0004	KAF8065624.1 (celF)	GH9	Extracellular	1166, 119.94 kDa	Green algae, Sea anemone Crustaceans,
Scenedesmus PABB0004	KAF8062061.1 (celZ)	GH9	Cell membrane	766, 77.74 kDa	Green algae, Termite Bivalve.
Scenedesmus quadricauda LGW0026011	2009	GH5	Cell Membrane (GH5 outside)	321, 34.33 kDa	Green algae, Bacteria
β- ΓΛΥ῝ΟΣΙΔΑΣΕΣ					

Organism	**ID/PBD**	**Conserved domains**	**Cell Location**	**Size (a**
Scenedesmus obliquus EN0004 v1.0	574933	GH1	Extracellular	565, 62.
Scenedesmus obliquus EN0004 v1.0	575894	GH1	Extracellular	795, 87.
Scenedesmus obliquus EN0004 v1.0	610267	GH1	Extracellular	596, 64.
Scenedesmus obliquus EN0004 v1.0	617109	GH1	Extracellular	484, 54.
Scenedesmus obliquus UTEX B 3031	2100	GH1	Extracellular	490, 54.
Scenedesmus obliquus UTEX B 3031	3494	GH1	Extracellular	589, 64.
Scenedesmus obliquus UTEX B 3031	8342	GH1	Extracellular	707, 78.
Scenedesmus obliquus UTEX B 3031	9031	GH1	Extracellular	389, 43.
Scenedesmus obliquus UTEX B 3031	10236	GH1	Extracellular	361, 40.
Scenedesmus obliquus UTEX B 3031	17466	GH1	Extracellular	814, 89.
Scenedesmus obliquus UTEX B 3031	23136	GH1	Extracellular	558, 61.
Scenedesmus obliquus UTEX B 3031	26740	GH1	Extracellular	560, 61.
Scenedesmus obliquus UTEX B 3031	32074	GH1	Extracellular	917, 10
Scenedesmus obliquus UTEX B 3031	32663	GH1	Membrane	573, 64.
Scenedesmus obliquus UTEX B 3031	35463	GH1	Extracellular	477, 54.
Scenedesmus sp. NREL 46B-D3 v1.0	88126	GH1	Extracellular	539, 59.
Scenedesmus sp. NREL 46B-D3 v1.0	1297936	GH1	Extracellular	512, 57.
Scenedesmus sp. NREL 46B-D3 v1.0	1298554	GH1	Membrane	489, 53.
Scenedesmus sp. NREL 46B-D3 v1.0	1298717	GH1	Membrane	488, 53.
Scenedesmus sp. NREL 46B-D3 v1.0	1507547	GH1	Extracellular	600, 66.
Scenedesmus sp. NREL 46B-D3 v1.0	1644545	GH1	Extracellular	461, 52.
Scenedesmus sp. NREL 46B-D3 v1.0	1509300	GH1	Extracellular	825, 91.
[Scenedesmus sp. PABB004]	KAF8062692.1	GH1	Chloroplastic Membrane	1210, 1
[Scenedesmus sp. PABB004]	KAF8062654.1	GH1	Extracellular	744, 75.
[Scenedesmus sp. PABB004]	KAF8060308.1	GH1 / PDI	Extracellular	2136, 2
[Scenedesmus sp. PABB004]	KAF8059426.1	GH1	Extracellular	1038, 1
S.quadricauda LWG 002611	9833	GH1	Mitochondrion	391, 43.
Scenedesmus obliquus var. DOE0013 v1.0	67487	GH1	Cytoplasm	62.11 k
Scenedesmus obliquus var. DOE0013 v1.0	177517	GH1	Extracellular	579, 61.
Scenedesmus obliquus var. DOE0013 v1.0	746725	GH1	Cytoplasm	513, 57.
Scenedesmus obliquus var. DOE0013 v1.0	797715	GH1	Cell membrane	563, 62.
Scenedesmus obliquus var. DOE0013 v1.0	882370	GH1	Extracellular	496, 54.
Scenedesmus obliquus var. DOE0013 v1.0	1019541	GH1	Cell membrane	255, 24.
EXOCELLULASES				
Scenedesmus obliquus EN0004 v1.0	352887	GH10	Cytoplasm, Soluble	635, 70.
Scenedesmus obliquus EN0004 v1.0	587573	GH10	Cytoplasm, Soluble	849, 93.
Scenedesmus obliquus UTEX B 3031	4623	GH10	Extracellular	498; 55.
Scenedesmus obliquus UTEX B 3031	16336	GH10	Cytoplasm, Soluble	440; 49.

Scenedesmus obliquus UTEX B 3031	10793	GH10	Extracellular	443; 49.
Scenedesmus obliquus var. DOE0013 v1.0	977091	GH10	Cytoplasm, Soluble	393; 44.
Scenedesmus obliquus var. DOE0013 v1.0	243254	GH10	Extracellular	495; 55.
Scenedesmus obliquus var. DOE0013 v1.0	752374	GH10	Cytoplasm, Soluble	839; 92.
Scenedesmus obliquus var. DOE0013 v1.0	750376	GH10	Cytoplasm, Soluble	410; 46.
Scenedesmus PABB0004	KAF8058849.1	GH10	Mitochondrion, Membrane	829; 90.
Scenedesmus quadricauda LGW0026011	547	GH10	Extracellular.	735; 79.

Fig 1

Fig.1 Domain architecture of the putative endoglucanases enzymes in the proteome of Scenedemaceae. The figure shows all predicted proteins containing domains annotated as glycosyl hydrolases in families GH9 (**A**), or GH5 (**B**), in the analyzed strains. Red and yellow squares represent signal peptides and transmembrane domains, respectively. CBM: carbohydrate binding domain. LPMO: Lytic polysaccharide monooxygenases

Fig 2A

2B

Fig.2 **A** GH9 family endoglucanase alignment. Alignment between Scenedesmus GH9 endoglucanases, and Nasta|1KS8_A and Trire|4BMF_A amino acid sequences used as the template for 3D modeling. Red asterisk above the alignment indicates catalytic residues, characterized in Nasta|1KS8. Cyan asterisk indicate cellulose binding residues and purple asterisk indicate conserved cysteines. Other conserved positions are shown in red and boxed in blue. CBM domain is framed in grey. Abbreviations: ScespPA4: *Scenedesmus* sp. PABB004, Scesp1: *Scenedesmus* sp. NREL 46B-D3 v1.0, Sceob1:*Scenedesmus obliquus* UTEX B 3031, SceobDOE: *Scenedesmus obliquus* var. DOE0013 v1.0, SceobE4: *Scenedesmus obliquus* EN0004 v1.0, Scequ2611: *Scenedesmus quadricauda* LWG 002611, Nasta:*Nasutitermes takasagoensis* and Trire: *Trichoderma Reesei* .

B Scequ2611|3068 and Scequ2611|4665 amino acid sequences alignment. Alignment between Scenedesmus GH9 endoglucanases, and two enzymes from close species. Red asterisk above the alignment indicates the catalytic residues. Other conserved positions are shown in red and boxed in blue. Abbreviations: ScespPA4: *Scenedesmus sp.*PABB004, Scesp1: *Scenedesmus sp* . NREL 46B-D3 v1.0, Scequ2611:*Scenedesmus quadricauda* LWG 002611, Monne: *Monoraphidium neglectum* and Rapsu: *Raphidocelis subcapitata*

Fig. 3

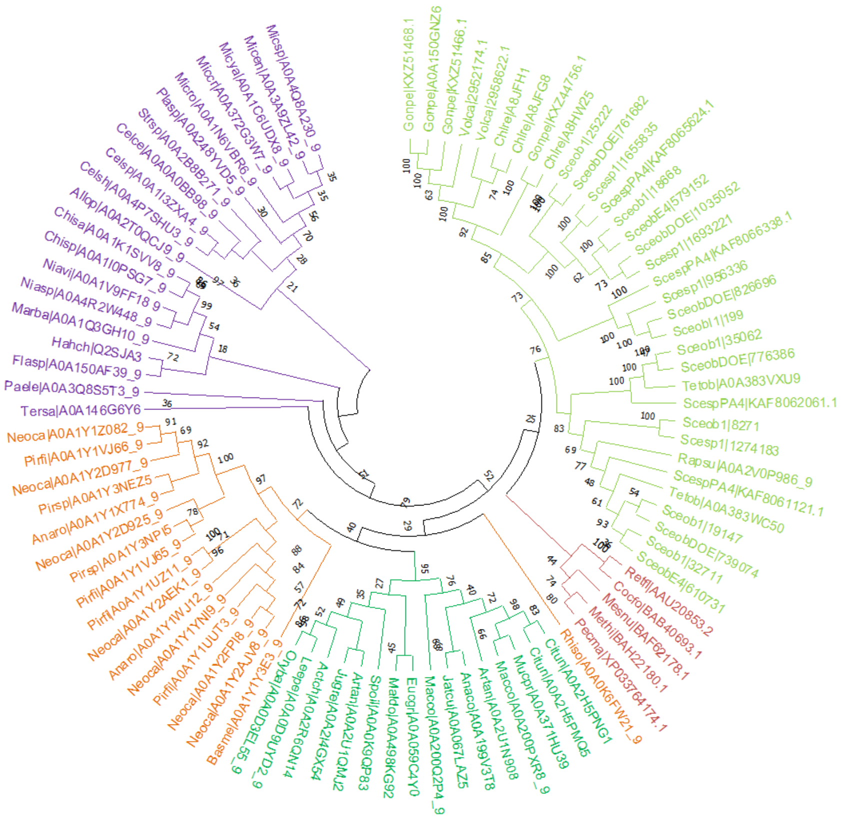

Fig. 3 GH9 family endoglucanase Phylogenetic tree. The tree is based on the alignment of the regions obtained with Gblock and it was built by Maximum Likelihood method in MEGA 6.06 version, under the LG + G model, suggested by MEGA. Phylogenies were determined by Bootstrap Analysis of 100 replicates. Branch lengths are proportional to distances. Bootstrap values are shown above branches. The dark green branches contain GH9 from plants, the lighter green ones correspond to GH9 from algae, the orange ones contain GH9 from fungi, the red ones correspond to invertebrates and those colored purple contain GH9 from bacteria. Abbreviations: Gonpe: *Gonium pectorale,* Volca:*Volvox carteri,* Chlre: *Chlamydomonas reinhardtii,* Sceob1:*Scenedesmus obliquus* UTEX B 3031, SceobDOE: *Scenedesmus obliquus* var. DOE0013 v1.0, Scesp1: *Scenedesmus* sp. NREL 46B-D3 v1.0, ScespPA4: *Scenedesmus* sp. PABB004, SceobE4:*Scenedesmus obliquus* EN0004 v1.0, Scequ2611: *Scenedesmus quadricauda* LWG 002611, Tetob: *Tetradesmus obliquus* , Rapsu:*Raphidocelis subcapitata,* Retfl: *Reticulitermes flavipes* , Copfo: *Coptotermes formosanus* , Mesnu: *Mesocentrotus nudus* , Methi*:* *Metaphire hilgendorfi* , Rhiso: *Rhizoctonia solani* , Citun: *Citrus unshiu* , Mucpr: Mucuna pruriens, Macco:*Macleaya cordata* , Artan: *Artemisia annua* , Anaco:*Ananas communis* , Jatcu: *Jatropha curcas,* Eucgr:*Eucalyptus grandis* , Maldo: *Malus domestica* , Spiol:*Spinacia oleracea* , Actch: *Actinidia chinensis var. chinensis* , Jugre: *Juglans regia* , Leepe: *Leersia perrieri,*Oryba: *Oryza barthii,* Basme: *Basidiobolus meristosporus,*Neoca: *Neocallimastix californiae,* Pirfi: *Piromyces finnis* , Anaro: *Anaeromyces robustus,* Pirsp: *Piromyces sp.*E2, Paele: Paenibacillus lentus, Flasp: Flammeovirga sp. SJP92, Hahch:*Hahella chejuensis* KCTC 2396, Marba: *marine bacterium*AO1-C, Niasp: *Niastella sp* . CF465, Niavi: *Niastella vici,*Chisp: *Chitinophaga sp.* YR573, Chisa: *Chitinophaga sancti* , Allop: *Allonocardiopsis opalescens,* Celsp: *Cellulomonas sp* ., Celsh: *Cellulomonas shaoxiangyii* , Celce: *Cellulomonas cellasea* DSM 20118, Strsp: *Streptomyces sp* . Ru87, Plasp:*Plantactinospora sp.* KBS50, Micro: *Microbispora rosea,*Miccr: *Micromonospora craniellae,* Micya: *Micromonospora yangpuensis* , Micen: *Micromonospora endolithica,* Micsp:*Micromonospora sp.* CNZ295

Fig. 4

Fig. 4 3D modeling of GH9 endoglucanase 32711 from Scenedesmus obliquus UTEX B 3031. **A.** Proposed model of Sceobl1|32711 (pink) superimposed to Nasta|1KS8_A GH9 and Trire| 4BMF_A CBM1 templates (cyan); putative catalytic and binding residues are shown. **B.** Inset of catalytic residues area. Abbreviations: Nasta: *Nasutitermes takasagoensis* and Trire:

Trichoderma Reesei

Fig. 5

Fig. 5.GH5 family endoglucanase alignment. Multiple alignment between Scenedesmus GH5 endoglucanases, and Pyrho|3W6M_A used as the template for 3D modeling. The multiple sequence alignment was performed using Clustal Omega program and ESPript 3.0. Red asterisk above the alignment indicates catalytic residues. Cyan asterisk indicate cellulose binding residues. Other conserved positions are shown in red and boxed in blue. CBM domain is framed in grey. Abbreviations: ScespPA:*Scenedesmus* sp. PABB004, Scesp1: *Scenedesmus* sp. NREL 46B-D3 v1.0, Sceob1: *Scenedesmus obliquus* UTEX B 3031, SceobDOE:*Scenedesmus obliquus* var. DOE0013 v1.0, SceobE4:*Scenedesmus obliquus* EN0004 v1.0, Scequ2611: *Scenedesmus quadricauda* LWG 002611, Pyrho: *Pyrococcus horikoshii*

Fig. 6

Hosted file

`image11.emf` available at https://authorea.com/users/500895/articles/581523-molecular-insight-into-cellulose-degradation-by-the-phototrophic-green-alga-scenedesmus

Fig. 6 GH5 family endoglucanase Phylogenetic tree. The tree is based on the alignment of the regions obtained with Gblock and it was built by Maximum Likelihood method in MEGA 6.06 version, under the WAG + G model suggested by MEGA. Phylogenies were determined by Bootstrap Analysis of 100 replicates. Branch lengths are proportional to distances. Bootstrap values are shown above branches. The dark green branches contain GH5 from plants, the lighter green ones correspond to GH5 from algae, the orange ones contain GH5 from fungi, and those colored purple contain GH5 from bacteria. Abbreviatures: Aciva:*Acidovorax valerianellae,* Xanal: *Xanthomonas albilineans* , Xanca: *Xanthomonas campestris pv. Campestris* , Xanor:*Xanthomonas oryzae pv. Oryzae,* Acici: *Acidovorax citrulli* , Xantr: *Xanthomonas translucens pv. Translucens,* Xylfa*: Xylella fastidiosa,* Acian: *Acidovorax anthurii* , Acisp*: Acidovorax sp. MR-S7* , Canma: *Candidatus Magnetobacterium bavaricum* , Thiba: *Thiotrichales bacterium,* Ulisp:*Uliginosibacterium sp.,* Halsp: *Halothiobacillus sp.,*Massp: *Massilia sp.,* Albte: *Albitalea terrae,* Crepo*: Crenothrix polyspora* , Metis: *Methylomagnum ishizawai* , Hydsp:*Hydrogenophaga sp. A37* , Polbr: *Polyangium brachysporum,*Abypr: *Abyssibacter profundi* , Rapsu: *Raphidocelis subcapitata* , Monne: *Monoraphidium neglectum,* Sceob1:*Scenedesmus obliquus* UTEX B 3031, Scesp1: *Scenedesmus* sp. NREL 46B-D3 v1.0, SceobDOE: *Scenedesmus obliquus* var. DOE0013 v1.0, Tetob: *Tetradesmus obliquus* , SceobE4: *Scenedesmus obliquus* EN0004 v1.0, Kleni: *Klebsormidium nitens,* Neoca*: Neocallimastix californiae,* Pirfi*: Piromyces finnis,*Pirsp*: Piromyces sp.,* Anaro: *Anaeromyces robustus,* Cocsu:*Coccomyxa subellipsoidea,* Chleu: *Chlamydomonas eustigma,*Monsp: *Monosporascus sp* ., Micco: *Micromonas commode,*Glyso: *Glycine soja,* Vigra*: Vigna radiata var. radiata,*Phaan: *Phaseolus angularis,* Arath: *Arabidopsis thaliana,*Braca: *Brassica campestris, Macco: Macleaya cordata*

Fig.7

Fig. 7 3D modeling of GH5 endoglucanase 14060 from Scenedesmus obliquus UTEX B 3031. A. Proposed model of Sceobl1|14060 (pink) superimposed to Pyrho|3W6M_A GH5 and Strco|2RTT_A CBM2 templates (cyan); putative catalytic and binding residues are shown. B. Inset of substrate binding residues area. C. Inset of catalytic residues area. Abbreviations: Strco:*Streptomyces coelicolor* ; Pyrho: *Pyrococcus horikoshii.*

Fig.8

Fig. 8 Domain architecture of the putative β-glucosidases enzymes in the proteome of Scenedemaceae. The figure shows all predicted proteins containing domains annotated as glycosyl hydrolases in families GH1, in the analyzed strains. Red, yellow, magenta and cyan boxes represent signal peptides, transmembrane domains, RAMA domain, and ER-PDI superfamily domain, respectively

Fig. 9A

Hosted file

`image14.emf` available at https://authorea.com/users/500895/articles/581523-molecular-insight-into-cellulose-degradation-by-the-phototrophic-green-alga-scenedesmus

Hosted file

`image15.emf` available at https://authorea.com/users/500895/articles/581523-molecular-insight-into-cellulose-degradation-by-the-phototrophic-green-alga-scenedesmus

9B

Φιγ. 9.A ΓΗ1 φαμιλψ β-γλυςοσιδασε αλιγνμεντ. (ὃλορ ονλινε, δουβλε ςολυμν). Multiple alignment between Scenedesmus GH1 endoglucanases, and Nanoc|5YJ7_A, sequence used as the template for 3D modeling. The multiple sequence alignment was performed using Clustal Omega program and ESPript 3.0. Red asterisk above the alignment indicates the catalytic residues. Abbreviations: ScespPA4:*Scenedesmus sp* . PABB004, Scesp1: *Scenedesmus sp* . NREL 46B-D3 v1.0, Sceob1: *Scenedesmus obliquus* UTEX B 3031, SceobDOE:*Scenedesmus obliquus* var. DOE0013 v1.0, SceobE4:*Scenedesmus obliquus* EN0004 v1.0, Scequ2611: *Scenedesmus quadricauda* LWG 002611, Nanoc: *Nannochloropsis oceanica* .**B. Scequ2611|3544 amino acid sequences alignment.** Alignment between Scenedesmus 3544 putative endoglucanase, and enzymes from close species. The multiple sequence alignment was performed using Clustal Omega program and ESPript 3.0. Red asterisk above the alignment indicates catalytic residues. Abbreviations: ScespPA4: *Scenedesmus sp.* PABB004, Scequ2611: *Scenedesmus quadricauda* LWG 002611, Monne: *Monoraphidium neglectum* and Rapsu: *Raphidocelis subcapitata*

Fig. 10

Hosted file

`image17.emf` available at https://authorea.com/users/500895/articles/581523-molecular-insight-into-cellulose-degradation-by-the-phototrophic-green-alga-scenedesmus

Fig. 10 GH1 family endoglucanase Phylogenetic tree. The tree is based on the alignment of the regions obtained with Gblock and it was built by Maximum Likelihood method in MEGA 6.06 version, under the LG + G model, suggested by MEGA. Phylogenies were determined by Bootstrap Analysis of 100 replicates. Branch lengths are proportional to distances. Bootstrap values are shown above branches. The dark green branches contain GH1 from plants, the lighter green ones correspond to GH1 from algae, the orange ones contain GH1 from fungi, and those colored purple contain GH1 from bacteria. Abbreviations: Sceob1:*Scenedesmus obliquus* UTEX B 3031, SceobDOE: *Scenedesmus obliquus* var. DOE0013 v1.0, Scesp1: *Scenedesmus* sp. NREL 46B-D3 v1.0, ScespPA4: *Scenedesmus* sp. PABB004, SceobE4:*Scenedesmus obliquus* EN0004 v1.0, Scequ2611: *Scenedesmus quadricauda* LWG 002611, Botbo: *Botryobasidium botryosum* , Aspud:*Aspergillus udagawae* , Rasem: *Rasamsonia emersonii* , Exoaq:*Exophiala aquamarine* , Aurna: *Aureobasidium namibiae* , Aurme: *Aureobasidium melanogenum* , Aurpu: *Aureobasidium pullulans* , Baupa: *Baudoinia panamericana* , Aciri:*Acidomyces richmondensis* , Horwe: *Hortaea werneckii* , Dotse: Dothistroma septosporum (strain NZE10 / CBS 128990) (Mycosphaerella pini), Zymtr: *Zymoseptoria tritici* , Cerze: *Cercospora Zeina* , Cerbe: *Cercospora berteroae* , Ceret: *Cercospora beticola* , Calfi: *Caldanaerobius fijiensis* , Actre:*Actinoplanes regularis* , Vulte: *Vulcaniibacterium tengchongense* , Vicva: *Victivallis vadensis* , Rubsp:*Rubrivirga sp.* , Lewag: *Lewinella agarilytica* , Ulvma:*Ulvibacterium marinum* , Arexa: *Arenicella xanthan* , Verba:*Verrucomicrobiae bacterium,* Sacde: *Saccharophagus degradans,* Lenar: *Lentisphaera araneosa,* Halhy:*Haliscomenobacter hydrossis,* Mansp: *Mangrovimonas sp.,*Urecr: *Urechidicola croceus,* Rosmi:

Roseivirga misakiensis, Polsp: *Polaribacter sp.,* Flaba: *Flaviramulus basaltis,* Aquag: *Aquimarina aggregate,* Confl:*Confluentibacter flavum,* Helan: *Helianthus annuus,* Orypu:*Oryza punctate,* Dauca: *Daucus carota subsp. Sativus,*Solly: *Solanum lycopersicum,* Artan: *Artemisia annua,*Oryba: *Oryza barthii,* Braol: *Brassica oleracea var. Oleracea,* Leepe: *Leersia perrieri,* Cinmi: *Cinnamomum micranthum f. kanehirae,* Anaco: *Ananas comosus,* Orypu:*Oryza punctate,* Arahy: *Arachis hypogaea,* Vitvi:*Vitis vinifera,* Cajca: *Cajanus cajan.*

Fig. 11

Φιγ. 11 ΓΗ1 β-γλυςοσιδασε 3Δ μοδελ 3D modeling of GH1 β-glucosidase 3494 from Scenedesmus obliquus UTEX B 3031. **A.**Proposed model of Sceobl1|3494 (pink) superimposed to Nanoc|5YJ7_A GH1 template (cyan); putative catalytic residues are shown. **B.** Inset of catalytic residues area. Abbreviations: Nanoc: *Nannochloropsis oceanica*

Fig. 12 A

B

Hosted file

`image20.emf` available at https://authorea.com/users/500895/articles/581523-molecular-insight-into-cellulose-degradation-by-the-phototrophic-green-alga-scenedesmus

Hosted file

`image21.emf` available at https://authorea.com/users/500895/articles/581523-molecular-insight-into-cellulose-degradation-by-the-phototrophic-green-alga-scenedesmus

Fig. 12A Domain architecture of the putative exoglucanases enzymes in the proteome of Scenedemaceae The figure shows all predicted proteins containing domains annotated as glycosyl hydrolases in families GH10, in the analyzed strains. Red boxes represent signal peptides. Yellow squares represent transmembrane domains

B GH10 family exoglucanase alignment. Multiple alignment between Scenedesmus GH10 exoglucanases, and Celfi|P07986, sequence used as the template for 3D modeling. The multiple sequence alignment was performed using Clustal Omega program and ESPript 3.0. Red asterisk above the alignment indicates catalytic residues. Abbreviations: ScespPA4: *Scenedesmus* sp. PABB004, Scesp1: *Scenedesmus* sp. NREL 46B-D3 v1.0, Sceob1:*Scenedesmus obliquus* UTEX B 3031, SceobDOE: *Scenedesmus obliquus* var. DOE0013 v1.0, SceobE4: *Scenedesmus obliquus* EN0004 v1.0, Scequ2611: *Scenedesmus quadricauda* LWG 002611, Celfi:*Cellulomonas fimi* .

Fig. 13

Hosted file

`image22.emf` available at https://authorea.com/users/500895/articles/581523-molecular-insight-into-cellulose-degradation-by-the-phototrophic-green-alga-scenedesmus

Fig. 13 GH10 family exoglucanase Phylogenetic tree . The tree is based on the alignment of the regions obtained with Gblock and it was built by Maximum Likelihood method in MEGA 6.06 version, under the WAG + G model, suggested by MEGA. Phylogenies were determined by Bootstrap Analysis of 100 replicates. Branch lengths are proportional to distances. Bootstrap values are shown above branches. The dark green branches contain GH10 from plants, the lighter green ones correspond to GH10 from algae, the orange ones contain GH10 from fungi, and those colored purple contain GH10 from bacteria. Abbreviations: Isodo:*Isoptericola dokdonensis,* Isosp: *Isoptericola sp.* Jonde:*Jonesia denitrificans* , Celbo: *Cellulomonas bogoriensis* , Actfe: *Actinotalea fermentans,* Celgi: *Cellulomonas gilvus* , Celsp: *Cellulomonas sp* ., Celbi: *Cellulomonas biazotea* , Celal: *Cellulomonas algicola,* Celfi: *Cellulomonas fimi*

, Micha: *Micromonospora haikouensis* Micsp: *Micromonospora sp* ., Micec: *Micromonospora echinaurantiaca,* Micco:*Micromonospora coxensis,* Celsp: *Cellulomonas sp.* , Celfi:*Cellulomonas flavigena* , Ktera: *Ktedonobacter racemifer* , Actsp: *Actinomadura sp* . GC306, Actda*: Actinomadura darangshiensis* , Stral: *Streptomyces alni,* Thebi:*Thermobispora bispora,* Phach: *Phanerodontia chrysosporium* , Neopa: *Neocallimastix patriciarum,* Arath: *Arabidopsis thaliana,* Nagal: *Naganishia albida,* Gibze: *Gibberella zeae* , Fusox: *Fusarium oxysporum f. sp. lycopersici* ,Aurpu: *Aureobasidium pullulans,* Humgr: *Humicola grisea var. Thermoidea,* Maggr: *Magnaporthe grisea,* Magor: *Magnaporthe oryzae,* Talfu: *Talaromyces funiculosus,* Agabi: *Agaricus bisporus* , Phach: *Phanerodontia chrysosporium,* Hypje:*Hypocrea jecorina* , Ustma: *Ustilago maydis* ,Clapu: *Claviceps purpurea,* Penca: *Penicillium canescens,*Pensi: *Penicillium simplicissimum,* Talpu: *Talaromyces purpureogenus,* Aspac: *Aspergillus aculeatus,* Aspni:*Aspergillus niger,* Aspka: *Aspergillus kawachii* , Rhior:*Rhizopus oryzae,* Pench: *Penicillium chrysogenum,* Theau:*Thermoascus aurantiacus,* Aspte: *Aspergillus terreus,*Aspor: *Aspergillus oryzae* , Neofu: *Neosartorya fumigata* ,Neofi: *Neosartorya fischeri,* Aspcl: *Aspergillus clavatus,* Aspfl: *Aspergillus flavus* , Emeni: *Emericella nidulans* , Aspor: Aspergillus *oryzae* , Aurpu:*Aureobasidium pullulans,* Nagal: *Naganishia albida.*

Fig. 14

Fig. 14 GH10 exoglucanase 3D model. 3D modeling of GH10 exoglucanase 4623 from *Scenedesmus obliquus* UTEX B 3031.**A.** Proposed model of Sceobl1|4623 (pink) superimposed to Celfi|P07986 GH10 template (cyan); putative catalytic residues are shown. **B.** Inset of catalytic residues area. Abbreviations: Celfi: *Cellulomonas fimi*

Supporting information:

Supplementary Figure 1: In a separate file

Data Availability Statement

Not applicable

Conflict of interests

The authors declare no potential financial or other interests that could be perceived to influence the outcomes of the research. No conflicts, informed consent, human or animal rights applicable. All authors declare agreement to authorship and submission of the manuscript for peer review.

Acknowledgements

This work was supported by Grants from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, PIP2015-0476), Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT, PICT2016-0350), and Agencia Santafecina de Ciencia Tecnología e Innovación (ASaCTeI IO2018-00098). JB, MVB and DGC are research members from CONICET. CND is assistant professor from Amity University Uttar Pradesh, Lucknow, India. MBV is a doctoral fellow from CONICET.

Authorship.

JB, MBV, DGC, MVB and CND designed the conception and delineation of the

study; and prepared the manuscript and reviewed it before submission. MBV

performed the *in silico* characterization. JB and MBV performed the acquisition of the data or analyzed such information. All authors read and approved the final manuscript.