Algal Research 34 (2018) 37–48

Contents lists available at ScienceDirect

Algal Research

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

Characterization of plant carbon substrate utilization by *Auxenochlorella protothecoides*

Brian W. Vogler[a,b], Shawn R. Starkenburg[a], Nilusha Sudasinghe[a], Jenna Y. Schambach[a], Joseph A. Rollin[c], Sivakumar Pattathil[d,1], Amanda N. Barry[a,*]

[a] Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87544, United States of America
[b] Department of Chemistry, Colorado School of Mines, Golden, CO 80401, United States of America
[c] National Renewable Energy Laboratory, Golden, CO 80401, United States of America
[d] Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602, United States of America

ARTICLE INFO

Keywords:
Algal biofuel
Mixotrophic
Plant substrate
Genome

ABSTRACT

Algae hold great potential as a source of renewable fuel due to their ability to produce refinery-compatible diesel and jet fuel precursors. Significant effort has been made to maximize productivity under photoautotrophic growth conditions; however, little progress has been made to discover and understand reduced carbon assimilation pathways or enzymatic degradation of complex carbon substrates in algae. We purport that utilization of plant-based carbon substrates in addition to photosynthesis (mixotrophic growth) for biochemical assimilation into biomass, biofuels, and bioproducts, can increase cultivation productivity and improve the economic viability of algal-derived biofuels. Herein we report that a freshwater production strain of microalgae, *Auxenochlorella protothecoides* UTEX 25, is capable of directly degrading and utilizing non-food plant substrates, such as switchgrass, for cell growth. Glycome profiling of plant substrates before and after addition to *A. protothecoides* cultures demonstrates the utilization of xyloglucans. Genomic, proteomic and transcriptomic analyses revealed the identity of many enzymes that are hypothesized to be involved in complex carbohydrate degradation, including several family 5 and 9 glycosyl hydrolases. This work paves the way for future designer engineering of plant-carbon utilization to further improve productivity of algal production strains.

1. Introduction

Auxenochlorella protothecoides UTEX 25, a freshwater microalga, produces large quantities of lipids [1–3], that can be directly converted to fuel. *A. protothecoides* can grow mixotrophically on simple sugars to increase biomass productivity [3], but feeding sugar to algae is not economically viable. Furthermore, adding sugar monomers to open ponds will inevitably cause rampant contamination by heterotrophic organisms. A potential cost-effective and energy-efficient alternative is the utilization of raw or minimally-treated (e.g., acid-treated, heat-treated) lignocellulosic feedstocks: trees, grasses, and agricultural residues [4]. These feedstocks are expected to increase in abundance over time which would further lower cost [5]. Whereas current technology allows these plant substrates to be processed into alcohols with *low* energy density, using these substrates as feedstocks for algae converts them to *high* energy density algal lipids with highly reduced carbon chains compatible with existing transportation fuel infrastructure (e.g., jet fuel, diesel, and gasoline). A recent investigation into hydrothermal liquefaction of blended pine-microalgal feedstocks suggests that blends of lignocellulosic and algal feedstocks may improve overall yields [6]. It is proposed that the synergistic effect of amine chemistry on lignocellulosic biopolymers results in the observed increase in conversion efficiency. This would allow for direct processing of the mixotrophic cultures with downstream production enhancements.

The extent to which algae can degrade the three polymers present in lignocellulose (cellulose, hemicelluloses, and lignin) is not known. A previous study has shown that *Chlamydomonas reinhardtii* is capable of degrading cellulose [7]. However, the degradation of complex lignocellulose has not been explored. Potential glycosyl hydrolases, enzymes that hydrolyze the glycosidic bonds in glycans such as cellulose and hemicelluloses, and their associated domains can be found in the deposited genome and transcriptome sequences of algae; however, no algal glycosyl hydrolases involved in the deconstruction of plant substrates have been identified. In the current study, we present the first

* Corresponding author.
E-mail address: abarry@lanl.gov (A.N. Barry).
[1] Current: Mascoma LLC (Lallemand Inc.), 67 Etna Road, Lebanon NH 03766, United States of America.

https://doi.org/10.1016/j.algal.2018.07.001
Received 26 February 2018; Received in revised form 15 June 2018; Accepted 1 July 2018
Available online 17 July 2018

example of algae degradation and utilization of untreated plant substrate, the putative genetic and molecular mechanism(s) behind this degradation, and identify potential glycosyl hydrolases that may be involved in plant deconstruction.

2. Materials and methods

2.1. Cultivation

Auxenochlorella protothecoides UTEX 25 was obtained from the UTEX Culture Collection of Algae (https://utex.org/). Wild-type *Chlamydomonas reinhardtii* (CC2677) was obtained from *Chlamydomonas* Genetic Center. Cultures were maintained on Sueoka's high salt (HS) media [8] on agar plates under constant light at room temperature. Cultures were replated on HS as necessary to maintain culture viability. For carboxymethylcellulose (CMC) degradation studies, cells were plated on HS media agar plates with 0.1% w/v CMC (Sigma) and grown in constant light or constant dark for 1 week. Culture plates were flooded with 0.2% w/v Congo Red (Sigma) in water, incubated for 20 min, decanted, and then destained with repeated washings of 1 M NaCl to determine the degree of CMC degradation.

In growth with plant substrate experiments, plant matter (raw *Pinacum virgatum* (switchgrass), 25% H_2O_2 pretreated switchgrass, ionic liquid pretreated switchgrass, raw *Eucalyptus grandis* (eucalyptus), ionic liquid pretreated eucalyptus, untreated disintegrated *Betula pendula* (silver birch) wood chips, soda-pulped disintegrated silver birch wood chips, liquid hot water pretreated *Zea mays* (corn) stover, ammonia fiber expansion (AFEX) pretreated eucalyptus top portion, raw *Solidago canadensis* (goldenrod) biomass, and extractive ammonia fiber expansion (EAFEX) pretreated corn stover) (all biomass provided to S. Pattathil courtesy of Idaho National Laboratory and Bruce Dale, Great Lakes Bioenergy Research Center, Michigan State University) was added to HS media. For initial glycomic analysis, triplicate 50 mL cultures of HS media with and without (control) 0.2% w/v plant substrate and with and without (control) *A. protothecoides*, inoculated from a liquid culture at an algae cell OD_{750} of 0.02, were grown at room temperature on a shaker in a 12/12 h light/dark cycle at ~80 μmol/$m^{-2}s^{-1}$ light intensity. Triplicate liquid samples of 50 mL were shipped to the University of Georgia Complex Carbohydrate Research Center for glycome analysis.

Hemicellulose (xylan and xyloglucan) enriched extracts were extracted from plant substrates with 4 M KOH as previously described [9]. For hemicellulose analyses, triplicate 100 mL cultures of HS media with and without (control) 0.01 g xylan extracts of raw switchgrass, ionic liquid pretreated switchgrass, 25% H_2O_2 pretreated switchgrass, raw eucalyptus, and ionic liquid pretreated eucalyptus with and without (control) *A. protothecoides* at a starting OD_{750} of 0.037 were grown in constant dark or constant minimal light (~25 μmol·m^{-2}·s^{-1}) and constant shaking at 22 °C for two weeks. 50 mL of each of these cultures was sent for glycome analyses. The remaining culture volume was centrifuged, and algae pellets were stored at −80 °C for proteomic analyses.

To examine growth with raw switchgrass, triplicate 50 mL cultures of *A. protothecoides* inoculated at around 4×10^5 cells/mL were grown with and without (control) switchgrass (0.2% w/v) at room temperature on a shaker in a 12/12 h light/dark cycle at ~80 μmol·m^{-2}·s^{-1} light intensity. Cell density was determined daily for all samples with a hemocytometer. All experiments were performed in triplicate and the standard deviations and a significance value (*p*-value) of the three biological replicates were determined. After 2 weeks of growth, culture samples were filtered to remove plant debris for metabolite 1D ^{1}H NMR analysis and fatty acid methyl ester (FAME) analysis for determination of algal lipids. Specific growth rate of flask cultures was calculated with the following equation: $\mu = \ln(N_2/N_1)/(t_2 - t_1)$, where μ is the specific growth rate, and N_1 and N_2 are the biomass at time 1 (t_1) and time 2 (t_2), respectively during the exponential growth phase.

For transcriptomic analysis, 50 mL cultures of HS with *A. protothecoides* at OD_{750} of 0.03 were grown in triplicate with and without (control) 0.1% w/v CMC (Sigma) in flasks under constant light or constant dark. After one week of growth, cells were pelleted and stored at −80 °C prior to RNA extraction and sequencing.

2.2. Metabolite 1D ^{1}H NMR

Analysis of cellulose and CMC conversion was conducted on an Avance 400 MHz NMR spectrometer (Bruker). Triplicate samples of HS media with switchgrass or CMC alone (control), algae alone (control), and switchgrass or CMC with algae, were incubated in constant light and shaking for one week. Cells and insoluble biomass were pelleted by centrifugation at 14,000 × *g* for 3 min and the supernatants were diluted by addition of 20% v/v D_2O. Trimethylsilylpropionic acid (Sigma) was used as an internal standard at 1 g/L. Water suppression was conducted using the Watergate W5 pulse sequence with double gradient echo [10], with a d1 relaxation time of 10 s, 4 dummy scans and 8 sample scans.

2.3. FAME analysis

Lipids were determined as FAME by gas chromatography coupled with flame ionization detection (GC/FID) according to Van Wychen et al. [11]. Acid-catalyzed transesterification to quantify the total fatty acids in the biomass was performed by treating 5–10 mg of freeze-dried algae biomass (from triplicate cultures) containing 25 μL of 10 mg/mL methyl tridecanoate (C13:0ME) as the internal standard with 200 μL of chloroform:methanol (2:1, v/v) and 300 μL of 0.6 M HCl:methanol. The samples were heated at 85 °C for 1 h and FAMEs were back extracted with 1 mL of hexane. The hexane extracts were analyzed with an Agilent 7890A Series GC/FID. 2 μL injections at a 10:1 split ratio were loaded onto a DB-WAX column (30 m length × 0.25 mm inner diameter × 0.25 μm film thickness (Agilent Technologies, Santa Clara, CA)). Helium was used as the carrier gas at a flow rate of 1 mL/min. Initial column temperature was 100 °C. Then, the temperature was ramped to 200 °C at 25 °C/min and held for 1 min and again ramped to 242 °C at 1.5 °C/min and held for 1 min (35 min total). The inlet and the detector temperatures were 250 and 280 °C, respectively. Chromatographic signals were matched to a GLC 461C 30-component FAME standard mix (Nu-Chek Prep, Inc., Elysian, MN), and FAME quantification was performed by internal calibration using C13:0ME as the internal standard.

2.4. Glycome profiling

To determine the type of cell-wall glycans in plant substrates before and after *A. protothecoides* growth, glycome profiling was performed as previously described [9]. Briefly, sequential extracts of the Alcohol Insoluble Residues (AIR) from various plant biomass samples was completed using increasingly harsh reagents to isolate cell wall components on the bases of the relative tightness with which they are integrated onto the cell walls followed by screening of these extracts with a comprehensive suite of cell wall glycan directed monoclonal antibodies (mAbs) that can monitor glycan epitopes comprised in most major non-cellulosic matrix glycans. The comprehensive suite of mAbs employed in glycome profiling were procured from laboratory stocks (CCRC, JIM and MAC series) at the Complex Carbohydrate Research Center (available through CarboSource Services; http://www.carbosource.net) or from BioSupplies (Australia) (BG1, LAMP).

2.5. Proteomic analysis

Single pellets (one per treatment) were shipped to Kendrick Labs (Madison, WI) for analysis by 2D gel electrophoresis. Two-dimensional electrophoresis was performed according to the carrier ampholine

method of isoelectric focusing [12, 13]. Isoelectric focusing was carried out in a glass tube of inner diameter 3.3 mm using pH 3–10 Isodalt Servalytes (Serva, Heidelberg, Germany) for 20,000 volt-hrs. 100 ng of an IEF internal standard, tropomyosin, was added to each sample. This protein migrates as a doublet with lower polypeptide spot of MW 33,000 and isoelectric point (pI) 5.2; an arrow on the stained gels marks its position. The tube gel pH gradient plot for this set of Servalytes was determined with a surface pH electrode. After equilibration for 10 min in buffer "O" (10% glycerol, 50 mM dithiothreitol, 2.3% SDS and 0.0625 M tris, pH 6.8), each tube gel was sealed to the top of a stacking gel that overlaid a 10% acrylamide slab gel (1.0 mm thick). SDS slab gel electrophoresis was carried out for about 5 h at 25 mA/gel. The following proteins (Sigma Chemical Co., St. Louis, MO and EMD Millipore, Billerica, MA) were used as molecular weight standards: myosin (220,000), phosphorylase A (94,000), catalase (60,000), actin (43,000), carbonic anhydrase (29,000), and lysozyme (14,000). These standards appear as bands at the basic edge of the silver-stained [14] 10% acrylamide slab gels. The gels were dried between sheets of cellophane paper with the acid edge to the left.

2.5.1. Computerized comparisons

Duplicate gels were obtained from each sample and were scanned with a laser densitometer (Model PDSI, Molecular Dynamics Inc., Sunnyvale, CA). The scanner was checked for linearity prior to scanning with a calibrated Neutral Density Filter Set (Melles Griot, Irvine, CA). The images were analyzed using Progenesis Same Spots software (version 4.5, 2011, TotalLab, UK) and Progenesis PG240 software (version 2006, TotalLab, UK). Computerized analysis for these pairs included image warping followed by spot finding, background subtraction (average on boundary), matching, and quantification in conjunction with detailed manual checking. Spot % was calculated from spot integrated density above background (volume) expressed as a percentage of total density above background of all spots measured. Differential expression was defined as fold-change of the spot percentages. For example, if corresponding protein spots from two different samples (e.g. photoautotrophic versus mixotrophic) have the same spot %, the difference is indicated as 1.0; if the spot % between samples is twice as large as wild type, the difference is 2.0 indicating 2-fold up regulation.

2.5.2. Molecular weight (MW) and isoelectric point (pI) measurements

The isoelectric point (pI) measurements were approximated based on the pH gradient plot for the batch of ampholines for conditions of 9 M urea and room temperature of 22 °C (Supplemental Fig. 1). The molecular weight and pI values for each spot were determined from algorithms applied to the reference image.

2.5.3. Protein digestion and peptide extraction

Spots 211, 263, 391, 466, 68, 433, 572, and 576 were selected for excision and sequencing in the lab of Dr. Costel Darie, Clarkson University. Proteins that were separated by SDS-PAGE/2D-PAGE and stained by Coomassie dye were excised, washed and the proteins from the gel were treated according to published protocols [15–17]. Briefly, the gel pieces were washed in high purity, high performance liquid chromatography (HPLC) grade water, dehydrated and cut into small pieces and destained by incubating in 50 mM ammonium bicarbonate, 50 mM ammonium bicarbonate/50% acetonitrile, and 100% acetonitrile under moderate shaking, followed by drying in a speed-vac concentrator. The gel bands were then rehydrated with 50 mM ammonium bicarbonate. The procedure was repeated twice. The gel bands were then rehydrated in 50 mM ammonium bicarbonate containing 10 mM dithiothreitol (DTT) and incubated at 56 °C for 45 min. The DTT solution was then replaced by 50 mM ammonium bicarbonate containing 100 mM Iodoacetamide for 45 min in the dark, with occasional vortexing. The gel pieces were then re-incubated in 50 mM ammonium bicarbonate/50% acetonitrile, and 100% acetonitrile under moderate shaking, followed by drying in speed-vac concentrator. The dry gel pieces were then rehydrated using 50 mM ammonium bicarbonate containing 10 ng/μL trypsin and incubated overnight at 37 °C under low shaking. The resulting peptides were extracted twice with 5% formic acid/50 mM ammonium bicarbonate/50% acetonitrile and once with 100% acetonitrile under moderate shaking. Peptide mixture was then dried in a speed-vac, solubilized in 20 μL of 0.1% formic acid/2% acetonitrile.

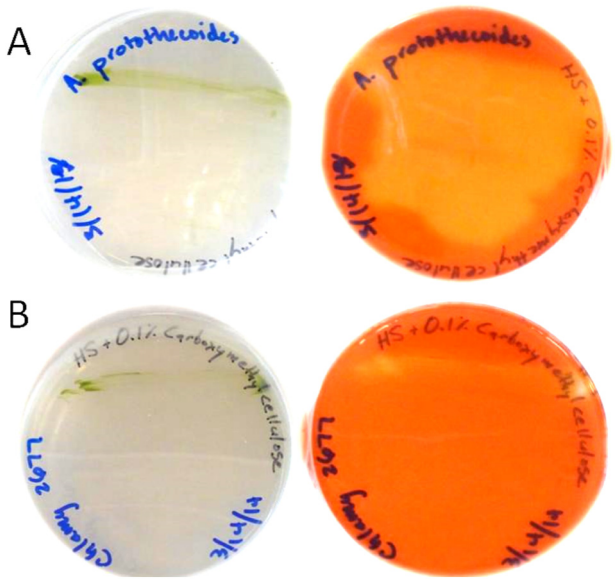

Fig. 1. Algae utilization of carboxymethylcellulose (CMC) in the absence of light. Reduction of the red stain indicates CMC consumption. Prestained and stained plates of (A) *A. protothecoides* and (B) *Chlamydomonas reinhardtii* are shown.

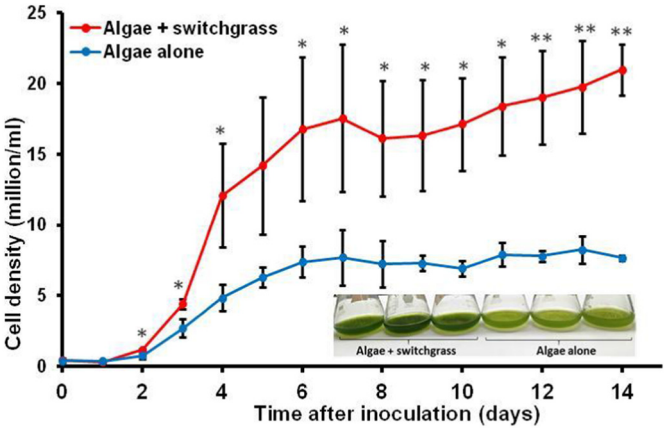

Fig. 2. Comparison of *A. protothecoides* growth with and without augmentation with raw switchgrass (0.2% w/v). Cultures were grown with a 12:12 light:dark cycle at ~80 μmol·m^{-2}·s^{-1} light with constant shaking at room temperature for two weeks. All experiments were performed in triplicate and the error bars represent the standard deviations of the three biological replicates. Significant differences in cell density are marked with asterisks (* for $p < 0.05$ and ** for $p < 0.01$). Triplicate cultures at 14 days are shown in the inset.

2.5.4. Nanoliquid chromatography and mass spectrometry (LC-MS/MS)

The peptides mixture was analyzed by reversed-phase nanoliquid chromatography (LC) and MS (LC-MS/MS) using a NanoAcuity UPLC (Micromass/Waters, Milford, MA) coupled to a Q-TOF Xevo G2 mass spectrometer (Micromass/Waters, Milford, MA), according to published procedures [16, 18, 19]. Briefly, the peptides were loaded onto a 100 μm × 10 mm NanoAquity BEH130 C18 1.7 μm UPLC column (Waters, Milford, MA) and eluted over a 60-min gradient of 2–80%

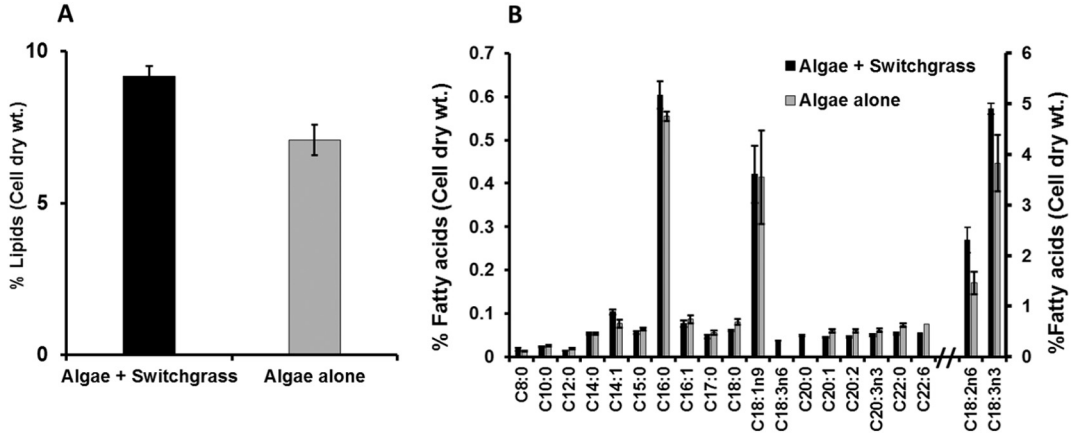

Fig. 3. Lipid contents (A) and fatty acid profiles (B) by GC/FID for *A. protothecoides* biomass grown with and without switchgrass. Data shown as the mean of the triplicate growth experiments and the error bars represent the standard deviations of the three biological replicates.

organic solvent (acetonitrile containing 0.1% formic acid (FA)) at a flow rate of 400 nL/min. The aqueous solvent was 0.1% FA in HPLC water. The column was coupled to a Picotip Emitter Silicatip nano-electrospray needle (New Objective, Woburn, MA). MS data acquisition involved survey MS scans and automatic data dependent analysis (DDA) of the top six ions with the highest intensity ions with the charge of 2+, 3+ or 4+. The MS/MS was triggered when the MS signal intensity exceeded 250 counts/s. In survey MS scans, the three most intense peaks were selected for collision-induced dissociation (CID) and fragmented until the total MS/MS ion counts reached 10,000 or for up to 6 s each. The entire procedure used was previously described [16, 18, 19]. Calibration was performed for both precursor and product ions using 1 pmol GluFib (Glu1-Fibrinopeptide B) standard peptide with the sequence EGVNDNEEGFFSAR and the monoisotopic doubly-charged peak with *m/z* of 785.84.

2.5.5. Data processing and protein identification

The raw data were processed using ProteinLynx Global Server (PLGS, version 2.4) software as previously described [18]. The following parameters were used: background subtraction of polynomial order 5 adaptive with a threshold of 30%, two smoothings with a window of three channels in Savitzky-Golay mode and centroid calculation of top 80% of peaks based on a minimum peak width of 4 channels at half height. The resulting pkl files were submitted for database search and protein identification to the in-house Mascot server (www.matrixscience.com, Matrix Science, London, UK) for database search using the following parameters: databases from NCBI (Plants), parent mass error of 0.5 Da with 1 ^{13}C, product ion error of 0.8 Da, enzyme used: trypsin, three missed cleavages, propionamide as cysteine fixed modification and methionine oxidized as variable modification. To identify the false negative results, we used additional parameters such as different databases or organisms, a narrower error window for the parent mass error (1.2 and then 0.2 Da) and for the product ion error (0.6 Da), and up to two missed cleavage sites for trypsin. In addition, the pkl files were also searched against in-house PLGS, version 2.4 (www.waters.com) using searching parameters similar to the ones used for Mascot search. The Mascot and PLGS database search provided a list of proteins for each gel band. To eliminate false positive results, for the proteins identified by either one peptide or a mascot score lower than 25, we verified the MS/MS spectra that led to identification of a protein.

2.6. Genomic sequencing, assembly, and annotation

The *A. protothecoides* UTEX 25 genome was sequenced using a combination of Illumina HiSeq 2000 [20] and 454 sequencing technologies [21]. A 1 × 100 base-pair Illumina shotgun library was prepared using standard TruSeq protocols and sequenced from bulk *A. protothecoides* genomic DNA on an Illumina GAII sequencer to generate 2 billion reads. Additional shotgun single-end and paired-end (11-kb insert) DNA libraries were prepared for sequencing on the 454 Titanium platform, which resulted in the generation of 1.16 and 1.15 million reads, respectively. The 454 single-end data and the 454 paired-end data were assembled together using Newbler (version 2.3, release 091027_1459). The Illumina-generated sequences were assembled separately with VELVET (version 1.0.13) [22]. The resulting consensus sequences from both the VELVET and Newbler assemblies were computationally shredded into 10-kb fragments and re-assembled with reads from the 454 paired-end library using parallel phrap (version 1.080812, High-Performance Software, LLC) [23]. The mitochondrial genome was identified in this final hybrid assembly using homologous blast searches against *Chlorella variabilis* and *Chlorella vulgaris* mitochondrial genomes. Misassemblies were corrected using Dupfinisher [24], and repeat resolution was performed in Consed to generate the final consensus sequences [25].

The resulting assembly was annotated using MAKER version 2.32.8_5.24 [26]. De novo assembled transcripts from all functional studies (see Transcriptome Analysis), and protein homology evidence from a closely related strain of *A. protothecoides* (Genbank Accession #GCF_000733215.1) were used to train and constrain structural gene models. All predicted genes were functionally annotated using InterProScan (version 5.21–60.0) by conducting homology searches against proteins in IntroPro, Uniprot and Swiss-Prot database using Blastp (version NCBI 2.2.28+). The complete genome sequences of the *A. protothecoides* UTEX 25 nuclear and chloroplast genomes have been deposited into GenBank under the accession number QOKY01000000. The annotation is publicly available at https://greenhouse.lanl.gov/. The mitochondrial genome was previously deposited under accession number KC631634.1. The general characteristics of the assembly and annotation are described in Supplemental Table 1.

From the genome sequence of *A. protothecoides* UTEX 25 and *Auxenochlorella sp.* 0710, thirty potential proteins with homology to the glycosyl hydrolase (GH) family of proteins were identified [27]. The identified proteins were scanned for conserved domains using NCBI's conserved domain database [28] to identify the GH family and were analyzed using TargetP 1.1 to predict subcellular localization [29]. The glycosyl hydrolases of families 5 and 9 were each used to search the characterized members of these families available on the Carbohydrate-Active enZYmes Database (CAZY; www.cazy.org), to identify and verify the specific reactions catalyzed. Using PhyML 3.2 with the Le & Gascuel substitution model and 100 bootstraps, we estimated the maximum likelihood phylogenies of these GH5 and GH9 sequences from the

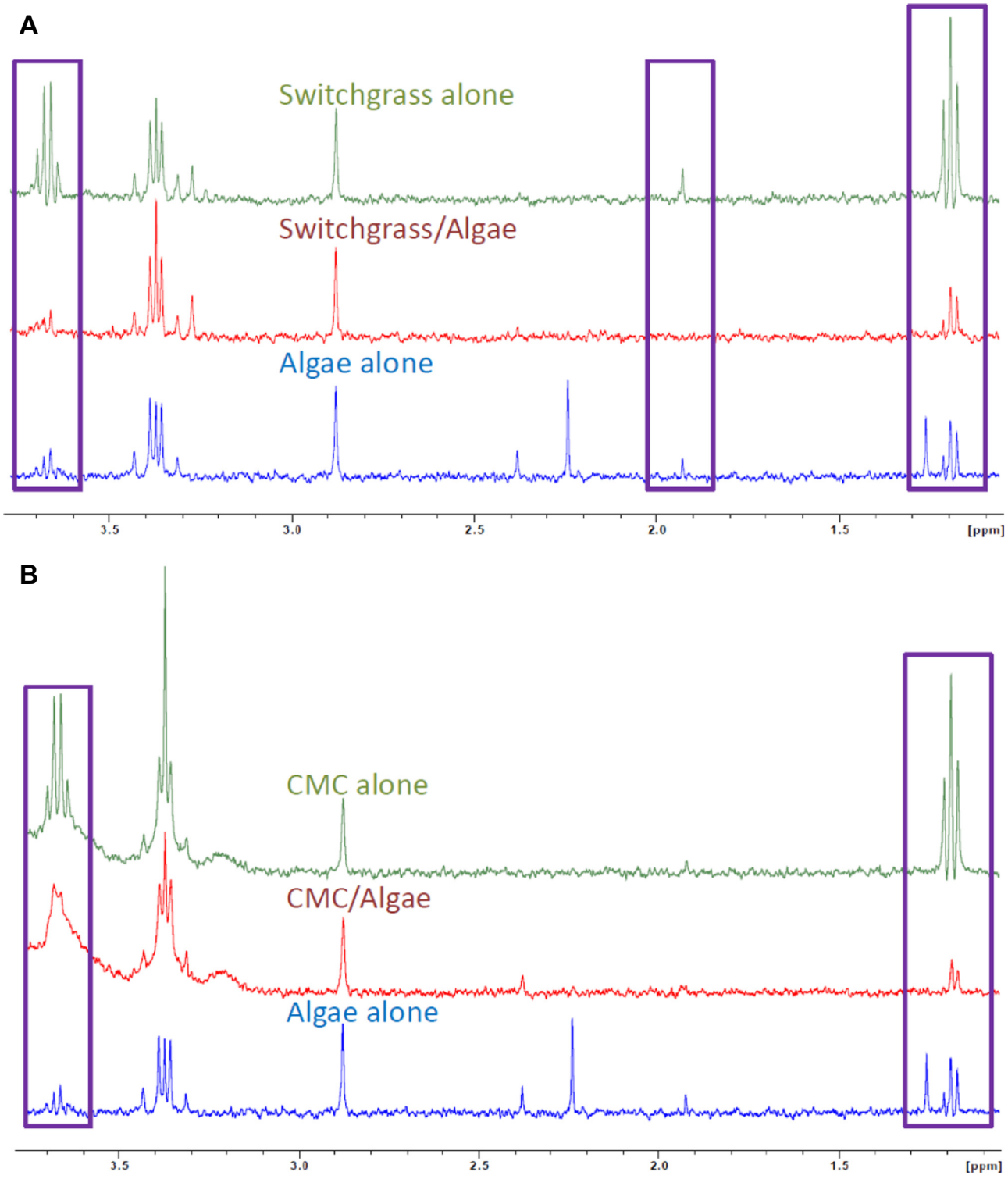

Fig. 4. Metabolite 1D ^{1}H NMR analysis of *A. protothecoides* cultures. HS culture medium was amended with (A) switchgrass or (B) CMC in the presence and absence of algae. Representative spectra from each treatment are shown.

predicted protein sequences [30].

2.7. Transcriptomic sequencing and analysis

As described above, cultures were grown in triplicate with and without (control) 0.1% CMC in HS media in flasks in both constant light and constant dark. In all cases, after five days, total RNA was extracted from 100 mg of cells within the algae pellet using the Direct-zol RNA-miniprep kit (ZYMO, P/N 2051) according to the manufacturer's instructions. Each total RNA sample was enriched for mRNA by hybridizing the poly(A) tail to oligo $d(T)_{25}$ probes covalently coupled to magnetic beads, followed by elution (NEB, P/N S1419S). The enriched mRNA fractions were prepared for Illumina sequencing using the ScriptSeq V.2 RNA-seq Library Preparation Kit (Epicentre, P/N SSV21106). The resulting sequencing library for each sample was denatured to a final concentration of 1.5 pM before being loaded onto the

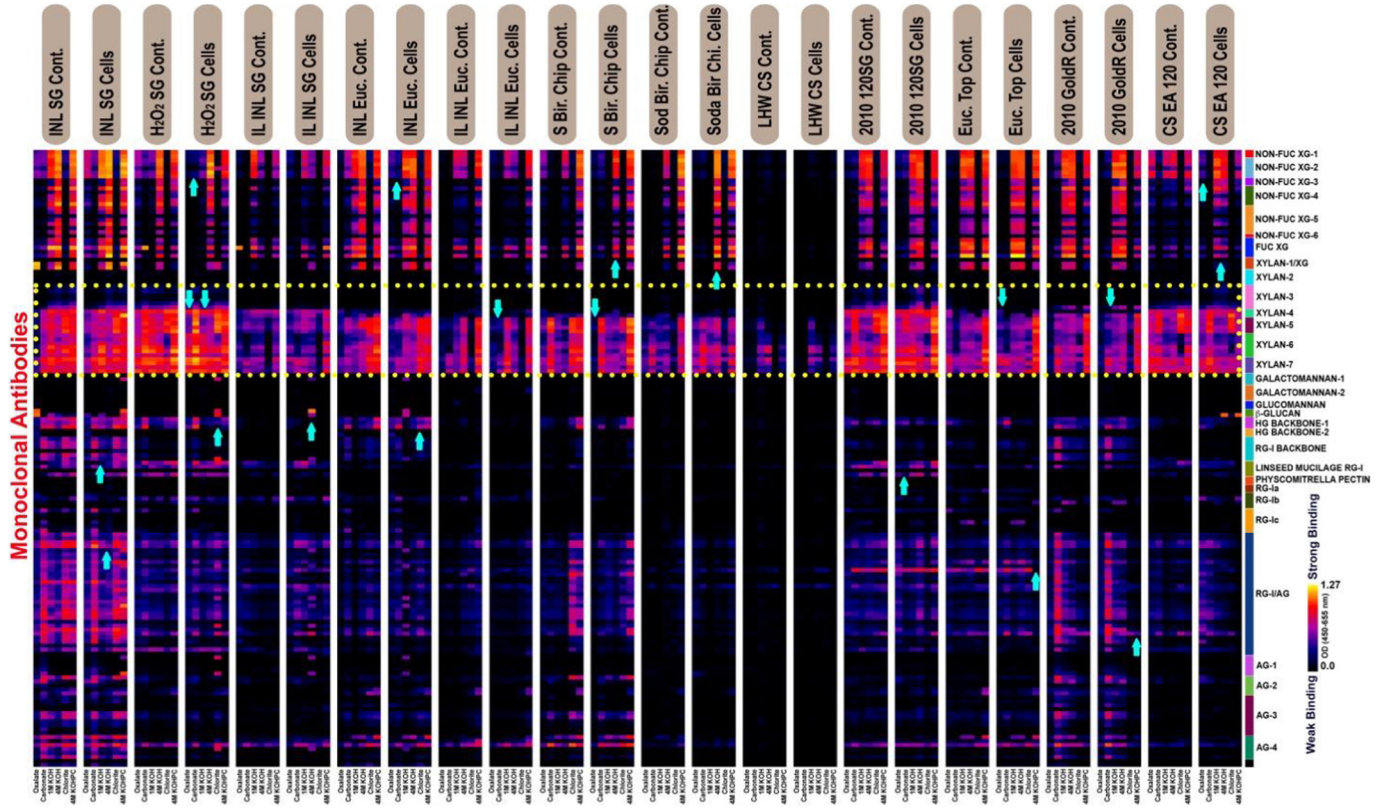

Fig. 5. Glycome profiling of various plant ligno-cellulosic substrates incubated in media with or without *A. protothecoides* for 2 weeks. Glycome profiling of the alcohol insoluble residues isolated from various plant biomass substrates were conducted as explained in Materials and methods section. The substrates used were: switchgrass (INL SG), 25% H_2O_2 pretreated switchgrass (H_2O_2 SG), ionic liquid pretreated switchgrass (IL INL SG), raw eucalyptus (INL Euc), ionic liquid pretreated eucalyptus (IL INL Euc), untreated disintegrated silver birch wood chips (S. Bir. Chip), soda-pulped disintegrated silver birch wood chips (Sod Bir. Chip), liquid hot water pretreated corn stover (LHW CS), AFEX pretreated eucalyptus top portion (Euc. Top), Goldenrod raw biomass (2010 GoldR), EAFEX pretreated corn stover (CS EA 120). Controls, indicated as "cont." represent media plus substrates without *A. protothecoides* while "cells" represent media plus substrates with *A. protothecoides*. The panel on the right indicates groupings of monoclonal antibodies (mAbs) based on their glycan specificity. Sequential extracts are depicted at the bottom panel. The binding strengths of glycan directed mAbs are expressed as a heatmap with color scheme, brightest yellow-red and dark blue colors depicts strongest and weakest binding, respectively. Glycan epitope regions of evidence for carbohydrate utilization/modification by algal growth are indicated by light blue arrows and the xylan region is shown in yellow dotted block. A representative sample from each treatment is shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

NextSeq Reagent Cartridge for a paired-end sequencing on a Nextseq 500 (2 × 150 bp), multiplexed at 6 samples per lane.

The recovered forward and reverse 2 × 151 bp Nextseq reads were trimmed for quality using FastQ Quality Control Software [31]. All reads with a minimum sliding window q-score of 15, and a minimum trimmed read length of > 29 base pairs were kept for mapping and differential expression analysis. Quality filtered reads were mapped with Bowtie2 using the default settings, per the manual. Abundance estimation (RPKM; fragments per kilobase of exon per million mapped reads) was calculated using RSEM. Differential expression of transcripts was determined with EdgeR. Transcripts that were differentially expressed > 2-fold with a p-value < 0.001 were considered statistically significant.

3. Results and discussion

3.1. Degradation of cellulose and plant substrates

To examine the degradation of cellulose by *A. protothecoides* UTEX 25, we grew *A. protothecoides* on CMC-containing agar plates kept under constant light (Supplemental Fig. 2) or constant dark (Fig. 1). Based on the clearing of Congo Red staining of the cellulose in the plate kept in the dark, *A. protothecoides* appears not only to actively degrade cellulose, but to do so more than *C. reinhardtii* (Fig. 1).

The diffuse pattern of degradation on the plate suggests that the cellulolytic enzyme is secreted from the organism. To explore the impact of plant substrate degradation by *A. protothecoides*, we examined the growth of the organism on raw switchgrass. By measuring cell counts over time, we observed a > 40% increase in growth rate over days 1–5 in the presence of switchgrass ($\mu = 1.22 \pm 0.32$ vs. $0.77 \pm 0.34\ day^{-1}$), resulting in an average of > 140% increase in cell count at stationary phase (days 9–14) (Fig. 2; Supplementary Fig. 2). The variability in cell density observed for switchgrass cultures was largely the result of one switchgrass culture appearing to lag heavily behind the others, likely due to the unavoidable heterogeneity of the untreated switchgrass substrate.

To determine the impact of mixotrophic growth with switchgrass on lipid accumulation, we examined the total algal lipids by FAME analysis. *A. protothecoides* grown with switchgrass show higher lipid accumulation compared to the cultures grown without (9% vs. 7%; Fig. 3A). Relatively lower lipid contents observed for *A. protothecoides* in the present study could possibly due to growing the cultures with no exogenous carbon dioxide supply. Increasing lipid concentration by supplementation of organic carbon in *A. protothecoides* and other *Chlorella* sp. in either heterotrophic or mixotrophic culturing has been observed previously [32–34], but this study is the first we are aware of that utilizes raw plant matter as the carbon source. Fatty acid profiles for the cultures grown with and without switchgrass were compared to examine the impact of the carbon source on the fatty acid distribution (Fig. 3B). Biomass generated from the cultures grown with switchgrass

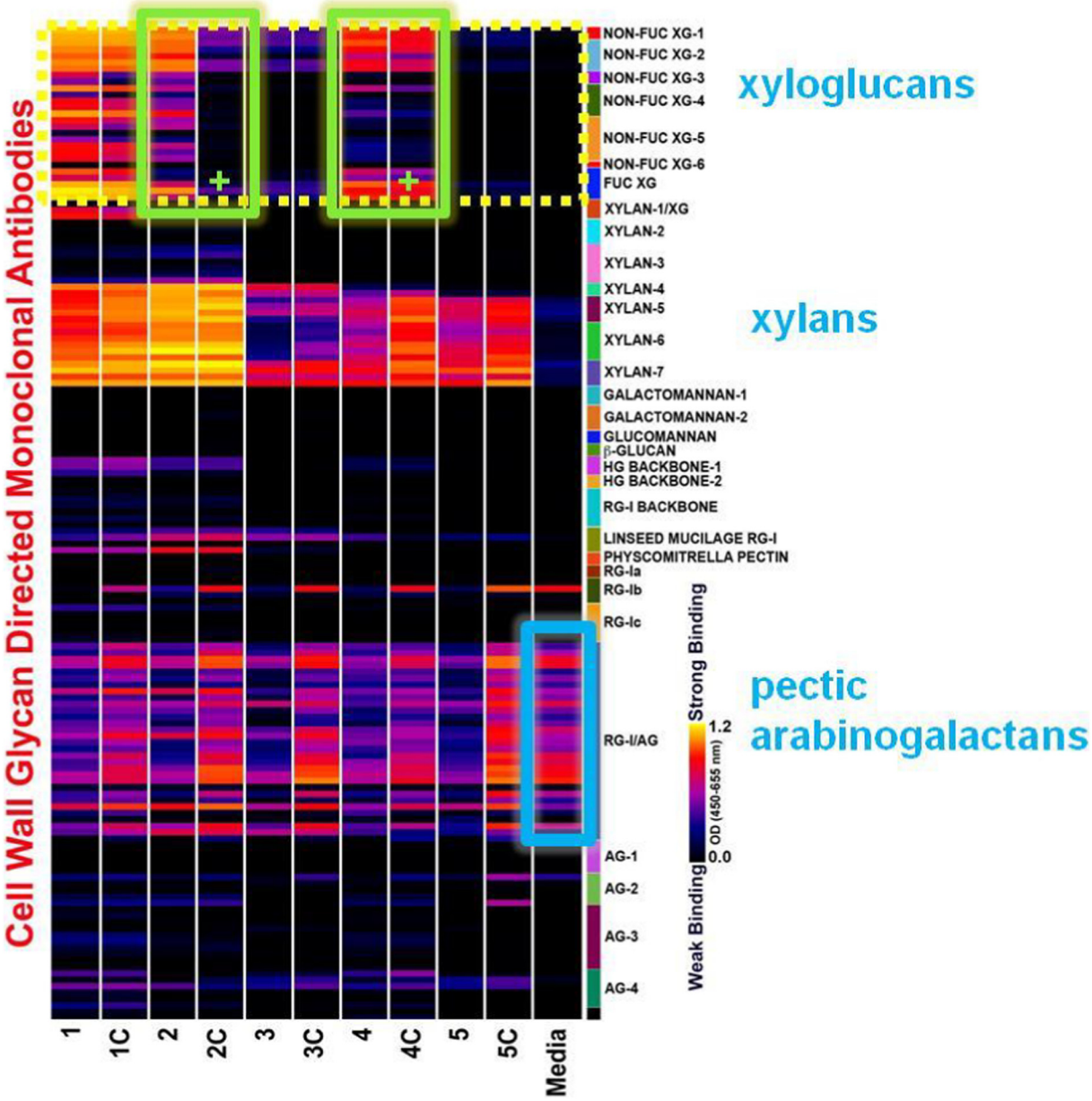

Fig. 6. Cell wall glycan-directed mAb screening of cell wall extracts in culture media incubated without (1–5) and with (1C-5C) algae. 4 M KOH extracts were isolated from raw switchgrass (1), ionic liquid pre-treated switchgrass (2), 25% H_2O_2-washed switchgrass (3), raw eucalyptus (4), and ionic-liquid pre-treated eucalyptus (5) to supplement with media for incubating with and without *A. protothecoides*. Lighted bars represent presence of sugar. Green-highlighted boxes show levels of xyloglucans are vastly reduced when algae is present (+). Blue-highlighted box shows algae produces a valuable coproduct arabinogalactan. A representative sample from each treatment is shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

showed slightly higher levels of C14:1, C16:0, C18:2n6 and C18:3n3 fatty acids compared to the cultures grown without switchgrass. Additionally, C18:3n6 and C20:0 are only observed in the biomass generated from the cultures grown in the presence of switchgrass.

To observe whether or not plant substrate structure was modified by *A. protothecoides*, we performed metabolite 1D ^{1}H NMR on cultures with substrate (switchgrass or CMC) alone, with algae alone, and with both substrate and algae. By examining metabolite 1D ^{1}H NMR, we observe a diminishing in peaks associated with single-hydrogen C—H bonds (0.8–1.5 ppm) and methylene bridge protons in cellulose (3.5–4.0 ppm) [35], when comparing cultures with switchgrass alone to cultures where algae and switchgrass are grown together (Fig. 4A). We see these same peaks disappearing with cultures grown with CMC with and without algae (Fig. 4B), indicating a similar mechanism of carbon utilization with both substrates.

With each substrate, the signal attributed to single-hydrogen C—H bonds from algae cultivated with substrate is less than or equal to the signals from either substrate or algae alone, when traces are roughly normalized. This allows the possibility that the substrate single-hydrogen C—H bonds are completely or almost completely incorporated into the algae, as we would otherwise expect a signal intensity between the two (algae alone and substrate alone). We also note two peaks, at 1.27 and 2.22 ppm, which are present in algae alone, but that disappear when the algae are cultured with either substrate. These phenomena are more difficult to explain, but may be the result of a chemical produced by the cells only in the absence of substrate. Alternatively, this chemical may be altered or degraded by interaction with our substrates.

To assess the range of non-cellulosic plant cell wall glycans (including most major hemicelluloses and pectic-polysaccharides) that are utilized or modified by *A. protothecoides*, we profiled its glycome (using

Table 1
A catalog of the potential GH-encoding genes from *A. protothecoides* sp. 0710 and their UTEX 25 homologs. Included are the glycosyl hydrolase families identified by the Conserved Domain Database and the respective E-value. From the Transcriptome Analysis, the fold-change (FC) (presented as a mean of triplicate samples) of each transcript is given during the dark and light periods in the presence of CMC (CMC Dark and CMC Light, respectively), relative to the unsupplemented culture, as well as *p*-values. Significant differences ($p < 0.01$) are highlighted in bold.

Gene ID		CDD GH Hit		Transcriptome Analysis			
				CMC Dark		CMC Light	
UTEX 25	sp. 0710	Family	E-value	*FC*	*p*-Value	*FC*	*p*-Value
APUTEX2500001730	KFM23057.1	GH2	0.0E+000	1.0	7.84E−01	0.9	4.40E−01
APUTEX2500000235	KFM29315.1	GH5	1.9E−020	0.5	**1.59E−05**	1.1	5.56E−01
APUTEX2500000742	KFM23002.1	GH5	7.1E−020	1.2	2.31E−01	1.3	9.65E−02
APUTEX2500000844	KFM25716.1	GH5	1.3E−022	1.3	6.66E−02	1.2	2.06E−01
APUTEX2500000985	KFM28016.1	GH5	3.5E−017	0.2	**8.26E−10**	1.2	6.12E−01
APUTEX2500002156	KFM22632.1	GH5	4.0E−025	1.0	8.95E−01	1.1	4.56E−01
APUTEX2500002671	KFM29130.1	GH5	2.7E−021	1.2	2.55E−01	1.1	6.37E−01
APUTEX2500003890	KFM28332.1	GH5	7.4E−013	0.9	7.76E−01	1.0	7.96E−01
APUTEX2500004827	KFM23138.1	GH5	8.2E−021	1.6	**7.48E−03**	0.9	7.69E−01
APUTEX2500000468	KFM25061.1	GH9	2.9E−084	1.1	6.18E−01	1.0	7.91E−01
APUTEX2500000981	KFM28012.1	GH9	3.4E−050	1.4	2.17E−02	0.9	6.95E−01
APUTEX2500003119	KFM24581.1	GH9	4.6E−098	1.0	9.80E−01	1.2	2.00E−01
APUTEX2500005369-A	KFM24183.1	GH9	1.1E−113	1.8	**2.09E−03**	1.3	1.70E−01
APUTEX2500005369-B	KFM24184.1	GH9	9.2E−104	0.9	6.35E−01	0.9	5.24E−01
APUTEX2500005369-C	KFM24182.1	GH9	2.8E−100	1.4	3.50E−02	0.9	4.54E−01
APUTEX2500005777	KFM24064.1	GH9	3.6E−067	1.4	1.07E−02	1.1	4.89E−01
APUTEX2500000143	KFM24770.1	GH16	3.0E−065	1.2	3.73E−01	1.2	3.66E−01
APUTEX2500002026	KFM25245.1	GH18	6.8E−009	1.5	**9.56E−03**	0.9	7.15E−01
APUTEX2500002706	KFM22671.1	GH20	2.9E−104	1.5	1.73E−01	1.0	1.00E+00
APUTEX2500005097	KFM28226.1	GH27	8.8E−128	0.7	5.57E−02	1.2	2.21E−01
APUTEX2500001996	KFM25289.1	GH31	0.0E+000	0.7	5.00E−02	1.0	1.00E+00
APUTEX2500003474	KFM25547.1	GH31	0.0E+000	0.5	**4.20E−04**	1.0	9.63E−01
APUTEX2500000041	KFM23987.1	GH32	1.7E−067	1.6	2.85E−02	1.1	5.19E−01
APUTEX2500000101	KFM28270.1	GH38	6.6E−065	0.9	6.64E−01	1.0	8.29E−01
APUTEX2500005781	KFM24068.1	GH43	3.9E−015	1.1	5.89E−01	0.7	1.02E−01
APUTEX2500000370	KFM24944.1	GH43	2.3E−005	1.1	4.15E−01	1.0	8.85E−01
APUTEX2500000577	KFM26117.1	GH47	3.7E−172	0.9	6.71E−01	1.2	2.64E−01
APUTEX2500002068	KFM28922.1	GH47	6.1E−111	1.0	7.84E−01	1.0	8.47E−01
APUTEX2500004355	KFM22865.1	GH47	0.0E+000	1.1	4.19E−01	1.1	4.14E−01
APUTEX2500001195	KFM26337.1	GH63	0.0E+000	1.4	1.45E−02	1.1	6.12E−01

a suite of monoclonal antibodies raised against most major matrix non-cellulosic plant cell wall glycans) in cultures amended with a panel of 12 untreated or pretreated plant biomass substrates. After active culturing, the abundance of specific cell wall glycan epitopes in the cell wall extracts from plant substrates varied as a result of the algal growth (Fig. 5; light blue arrows in yellow dotted blocks).

Algal growth specifically altered the abundance of epitopes contained in hemicellulose glycans like non-fucosylated and fucosylated xyloglucans, substituted and unsubstituted xylan. Changes were also noted in pectic backbone glycan epitopes including homogalacturonan and rhamnogalacturonan backbones, and in the abundance of some pectic arabinogalactan epitopes. In general, the depletion of these epitopes was noted in a majority of the biomass tested. Further, in most switchgrass biomass substrates studied, non-cellulosic glycan epitopes generally showed marginally increased extractability among plant biomass substrates incubated with algae indicating that *A. protothecoides* growth induced modifications in the plant cell wall structure, causing changes in the extractability of epitopes, including those of xylans, xyloglucans, pectic polysaccharides, glucomannan and β-glucan.

To further examine which specific glycans are utilized, we grew *A. protothecoides* on specific plant carbohydrate extracts isolated from switchgrass and eucalyptus (4 M KOH extracts enriched in hemicelluloses, such as xylans and xyloglucans). When we repeated mAb-based screening that included hemicellulose-specific antibodies, we found that the predominant components utilized by the algae cells were xyloglucans (Fig. 6), suggesting that *A. protothecoides* actively expresses one or more xyloglucanases, enzymes that facilitates the hydrolysis of cellulose in wood. This finding is consistent with the fact that this organism was first isolated from woody material [36]. Unexpectedly and outside the scope of this study, mAb-based screening also revealed the presence of pectic arabinogalactans in the *A. protothecoides* cell wall, a valuable thickening agent for food [37] (Fig. 6).

3.2. Genome analysis

To determine the genetic basis of the observed xyloglucanase activity, we sequenced, assembled and annotated the genome of *A. protothecoides* UTEX 25. Homology searches with publicly available xyloglucanase protein and transcript sequences identified at least 30 potential proteins in *A. protothecoides* UTEX 25 with homology to the glycosyl hydrolase (GH) family of proteins (Table 1).

All of these proteins are also conserved in the genome of a closely related strain of *A. protothecoides* [27]. Proteins involved in ER lipid and glycoprotein shuttling and synthesis (glycosyl families 16, 31, and 32) were identified along with three GH family 47 homologs that potentially function as alpha-mannosidases. Other GH or GH-like proteins identified include proteins homologous to GH family 20 protein (a potential hexosamidase), GH family 38 (a potential alpha-mannosidase), and GH family 63, potentially involved in the transport of phosphorylated mannosylglycerate within the cell, with a C-terminal adenylosuccinate lyase. Furthermore, a GH family 18 homolog was identified that contains a D-X-X-D-X-D-X-Q motif. Whereas chitinase members of the GH18 family contain a D-X-X-D-X-D-X-E motif, the Glu to Gln substitution is indicative of xylanase inhibitors [38]. Two of these potential glycosyl hydrolases include sequences with high homology (E-value $< 10^{-30}$) to peptidases, likely involved in other cellular processes.

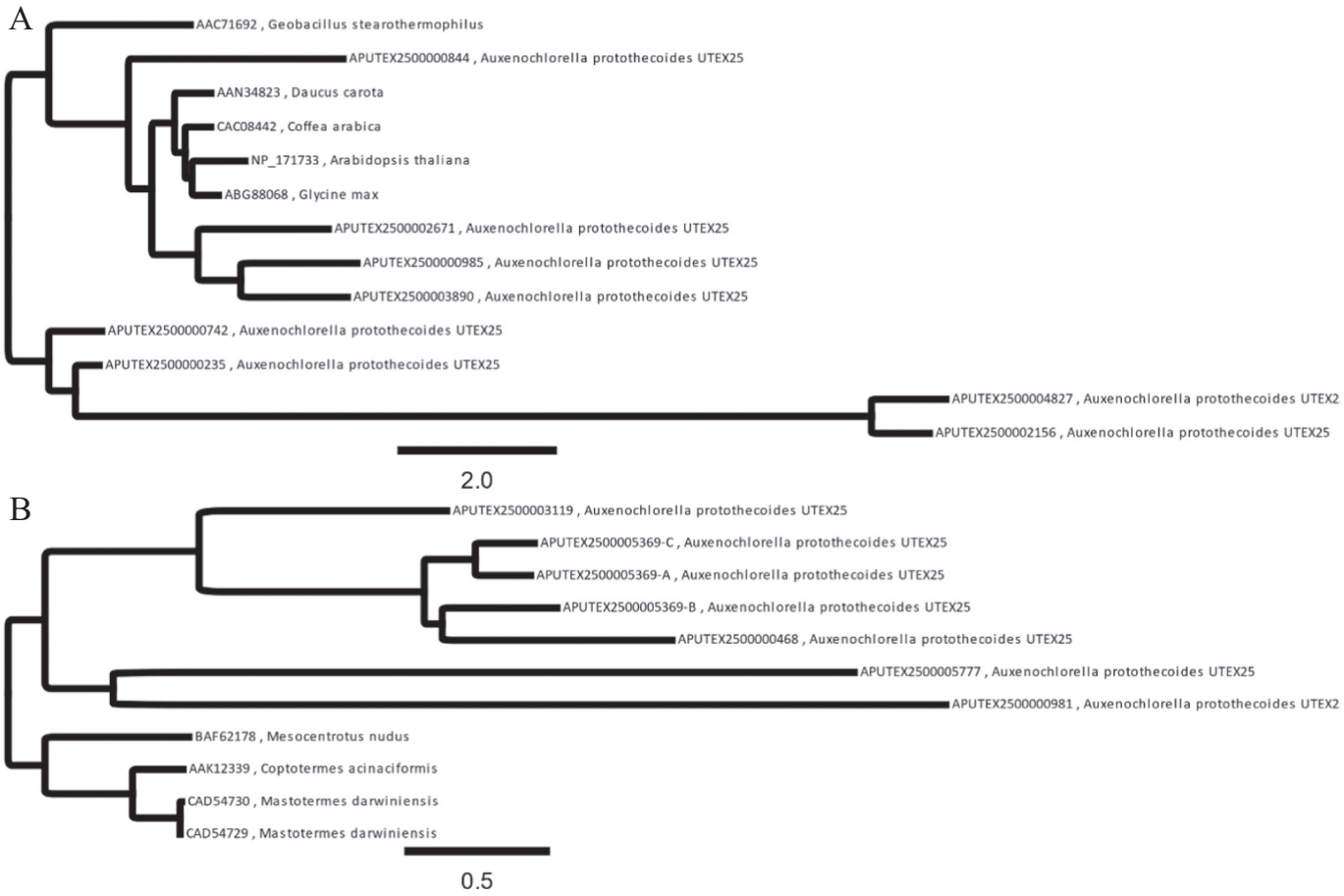

Fig. 7. Phylogenetic analysis of (a) Glycosyl Hydrolase Family 5 and (b) Glycosyl Hydrolase Family 9 proteins and their nearest characterized neighbors.

The largest fraction of GH and GH-like protein homologs identified in *A. protothecoides* are most similar to the family 5 (GH5) and family 9 (GH9) type of glycosyl hydrolases. Two of the GH5 family of proteins contain a PAN/APPLE-like domain, which are functionally versatile mediators of protein-protein and protein-carbohydrate interactions [39]. Finally, one protein (APUTEX2500005777) includes a CESA-CelA-like domain, of the glucosyl transferase family, usually involved in the elongation of the glucan chain of cellulose. Maximum likelihood phylogenetic trees of these two families with the characterized enzymes from the nearest neighbor are all of only two enzyme commission (EC) numbers (Fig. 7).

For GH5, all characterized homologs were EC 3.2.1.78, indicating mannan endo-1,4-beta-mannosidase activity. The GH9 homologs were all EC 3.2.1.4, indicating endocellulase activity. The characterized members of the GH 5 and 9 families have over 20 different catalytic functions, including xylanase activities (ECs 3.2.1.8, 3.2.1.37, 3.2.1.151), which were notably absent from the nearest homologs.

The sequences of the predicted GH5 and GH9 proteins were further analyzed for the presence of active sites. In family 5, each sequence was found to have the conserved Asn in the active site. More broadly, the complete PROSITE motif ([LIV]-[LIVMFYWGA](2)-[DNEQG]-[LIVMGST]-{SENR}-N-E-[PV]-[RHDNSTLIVFY]) was also present in all members, when allowing for two mismatches, with five having only one incorrect amino acid and member completely conserved (Supplemental Fig. 3). GH family 9 has two active site signatures ([STV]-x-[LIVMFY]-[STV]-x(2)-G-x-[NKR]-x(4)-[PLIVM]-H-x-R and [FYW]-x-D-x(4)-[FYW]-x(3)-E-x-[STA]-x(3)-N-[STA]), with three residues annotated as active sites. The first signature was not completely conserved in any of the members (and completely absent from one), but the active His residue is present in four of the seven members. The second signature is found in all but one member, with complete conservation in four of the seven, including the active Asp and Glu residues. Notably, coding sequences in the UTEX 25 gene models for most of the glycosyl hydrolases differed slightly from the sequences catalogued in NCBI, with additional exons or introns.

3.3. Proteomics and transcriptomics

To further characterize the enzymes involved in cellulolytic degradation and to explore the molecular mechanisms of plant substrate utilization, we analyzed the proteome and transcriptome of *A. protothecoides* UTEX 25 cultures grown in the presence or absence of either raw or H_2O_2-treated switchgrass. With respect to the proteome (2D gel analysis; Supplemental Fig. 4), several proteins were up-regulated in the presence of plant substrates compared to algae grown photoautotrophically (Supplemental Table 2). An attempt at analysis of supernatant failed to detect proteins, due likely to dilution or differences in liquid cultivation compared to solid cultivation on CMC. Eight of the proteins with the highest degree of differential expression were selected for identification by LC-MS/MS (Table 2).

The most differentially expressed algal protein in the presence of plant substrate (Spot 211) is predicted to have an NAD-binding domain and may be a potential oxidoreductase. The remaining seven spots contained peptides from the 50S ribosomal protein L9, a putative RNA-binding protein, putative plastid-lipid-associated chloroplastic protein 4, 60S ribosomal protein L6, Protein YOP1, Chlorophyll *a-b* binding protein 3C, ATP-dependent Clp protease proteolytic subunit-related protein 4, 40S ribosomal protein S8, 60S ribosomal protein L19-2, Chlorophyll *a-b* binding protein of LHCII type I, Ribulose bisphosphate carboxylase small chain 7, Fructokinase-1, photosystem II CP43

Table 2
The differentially expressed spots selected for proteomic analysis. Fold changes are calculated from average spot percentages (individual spot density divided by total density of all measured spots for duplicate 2D gels).

Spot no.	pI	MW (g)	Algae alone	Algae + raw plant	Algae + H_2O_2 plant	Algae + raw plant vs. algae alone		Algae + H_2O_2 plant vs. algae alone		Protein identified	Mascot score	Number of peptides ID'd	NCBI accession #	MW (g)
			Ave. %	Ave. %	Ave. %	Fold Change	*p*-Value	Fold Change	*p*-Value					
211	7.5	41,729	0.014	0.087	0.048	6.40	0.020	3.51	0.548	Hypothetical protein F751_1382	233	11	gi|760439869	40,424
										Ribulose bisphosphate carboxylase/ oxygenase activase	167	8	gi|760447469	85,706
263	5.3	37,867	0.032	0.053	0.039	1.69	0.031	1.22	0.116	Fructokinase-1	177	6	gi|760445347	36,020
										Photosystem II CP43 chlorophyll apoprotein	93	7	gi|570364819	50,611
										Chlorophyll *a-b* binding protein of LHCII type I	76	6	gi|760441981	26,783
391	6.5	28,763	0.005	0.008	0.006	1.70	0.018	1.34	0.499	Putative RNA-binding protein, partial	88	5	gi|760442443	15,939
										Putative RNA-binding protein, partial	53	3	gi|760442441	8465
466	7.5	23,833	0.007	0.034	0.020	5.12	0.037	3.07	0.230	Putative plastid-lipid-associated protein 4	98	2	gi|760439905	20,201
68	5.2	86,138	0.002	0.004	0.004	2.06	0.174	2.03	0.049	Hypothetical protein B456_007G254600	29	1	gi|763777340	11,450
433	6.4	26,103	0.225	0.488	0.479	2.17	0.346	2.13	0.005	Chlorophyll *a-b* binding protein of LHCII type I	455	21	gi|760441981	26,783
										60S ribosomal protein L6	226	8	gi|760444095	23,966
										Protein YOP1	159	8	gi|760449377	22,162
										Chlorophyll *a-b* binding protein 3C	104	4	gi|760440197	29,042
										ATP-dependent Clp protease proteolytic subunit-related protein 4	82	3	gi|760450513	33,229
										40S ribosomal protein S8	63	3	gi|760440193	22,103
										60S ribosomal protein L19–2	59	2	gi|760448873	23,181
572	5.0	16,102	0.054	0.055	0.119	1.02	0.963	2.21	0.033	Chlorophyll *a-b* binding protein of LHCII type I	306	11	gi|760441981	26,783
										Ribulose bisphosphate carboxylase small chain 7	76	6	gi|760437337	23,166
576	7.3	16,015	0.005	0.009	0.025	1.81	0.425	4.67	0.159	Photosystem I reaction center subunit II	116	4	gi|760446907	22,125
										50S ribosomal protein L9	95	7	gi|760447105	20,104

chlorophyll apoprotein, and Photosystem I reaction center subunit II. Interestingly, spot 68 had homology to a hypothetical protein from *Gossypium raimondii*, a species of cotton plant endemic to northern Peru. However, the Mascot score for this protein was very low (29), based upon a single peptide, and could be a false identification.

Transcriptomic analysis of *A. protothecoides* grown with and without CMC in either constant light or constant dark revealed several potential proteins targets for cellulolytic activity (Table 1; Supplemental Table 3). One of the most significantly differentially expressed transcripts in *A. protothecoides* grown in the presence of CMC in the dark was identified as a potential sugar transporter (APUTEX2500005548), up-regulated approximately 35× compared to phototrophically grown algae. Major facilitator superfamily members like this sugar transporter have been shown to enable the consumption of cellobiose, cellotriose, and cellotetraose in *N. crassa* [41]. We also observed significant upregulation of another potential sugar transporter (APUTEX2500002726) and F family ABC transporter (APUTEX2500005394), likely also involved in product uptake. Not unexpectedly, other identified transcripts up-regulated in the presence of CMC include genes involved in glycolysis and carbon storage. Consistent with the proteomics results, an increase in transcripts for a50S ribosomal protein subunit (APUTEX2500005089) and Fructokinase-1 (APUTEX2500004016) were observed. Notably, the potential oxidoreductase, Spot 211, discovered by proteomics was not significantly (within a p-value < 0.01) up-regulated in the presence of CMC (APUTEX2500005282). A member of the glycosyl hydrolase family 63 (APUTEX2500001195), annotated as a mannosylglycerate transporter (APUTEX2500001195), had higher levels of transcription in the presence of CMC in the dark. Most interesting, we observed the significant upregulation of three potential glycosyl hydrolases (APUTEX2500005369A, APUTEX2500005777, and APUTEX2500004827) transcripts from the GH5 and GH9 families (previously known as cellulase families A and E [40]), in cultures grown in the dark with CMC (Table 1). One of these family 9 glycosyl hydrolases (APUTEX2500005777) encodes the CESA cellulose synthase domain N-terminal to the glycosyl hydrolase. One of the family 5 glucosyl hydrolases (APUTEX2500004827) is one of the two glycosyl hydrolases identified earlier with a PAN/APPLE domain C-terminal to the glycosyl hydrolase. Six other potential glycosyl hydrolases sequences (APUTEX2500000370, APUTEX2500000844, APUTEX2500000742, APUTEX2500002671, APUTEX2500005369C, and APUTEX2500000981) with homology to proteins from families GT64, GH5, and GH9 were slightly induced in the presence of cellulose albeit not significantly. We also observed a significant induction of a transcript encoding the putative xylanase inhibitor (APUTEX2500002026), in the presence of CMC when grown in the dark. We hypothesize that this protein may serve a protective role, inhibiting hydrolytic activity within the cell.

Glycosyl hydrolases have been identified in other algae species and are generally thought to be involved in cell wall biosynthesis or turnover [42]. This is the first identification of algal glycosyl hydrolases potentially involved in plant substrate degradation. Research is ongoing to characterize these proteins and their role in plant deconstruction.

4. Conclusions

We discovered that the algae *A. protothecoides* UTEX 25 is capable of directly degrading and utilizing non-food plant substrates, such as switchgrass, for cell growth. *A. protothecoides* is also known to be a high-lipid producing green microalgae [1]. Although *A. protothecoides* was known to utilize sugars, we have demonstrated that growth can be enhanced in the presence of raw plant biomass under mixotrophic conditions and that enzymes with cellulolytic activity are actively secreted. Notably, *A. protothecoides* is substantially more effective at degrading cellulose than the only other known cellulose-degrading algae (Fig. 1) [7]. We also discovered that *A. protothecoides* has a faster growth rate on plant substrates compared to media without an organic C source (Fig. 2) and produces more lipids (Fig. 3). By examining both metabolite 1D ^{1}H NMR and profiling the plant glycome, we observed that the structure of plant substrate is modified by the algae. In NMR, a diminishing in peaks associated with single-hydrogen C—H bonds (0.8–1.5 ppm) and methylene bridge protons in cellulose (3.5–4.0 ppm) [31], when comparing cultures with plant substrate alone to cultures where algae and substrate are grown together (Fig. 4) is observed. Glycome profiling revealed that the predominant components utilized by the algae were xyloglucans (Fig. 6), suggesting that *A. protothecoides* actively expresses one or more xyloglucanases. Genomic, transcriptomic, and proteomic analyses identified many members of the glycosyl hydrolase family, including 6 potential glycosyl hydrolases from families 5 and 9 are encoded in the genome and were significantly up-regulated in the presence of plant substrate (Table 1). This research identifies potential gene targets for future genetic engineering of plant substrate degradation and calls attention to the potential benefits of growing algae with plant substrate at larger scales to generate greater algae biomass and lipid production.

Acknowledgements

This work was supported by a grant to A.N.B. from the Laboratory Directed Research and Development Early Career Research Program at Los Alamos National Laboratory (20170533ECR) and funds provided to S.R.S. by the Department of Energy's Bioenergy Technologies Office, under contract DE-EE0003046 to the National Alliance for Advanced Biofuels and Bioproducts. Special thanks to Geoffrey Smith, Erik Hanschen, Blake Hovde, Olga Chertkov, Cheryl Gleasner, Kim McMurry, Oliver Oviedo, and Xiang Li Zhang for technical assistance.

A.N.B. designed and contributed to all experiments, B.W.V. contributed to glycosyl hydrolase enzyme sequence and structural analyses, S.R.S completed all genomic and transcriptomic sequencing, N.S. contributed to cell growth and FAME data collection, J.Y.S. contributed to plate experiments, J.A.R. contributed to NMR data collection, and S.P. contributed to glycome data collection. All authors contributed to data analysis and to the writing of the article. 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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.algal.2018.07.001.

References

[1] J. Olivares, National Alliance for Advanced Biofuels and Bioproducts Synopsis (NAABB) Final Report, US DOE-EERE Biotechnol. Off., 2014.
[2] H. Xu, X. Miao, Q. Wu, High quality biodiesel production from a microalga *Chlorella protothecoides* by heterotrophic growth in fermenters, J. Biotechnol. 126 (2006) 499–507.
[3] H. Rismani-Yazdi, K.H. Hampel, C.D. Lane, B.A. Kessler, N.M. White, K.M. Moats, F.C.T. Allnutt, High-productivity lipid production using mixed trophic state cultivation of *Auxenochlorella* (*Chlorella*) protothecoides, Bioprocess Biosyst. Eng. 38 (2015) 639–650.
[4] K. Miazek, C. Remacle, A. Richel, D. Goffin, Effect of lignocellulose related compounds on microalgae growth and product biosynthesis: a review, Energies 7 (2014) 4446–4481, https://doi.org/10.3390/en7074446.
[5] M.H. Langholtz, B.J. Stokes, L.M. Eaton, 2016 Billion-Ton Report: Advancing Domestic Resources for a Thriving Bioeconomy, EERE Publication and Product Library, 2016.
[6] J.M. Jarvis, J.M. Billing, Y.E. Corilo, A.J. Schmidt, R.T. Hallen, T.M. Schaub, FT-ICR MS analysis of blended pine-microalgae feedstock HTL biocrudes, Fuel 216 (2018) 341–348, https://doi.org/10.1016/j.fuel.2017.12.016.
[7] O. Blifernez-Klassen, V. Klassen, A. Doebbe, K. Kersting, P. Grimm, L. Wobbe, O. Kruse, Cellulose degradation and assimilation by the unicellular phototrophic eukaryote *Chlamydomonas reinhardtii*, Nat. Commun. 3 (2012) 1214, , https://doi.org/10.1038/ncomms2210.
[8] N. Sueoka, Mitotic replication of deoxyribonucleic acid in *Chlamydomonas reinhardi*, Proc. Natl. Acad. Sci. 46 (1960) 83–91.

[9] S. Pattathil, U. Avci, J.S. Miller, M.G. Hahn, Immunological approaches to plant cell wall and biomass characterization: glycome profiling, Biomass Convers, Humana Press, Totowa, NJ, 2012, pp. 61–72, , https://doi.org/10.1007/978-1-61779-956-3_6.
[10] M. Liu, X. Mao, C. Ye, H. Huang, J.K. Nicholson, J.C. Lindon, Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy, J. Magn. Reson. 132 (1998) 125–129, https://doi.org/10.1006/jmre.1998.1405.
[11] S. Van Wychen, L.M.L. Laurens, Determination of Total Lipids as Fatty Acid Methyl Esters (FAME) by In Situ Transesterification, Laboratory Analytical Procedure (LAP), National Renewable Energy Laboratory (NREL), Golden, CO, 2013.
[12] P.H. O'Farrell, High resolution two-dimensional electrophoresis of proteins, J. Biol. Chem. 250 (1975) 4007–4021.
[13] A. Burgess-Cassler, J.J. Johansen, D.A. Santek, J.R. Ide, N.C. Kendrick, Computerized quantitative analysis of coomassie-blue-stained serum proteins separated by two-dimensional electrophoresis, Clin. Chem. 35 (1989) 2297–2304.
[14] B.R. Oakley, D.R. Kirsch, N.R. Morris, A simplified ultrasensitive silver stain for detecting proteins in polyacrylamide gels, Anal. Biochem. 105 (1980) 361–363.
[15] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels, Anal. Chem. 68 (1996) 850–858, https://doi.org/10.1021/ac950914h.
[16] C.C. Darie, K. Deinhardt, G. Zhang, H.S. Cardasis, M.V. Chao, T.A. Neubert, Identifying transient protein–protein interactions in EphB2 signaling by blue native PAGE and mass spectrometry, Proteomics 11 (2011) 4514–4528, https://doi.org/10.1002/pmic.201000819.
[17] I. Sokolowska, A.G. Woods, M.A. Gawinowicz, U. Roy, C.C. Darie, Identification of potential tumor differentiation factor (TDF) receptor from steroid-responsive and steroid-resistant breast cancer cells, J. Biol. Chem. 287 (2012) 1719–1733, https://doi.org/10.1074/jbc.M111.284091.
[18] I. Sokolowska, C. Dorobantu, A.G. Woods, A. Macovei, N. Branza-Nichita, C.C. Darie, Proteomic analysis of plasma membranes isolated from undifferentiated and differentiated HepaRG cells, Proteome Sci. 10 (2012) 47, https://doi.org/10.1186/1477-5956-10-47.
[19] I. Sokolowska, M.A. Gawinowicz, A.G.N. Wetie, C.C. Darie, Disulfide proteomics for identification of extracellular or secreted proteins, Electrophoresis 33 (2012) 2527–2536, https://doi.org/10.1002/elps.201200182.
[20] S. Bennett, Solexa Ltd, Pharmacogenomics 5 (2004) 433–438, https://doi.org/10.1517/14622416.5.4.433.
[21] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.-J. Chen, Z. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L.I. Alenquer, T.P. Jarvie, K.B. Jirage, J.-B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P. Yu, R.F. Begley, J.M. Rothberg, Genome sequencing in microfabricated high-density picolitre reactors, Nature 437 (2005) 376–380, https://doi.org/10.1038/nature03959.
[22] D.R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs, Genome Res. 18 (2008) 821–829, https://doi.org/10.1101/gr.074492.107.
[23] B. Ewing, L. Hillier, M.C. Wendl, P. Green, Base-calling of automated sequencer traces UsingPhred. I. Accuracy assessment, Genome Res. 8 (1998) 175–185, https://doi.org/10.1101/gr.8.3.175.
[24] C. Han, P. Chain, Finishing repetitive regions automatically with Dupfinisher, BIOCOMP, 2006, pp. 142–147.
[25] D. Gordon, C. Abajian, P. Green, Consed: a graphical tool for sequence finishing, Genome Res. 8 (1998) 195–202, https://doi.org/10.1101/gr.8.3.195.
[26] C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects, BMC Bioinf. 12 (2011) 491, https://doi.org/10.1186/1471-2105-12-491.
[27] C. Gao, Y. Wang, Y. Shen, D. Yan, X. He, J. Dai, Q. Wu, Oil accumulation mechanisms of the oleaginous microalga Chlorella protothecoides revealed through its genome, transcriptomes, and proteomes, BMC Genomics 15 (2014) 582, https://doi.org/10.1186/1471-2164-15-582.
[28] A. Marchler-Bauer, M.K. Derbyshire, N.R. Gonzales, S. Lu, F. Chitsaz, L.Y. Geer, R.C. Geer, J. He, M. Gwadz, D.I. Hurwitz, C.J. Lanczycki, F. Lu, G.H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, S.H. Bryant, CDD: NCBI's conserved domain database, Nucleic Acids Res. 43 (2015) D222–D226, https://doi.org/10.1093/nar/gku1221.
[29] O. Emanuelsson, H. Nielsen, S. Brunak, G. von Heijne, Predicting subcellular localization of proteins based on their N-terminal amino acid sequence, J. Mol. Biol. 300 (2000) 1005–1016, https://doi.org/10.1006/jmbi.2000.3903.
[30] S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0, Syst. Biol. 59 (2010) 307–321, https://doi.org/10.1093/sysbio/syq010.
[31] C.-C. Lo, P.S.G. Chain, Rapid evaluation and quality control of next generation sequencing data with FaQCs, BMC Bioinf. 15 (2014), https://doi.org/10.1186/s12859-014-0366-2.
[32] T. Heredia-Arroyo, W. Wei, B. Hu, Oil accumulation via heterotrophic/mixotrophic *Chlorella protothecoides*, Appl. Biochem. Biotechnol. 162 (2010) 1978–1995.
[33] J. O'Grady, J.A. Morgan, Heterotrophic growth and lipid production of *Chlorella protothecoides* on glycerol, Bioprocess Biosyst. Eng. 34 (2011) 121–125.
[34] M. Wan, P. Liu, J. Xia, J.N. Rosenberg, G.A. Oyler, M.J. Betenbaugh, Z. Nie, The effect of mixotrophy on microalgal growth, lipid content, and expression levels of three pathway genes in *Chlorella sorokiniana*, Appl. Microbiol. Biotechnol. 91 (2011) 835–844.
[35] A. Isogai, NMR analysis of cellulose dissolved in aqueous NaOH solutions, Cellulose 4 (1997) 99–107, https://doi.org/10.1023/A:1018471419692.
[36] U.C.C. of Algae, UTEX 25 *Chlorella protothecoides*, UTEX Cult. Collect. Algae. (n.d.). https://utex.org/products/utex-0025 (accessed February 19, 2018).
[37] E.G. Maxwell, N.J. Belshaw, K.W. Waldron, V.J. Morris, Pectin – an emerging new bioactive food polysaccharide, Trends Food Sci. Technol. 24 (2012) 64–73, https://doi.org/10.1016/j.tifs.2011.11.002.
[38] M. Hennig, J.N. Jansonius, A.C. Terwisscha van Scheltinga, B.W. Dijkstra, B. Schlesier, Crystal structure of concanavalin B at 1.65 Å resolution. An "inactivated" chitinase from seeds of *Canavalia ensiformis*, J. Mol. Biol. 254 (1995) 237–246, https://doi.org/10.1006/jmbi.1995.0614.
[39] P.J. Brown, A.C. Gill, P.G. Nugent, J.H. McVey, F.M. Tomley, 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. 497 (2001) 31–38.
[40] B. Henrissat, M. Claeyssens, P. Tomme, L. Lemesle, J.-P. Mornon, Cellulase families revealed by hydrophobic cluster analysis, Gene 81 (1989) 83–95, https://doi.org/10.1016/0378-1119(89)90339-9.
[41] J.M. Galazka, C. Tian, W.T. Beeson, B. Martinez, N.L. Glass, J.H.D. Cate, Cellodextrin transport in yeast for improved biofuel production, Science 330 (2010) 84–86, https://doi.org/10.1126/science.1192838.
[42] M.J. Scholz, T.L. Weiss, R.E. Jinkerson, J. Jing, R. Roth, U. Goodenough, M.C. Posewitz, G.H. Gerken, Ultrastructure and composition of the *Nannochloropsis gaditana* cell wall, Eukaryot. Cell 13 (2014) 1450–1464, https://doi.org/10.1128/EC.00183-14.