2002

Cellulose Synthase (*CesA*) Genes in the Green Alga *Mesotaenium caldariorum*

Alison W. Roberts
University of Rhode Island, aroberts@uri.edu

Eric M. Roberts
University of Rhode Island

See next page for additional authors

Citation/Publisher Attribution
Roberts, A. W., Roberts, E. M., & Delmer, D. P. (2002). Cellulose Synthase (*CesA*) Genes in the Green Alga *Mesotaenium caldariorum*. *Eukaryotic Cell, 1*(6), 847-855. doi: 10.1128/EC.1.6.847-855.2002
Available at: http://dx.doi.org/10.1128/EC.1.6.847-855.2002

EUKARYOTIC CELL, Dec. 2002, p. 847–855 Vol. 1, No. 6
1535-9778/02/$04.00+0 DOI: 10.1128/EC.1.6.847–855.2002

Cellulose Synthase (*CesA*) Genes in the Green Alga *Mesotaenium caldariorum*

Alison W. Roberts,[1]* Eric M. Roberts,[1] and Deborah P. Delmer[2]†

Department of Biological Sciences, University of Rhode Island, Kingston, Rhode Island,[1] and Section of Plant Biology, University of California, Davis, Davis, California[2]

Received 30 April 2002/Accepted 10 August 2002

Cellulose, a microfibrillar polysaccharide consisting of bundles of β-1,4-glucan chains, is a major component of plant and most algal cell walls and is also synthesized by some prokaryotes. Seed plants and bacteria differ in the structures of their membrane terminal complexes that make cellulose and, in turn, control the dimensions of the microfibrils produced. They also differ in the domain structures of their *CesA* gene products (the catalytic subunit of cellulose synthase), which have been localized to terminal complexes and appear to help maintain terminal complex structure. Terminal complex structures in algae range from rosettes (plant-like) to linear forms (bacterium-like). Thus, algal *CesA* genes may reveal domains that control terminal complex assembly and microfibril structure. The *CesA* genes from the alga *Mesotaenium caldariorum*, a member of the order Zygnematales, which have rosette terminal complexes, are remarkably similar to seed plant *CesA*s, with deduced amino acid sequence identities of up to 59%. In addition to the putative transmembrane helices and the D-D-D-QXXRW motif shared by all known *CesA* gene products, *M. caldariorum* and seed plant CesAs share a region conserved among plants, an N-terminal zinc-binding domain, and a variable or class-specific region. This indicates that the domains that characterize seed plant CesAs arose prior to the evolution of land plants and may play a role in maintaining the structures of rosette terminal complexes. The *CesA* genes identified in *M. caldariorum* are the first reported for any eukaryotic alga and will provide a basis for analyzing the *CesA* genes of algae with different types of terminal complexes.

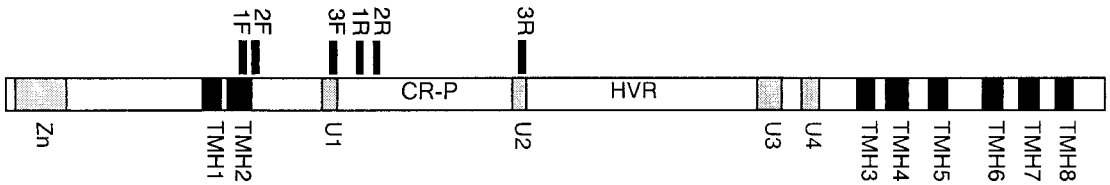
The fundamental unit of cellulose is the microfibril, consisting of a bundle of parallel chains of β-1,4-glucan that are hydrogen bonded to one another, forming a crystalline array (4, 25). Although cellulose is best known as the major component of plant and most algal cell walls, the *CesA* genes encoding the putative catalytic subunit of cellulose synthase (EC 2.4.1.12) were first identified in the cellulose-producing bacterium *Acetobacter xylinus* (39, 47). *CesA* genes in other prokaryotes and several seed plants, including cotton, *Arabidopsis thaliana*, maize, rice, and poplar (36) have subsequently been characterized. All share a common domain structure that includes putative transmembrane helices (TMH) and a cytoplasmic loop consisting of four conserved regions (U1 to U4), each containing a D residue or QXXRW sequence predicted to be involved in substrate binding and catalysis (D-D-D-QXXRW motif). An N-terminal zinc-binding domain, a strongly conserved region (CR-P) between U1 and U2, and a more variable region between U2 and U3 are found only in plant CesAs (11).

The predicted transmembrane nature of the CesA protein is consistent with the results of earlier freeze fracture electron microscopy studies showing nascent cellulose microfibrils associated with arrays of integral plasma membrane protein particles known as terminal complexes (reviewed in references 5, 10, and 15). Terminal complexes consisting of linear arrays of particles were first identified in the green alga *Oocystis apiculata* (7) and the cellulose-producing bacterium *Acetobacter xylinus* (8). Hexagonal arrays of particles termed rosettes were later identified as the terminal complexes of land plants (30). The arrangement of particles within terminal complexes appears to determine the dimensions of the cellulose microfibrils they produce (10, 14, 17, 19, 22, 44). For example, rosettes produce microfibrils composed of about 36 glucan chains but the large linear terminal complexes of giant marine green algae produce microfibrils containing up to 1,000 glucan chains (11). Recently, the role of terminal complexes in cellulose synthesis was demonstrated more directly by labeling freeze fracture replicas of mung bean rosettes with antibodies raised against a *CesA* gene product (24).

Although the factors that determine terminal complex structure, and thus microfibril dimensions, remain unknown, several lines of evidence indicate that *CesA* gene products play a direct role in maintaining the association of the particles that compose terminal complexes. The *rsw1* mutation in *Arabidopsis thaliana*, which results in a single amino acid substitution in the cytoplasmic domain of a cellulose synthase (*Arabidopsis thaliana* CesA1 [*At*CesA1]), disrupts assembly of crystalline cellulose microfibrils and leads to accumulation of noncrystalline β-1,4-glucan. Freeze fracture of *rsw1* mutants showed that the rosettes are dissociated (2). It has also been shown that the products of two cotton *CesA* genes (*Gossypium hirsutum CesA1* [*GhCesA1*] and *GhCesA2*) can associate in vitro through their zinc-binding domains, indicating a role for this domain in terminal complex assembly (26). The *Acetobacter* CesA proteins, which assemble as a linear terminal complex, lack the zinc-binding domain and two other domains found in all seed plant CesA proteins (11). These observations indicate that comparing the *CesA* genes of organisms with different types of termi-

* Corresponding author. Mailing address: Department of Biological Sciences, University of Rhode Island, Kingston, RI 02881. Phone: (401) 874-4098. Fax: (401) 874-5974. E-mail: aroberts@uri.edu.

† Present address: Food Security, The Rockefeller Foundation, New York, NY 10018.



1F/1R; 326 bp
1F/2R; 371 bp
1F/3R; 766 bp
3F/3R; 528 bp
3F/1R; 88 bp
3F/2R; 133 bp
2F/1R; 293 bp
2F/2R; 338 bp
2F/3R; 733 bp

Primer	Deduced amino acid sequence	Nucleotide sequence (5'-3')
IF	(C,S)E(I,V)WF	AATgAATTCWgYgARRTHTggTTY
2F	DQFPKW	TTAgAATTCgAYCARTTYCCNAARTgg
3F	CY(I/M/V)SDDG	ggTgYTAYRTNWSHgAYgAYgg
1R	(K,R)WVPF	AATggATCCAANggNACCCAYYT
2R	PE(W,F)YF	TTAggATCCAARTAVMAYTCNgg
3R	(L,F)(N,I)(L,V)DCD	ggTgRTCRCARTCVASRWTVA

FIG. 1. Domain structure of *GhCesA1*(GenBank accession number U58283) showing positions of degenerate primers (sequences listed in the box) designed to amplify *CesA* fragments from *M. caldariorum* strain 41 (UTEX Algal Culture Collection). Bars indicate the predicted products along with their sizes in base pairs. The zinc-binding domain (Zn), putative TMH, U domains, CR-P, and hypervariable region (HVR, also know as the CSR) are labeled.

nal complexes may reveal domains that control terminal complex assembly and thus microfibril structure.

The origin of rosettes is thought to be a crucial event in the evolution of land plants because it is linked to fundamental changes in cytokinesis and intercellular communication that provided the basis for the origin of the complex body plan (16). Green algae demonstrate the greatest diversity in terminal complex structure, and the history of the evolution of rosettes from linear terminal complexes appears to be preserved within this group (6, 21, 44). According to a recent classification, the monophyletic group Charophyta includes the land plants and six orders of green algae, including the Zygnematales, the Coleochaetales, and the Charales, which are thought to be the closest relatives of land plants (23). Within the Charophyta, all species examined have six-particle rosettes except for *Coleochaete scutata* (44), which has a unique eight-particle terminal complex (32). Other green algae have linear terminal complexes (44). *Mesotaenium caldariorum* is in the order Zygnematales, which diverged from the land plant lineage before the Coleochaetales and the Charales (23). Thus, characterization of *M. caldariorum CesA*s (*McCesA*s) will reveal the extent of *CesA* divergence since plants colonized the land and provide a basis for analyzing the *CesA* genes from algae with different types of terminal complexes, including *C. scutata*, with its apparently derived eight-particle terminal complex, and chlorophyte green algae, with presumably more primitive linear terminal complexes.

Here we report the identification of *CesA* genes in the unicellular charophycean alga *M. caldariorum*, the first reported from a eukaryotic alga.

MATERIALS AND METHODS

Culturing. A culture of the unicellular green alga, *M. caldariorum* strain 41 was obtained from the UTEX Algal Culture Collection (University of Texas, Austin). Suspension cultures were grown as described previously (27).

Primer design. The design of three forward and three reverse degenerate primers was based on regions of amino acid conservation among polypeptides deduced from *CesA* genes of *Acetobacter* and seed plants. The positions of the primers with respect to regions coding for characterized domains of *Gh*CesA1 (GenBank accession number U58283) are illustrated in Fig. 1. Primers 1F and 3R were based on primers UGF and DOMB1R, designed by Doblin et al. (12). For UGF, the redundancy at positions 12 and 13 was increased to account for additional Ser codons (AGC, AGT), and a G, the second nondegenerate base coding for Gly, was added to the 3′ end. The DOMB1R primer was modified by changing the clamp from CC to GG and increasing the degeneracy at position 18 to account for the substitution of Ile for Asn in some *Acetobacter* species (Fig. 1). Cloning sites were removed from both primers. Two additional forward primers (2F and 3F) were based on deduced amino acid sequences within domain U1 of *Gh*.CesA1 proteins. Additional reverse primers (1R and 2R) were based on sequences within the CR-P region. The sequences of all primers are listed in Fig. 1.

PCR cloning. Genomic DNA was isolated from cultured *M. caldariorum* cells by a rapid cetyltrimethylammonium bromide extraction method (28). Phage suspension from a genomic library of *M. caldariorum* DNA cloned into λGEM-11 *Bam*HI arms (27) was also used as a template.

Genomic DNA was amplified directly with various combinations of forward and reverse primers (Fig. 1) at an annealing temperature of 56°C with 2.5 mM $MgCl_2$ through 35 cycles. A genomic library phage suspension was amplified with primers 1F and 3R, and the products were subjected to nested and half-nested PCR with various combinations of forward and reverse primers under the same conditions used to amplify genomic DNA. Amplified fragments were gel purified and cloned into pCR-TOPO 2.1 (Invitrogen Corp., Carlsbad, Calif.) in accordance with the manufacturer's instructions. Plasmid DNA was isolated, and both strands were sequenced by primer walking and the BigDye dRhodamine Terminator method (PE Biosystems, Foster City, Calif.). Sequences were edited and assembled with Sequencher (version 4.0.5; Gene Codes Corp., Ann Arbor, Mich.).

Library screening. Probes were synthesized by incorporation of digoxigenin-dUTP by PCR with cloned *McCesA* fragments as templates and with the same primers that were originally used to amplify the cloned sequences. These probes were used to screen 300,000 plaques from the *M. caldariorum* genomic library. Phage DNA was isolated and subcloned by standard protocols (38). Plasmid DNA was isolated and sequenced as described above.

Sequence analysis. The cDNA and polypeptide sequences were deduced from genomic sequences by using NetGene2 (18) and GenScanW (9).

For phylogenetic analysis, *Mc*CesAs were compared with predicted amino acid sequences corresponding to *CesA* genes obtained from GenBank (http://www.ncbi.nlm.nih.gov) for *Arabidopsis thaliana* (accession numbers AF027172, AF027173, and AB018111 [2]; AB006703, AF016893, and AF062485; AF091713 [42]; and AL035526, AC007019, and AC006300), maize (*Zea mays*, AF200525, AF200526, AF200528, AF200529, AF200530, AF200531, AF200532, and

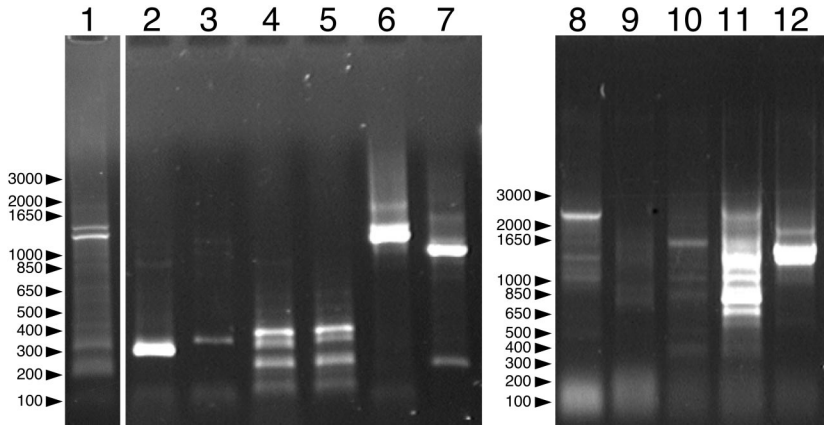


FIG. 2. PCR amplification of isolated genomic DNA and a genomic DNA library phage suspension from *M. caldariorum* strain 41 (UTEX Algal Culture Collection) with degenerate primers. Lane 1 shows amplification of a genomic DNA library phage suspension with the primer pair 1F-3R. Lanes 2 to 7 show the results of nested PCR with the products in lane 1 used as the templates with primer pairs 2F-1R, 2F-2R, 1F-1R, 1F-2R, 2F-3R, and 3F-3R, respectively. Lanes 8 to 12 show amplification of isolated genomic DNA with primer pairs 1F-1R, 1F-2R, 1F-3R, 3F-3R, and 2F-3R. The major bands in lanes 2, 7, and 12 were purified and cloned.

AF200533 [20]), poplar (*Populus tremuloides*, AF072131 [48] and hybrid *Populus tremula*/*Populus*. *alba*, AF081534), cotton (*G. hirsutum*, U58283 and U58284/AF254895 [34] and AF150630 [24]), tobacco (*Nicotiana tabacum*, AF304374 [12]), and *Anabaena* sp. strain PCC 7120 (BAB75456.1 = contig 326). The *CesA* sequence from *Nostoc punctiforme* (contig 499) was obtained from the Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/JGI_microbial/html/). Sequences were edited before alignment, as described in Results. Phylograms were constructed from the aligned sequences by using the heuristic search method in PAUP* (version 4.1b10; Sinauer Associates, Sunderland, Mass.) and were tested by bootstrap analysis (by the parsimony method, with 1,000 replicates). Trees were printed with TreeView (33).

Nucleotide sequence accession numbers. The nucleotide sequences of *McCesA1* and *McCesA2* have been deposited in GenBank under accession numbers AF525360 and AF525361, respectively.

RESULTS

Degenerate primers based on conserved regions of the deduced amino acid sequences of plant and prokaryote *CesA* genes were used to amplify *CesA* gene fragments from isolated genomic DNA and a genomic DNA library phage suspension from *M. caldariorum*. Primer pair 1F-3R amplified two major fragments and several minor fragments from the phage suspension (Fig. 2, lane 1). To test for specificity of amplification, the products of this reaction were subjected to fully nested and half-nested PCR (Fig. 1, lanes 2 to 7). Fully nested PCR with primer pairs 2F-1R and 2F-2R amplified fragments of about 300 and 350 bp, respectively (Fig. 2, lanes 2 and 3). This is close to the expected product sizes of 293 and 338 bp that were calculated from the *GhCesA1* sequence and verified by amplification of cloned *GhCesA1* with primer pairs 2F-1R and 2F-2R (data not shown). Half-nested PCR with primer pairs 1F-1R and 1F-2R produced numerous bands (Fig. 2, lanes 4 and 5), including those close to the expected product sizes of 326 and 370 bp, respectively. Half-nested PCR with primer pairs 2F-3R and 3F-3R produced strong bands at about 1 and 1.2 kb (Fig. 2, lanes 6 and 7). These exceed the expected values of 733 and 528 bp, presumably due to the presence of one or more introns, since the products differ from each other by about the expected 205 bp. Direct amplification of genomic DNA with primer pairs 1F-1R and 1F-2R did not produce products of the expected sizes (Fig. 1, lanes 8 and 9). However, at least some of the products of amplification with primer pairs 1F-3R, 3F-3R, and 2F-3R were similar in size to those resulting from amplification of the genomic DNA library phage suspension with the same primers.

The major bands from lanes 2, 7, and 12 (Fig. 2) were excised, purified, and cloned into pCR-TOPO 2.1. When the inserts were excised, 11 clones derived from the product in lane 2 (Fig. 2) appeared identical, but the major product in lane 7 produced two distinct classes of inserts and the product in lane 12 produced three distinct classes, including one containing an internal restriction site. A single representative of each of the six distinct clones was sequenced and compared to sequences in GenBank with BLASTX (1). The predicted products of two clones derived from amplification of genomic DNA with the primer pair 2F-3R were similar to *Gh*CesA1, spanning the regions upon which the primers were based (Fig. 1). Although the deduced polypeptides share 84% (*M. caldariorum* clone 1 [*Mc*1] compared with *Zea mays* CesA4 [*Zm*CesA4] [accession number AF200528]) and 73% (*Mc*2 compared with *Zm*CesA5 [accession number AF200529]) amino acid identity with known CesAs encoded within three open reading frames, they lack similarity in the amino acids encoded by the regions spanning nucleotides 409 to 705 and 970 to 1307 (*Mc*1) and nucleotides 394 to 624 and 889 to 1162 (*Mc*2). Prediction of intron-exon boundaries with NetGene2 (18) supports the hypothesis that these regions represent introns (Fig. 3). The spliced sequences have open reading frames of 733 and 718 bp, respectively, and their predicted amino acid sequences share 76% identity. *Mc*3 and *Mc*4, derived from amplification of a genomic library suspension with primer pairs 2F-2R and 3F-3R, respectively, were very similar to *Mc*2, differing by 17 bp within their 1,224-bp consensus sequence (data not shown). Together, the four clones represent at least two distinct *McCesA* sequences (Fig. 3).

The cloned *Mc*1 and *Mc*2 fragments were used to synthesize probes for screening an *M. caldariorum* genomic library. A total of 300,000 plaques were screened, 103 plaques were se-

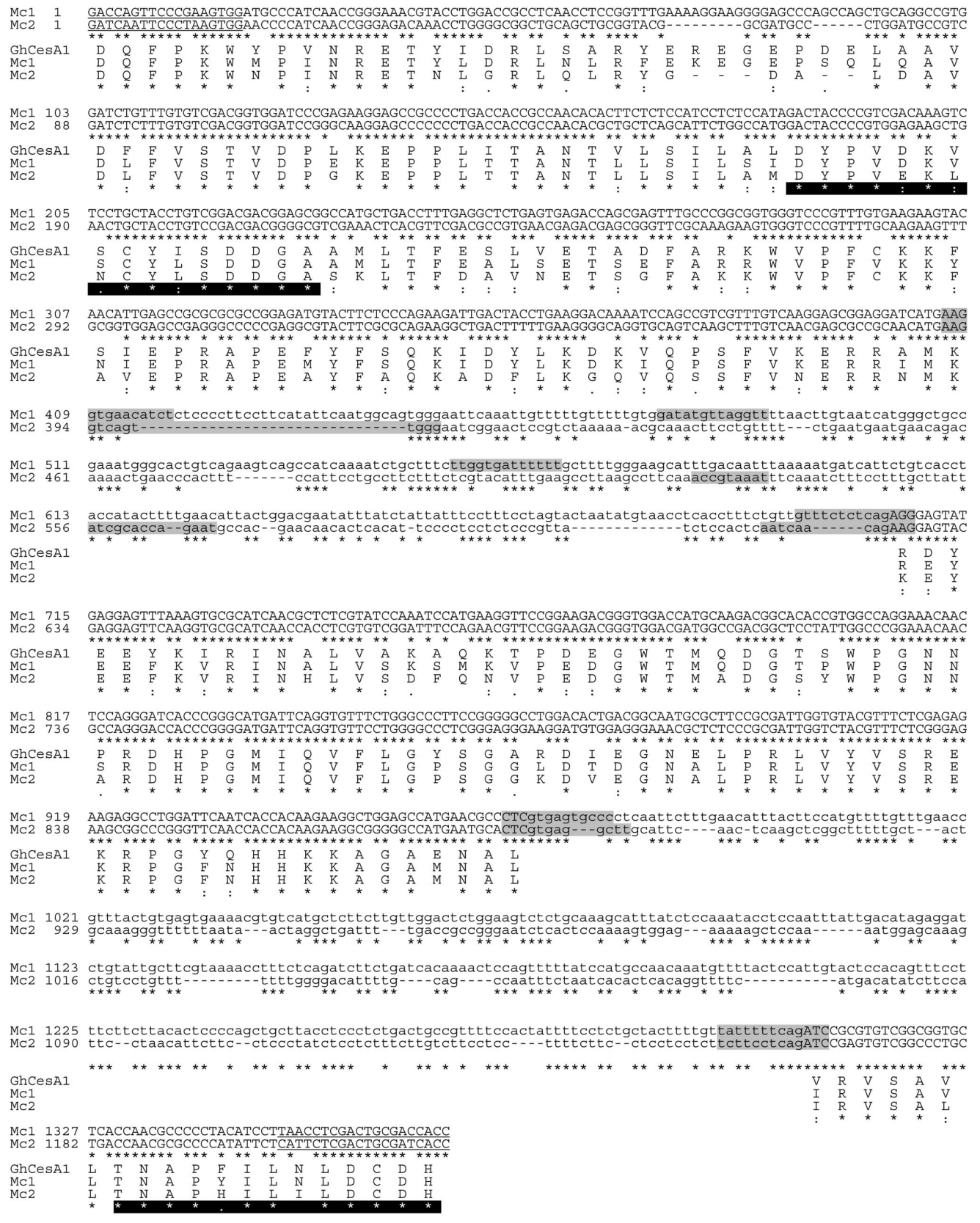


FIG. 3. Nucleotide sequence alignment of *CesA* gene fragments *Mc*1 and *Mc*2 amplified from *M. caldariorum* strain 41 (UTEX Algal Culture Collection) genomic DNA. Nucleotides corresponding to the primers used for amplification are highlighted in gray, intron-exon boundaries predicted by NetGene2 (18) are underlined, and predicted introns are shown in lowercase letters. Deduced amino acid sequences corresponding to both fragments are shown in alignment with the amino acid sequence deduced from a *GhCesA1* cDNA sequence. Putative catalytic domains U1 and U2 are highlighted in black. Asterisks indicate identity; colons and periods indicate full conservation of strong and weak groups, respectively (43).

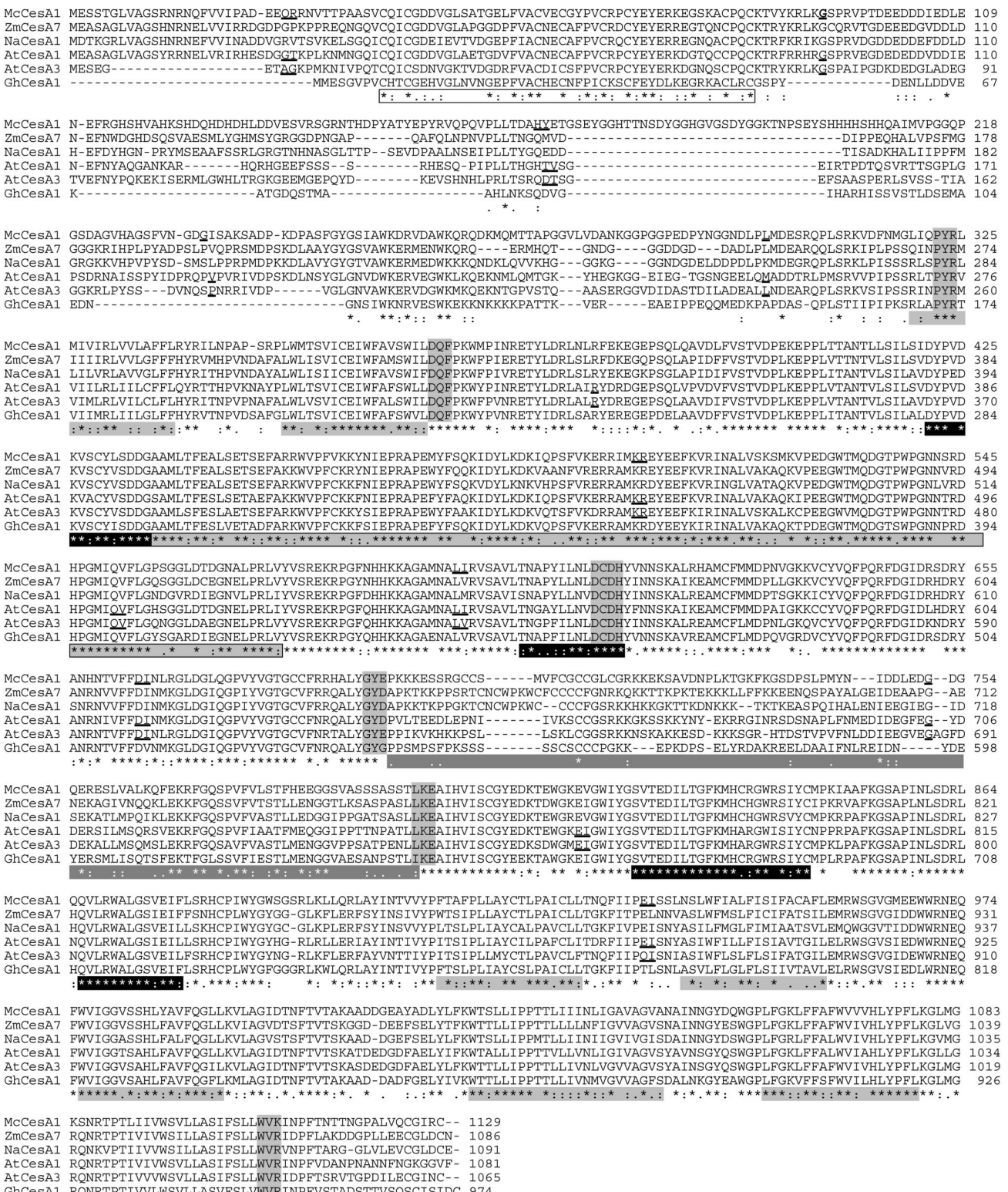


FIG. 4. Full-length deduced amino acid sequence of *M. caldariorum* strain 41 (UTEX Algal Culture Collection) CesA1 (GenBank accession number AF525360) aligned with deduced amino acid sequences of CesAs from maize (*Zm*CesA7 [AF200531]), tobacco (*Nicotiana tabacum* CesA1 [*Na*CesA1] [AF304374]), *Arabidopsis thaliana* (*At*CesA1 [AF027172], *At*CesA3 [AB018111]) and cotton (*Gh*CesA1 [U58283]). *Mc*CesA1 was deduced from a full-length nucleotide sequence isolated from a genomic library. The cDNA sequence was deduced with GenScanW (9). Intron insertion sites are underlined for *Mc*CesA1 and angiosperm sequences for which the genomic sequence is known. Domains are indicated below the sequences as follows: the zinc-binding domain with an unshaded box, putative TMH with light gray shading without a box, U domains with a black background, the CR-P with light gray shading within a box, and the variable region or CSR with dark gray shading without a box. Intron insertion sites are underlined in sequences for which they are known. Shaded amino acids indicate sites at which the sequences were edited when alignments for phylogenetic analysis were constructed. Asterisks indicate identical residues; colons and periods indicate full conservation of strong and weak groups, respectively (43).

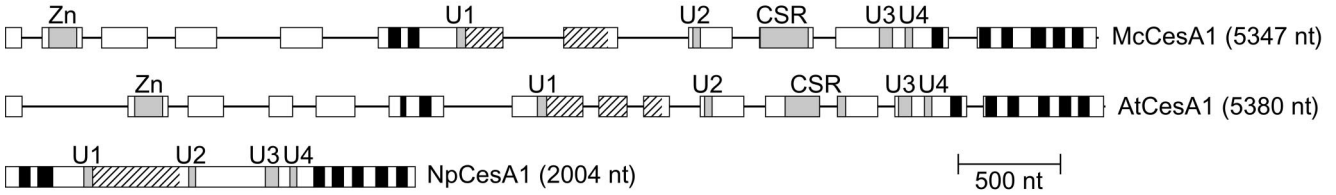


FIG. 5. Gene structure of *McCesA1* (GenBank accession number AF525360), *AtCesA1*(AF027172), and *Nostoc punctiforme CesA1* (*Np*CesA1) (contig 499). Boxes represent exons and lines represent introns. Putative TMH as predicted by HMMTOP (45) are shown in solid black, the CR-P is shown with diagonal hatching, and other domains are labeled.

lected, and 10 clones were purified. Phage DNA was isolated from each of these clones, and the inserts were excised with *Bam*HI, revealing five distinct restriction patterns. One clone was subcloned, sequenced in its entirety, and assembled. Comparison to sequences in GenBank with BLASTX revealed eight open reading frames with high similarity to plant *CesAs*. Start and stop codons were identified in frame at the N-terminal and C-terminal ends. Prediction of splicing sites by using GenScanW with both *Arabidopsis* and maize parameter matrices (9) indicated the presence of 11 exons and 10 introns, and the spliced gene produced an open reading frame of 3,390 bp. This gene was similar to that for *Mc*1, differing by only nine base substitutions and a 9-bp insert within their 1,377-bp consensus, and was named *McCesA1*. Two additional genomic clones were partially subcloned and sequenced. One was very similar to *McCesA1*, differing by a single deletion and two base substitutions, including a T→C substitution that produced an additional *Bam*HI site. The other clone was also similar to *Mc*1, differing by seven base substitutions within their 1,368-bp consensus sequences. Genomic clones corresponding to *Mc*2 to *Mc*4 were retrieved neither in the first screen nor when the genomic library was rescreened with only the probe based on *Mc*2. Of nine additional clones that were partially sequenced, three were nearly identical to *McCesA1*, three were more similar to *Mc*1, and three were similar to *McCesA1* but had additional deletions (data not shown). The designation *McCesA2* was assigned to *Mc*2, which represents *Mc*2 to *Mc*4.

By using ClustalX software (43), the predicted *Mc*CesA1 protein was compared with proteins representing different subfamilies of seed plant CesAs (20). The hypothetical *Mc*CesA1 protein of 1,130 amino acids contains all domains characterized in plant CesAs, as highlighted in Fig. 4. These include the zinc-binding domain near the N terminus (26). As predicted by HMMTOP (45), *Mc*CesA1 contains eight putative TMH. The cytoplasmic domain between the second and third TMH includes the four putative substrate-binding domains, U1 to U4, which are highly conserved in all known CesAs. Between U1 and U2 is the CR-P, a conserved region in plants (34) that is absent in bacterial CesAs (11) and is poorly conserved in some cyanobacteria (31) and the slime mold *Dictyostelium discoideum* (3). The *Mc*CesA1 CR-P is very similar to those of plant CesAs (up to 87% identity with *A. thaliana* CesA1 [*At*CesA1] [accession number AF027172]) and bears only slight similarity to those of cyanobacterial CesAs (13% identity with that of *Nostoc punctiforme*, contig 499).

*Mc*CesA1 is also similar to seed plant CesAs in regions that are not universally conserved (Fig. 4). These include the hypervariable region between U2 and U3 (34), also known as the class-specific region (CSR) (46). Like those of seed plants, the *Mc*CesA1 CSR contains basic residues at the N terminus and acidic residues at the C terminus, including DDXED and EXE motifs (amino acids 747 to 751 and 756 to 758, respectively). It also contains three K motifs (centered on amino acids 697, 720, and 735) and a cysteine-rich region (amino acids 703 to 715) and shares up to 42% amino acid identity with the CSRs of plant CesAs (*Zm*CesA1 [AF200525]). The region between the zinc-binding domain and the first TMH is also highly variable among the known CesAs. *Mc*CesA has the longest N terminus, including a unique 28-residue block and blocks corresponding to all of the sequence blocks found in the N-terminal regions of other plant CesA proteins.

McCesA1 joins 11 other *CesA* genomic sequences in which intron-exon boundaries are conserved (36, 37). All *McCesA1* intron-exon junctions are also found in *AtCesA1* and *AtCesA3* (Fig. 4 and 5). Within the region corresponding to *McCesA1* exon 6, these *Arabidopsis* genes have an additional intron, which is present in all other *Arabidopsis CesA*s except *AtCesA4*, *AtCesA5*, and *AtCesA9*. A second additional intron within the region corresponding to *McCesA1* exon 7 is present in all *Arabidopsis CesA*s except *AtCesA7*, and a third additional intron in the region corresponding to *McCesA1* exon 10 is present in all *Arabidopsis CesA*s. In *McCesA1* and all *CesA*s examined, the C-terminal exon contains TMH-4 through TMH-8 and the penultimate exon contains H-3, H-4, and TMH-3 (Fig. 5).

Figure 6A shows a parsimony phylogram corresponding to the bootstrap consensus tree for deduced amino acid sequences encoded by *McCesA1* and selected seed plant *CesA*s, rooted with deduced amino acid sequences encoded by two cyanobacterial *CesA*s. Prior to alignment with ClustalX (43), the sequences were edited to remove the poorly conserved N terminus upstream of the (P/L/S)(Y/F)R consensus sequence, the variable region between the G(Y/F)(D/E/S/G) and (L/I)(K/R)E consensus sequences, and the C terminus downstream of the WV(R/K) consensus sequence (Fig. 4). The analysis shows a high similarity between *Mc*CesA1 and seed plant CesAs (Fig. 6A). Although the possibility of higher-level groupings is strongly supported (bootstrap values, 72 to 100%), the early divergence of *Mc*CesA1 and separation of the seed plant *CesA*s into two major clades is supported only weakly (bootstrap values, 52 to 68%). An analysis including *Mc*CesA2 was carried out with the conserved region from the DQF consensus sequence directly following TMH-2 to the DCDH consensus sequence of U2 (Fig. 4). The unrooted phylogram corresponding to the bootstrap consensus tree shows strong support for an *M. caldariorum* clade that is separate from that

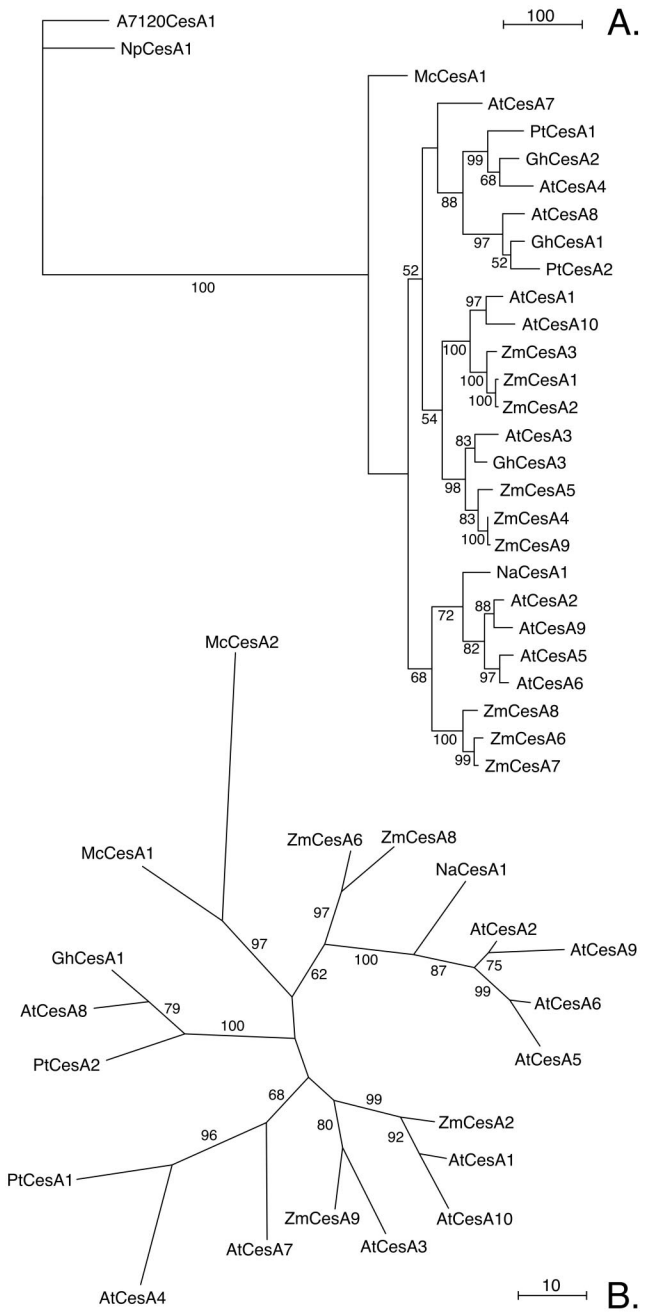


FIG. 6. Parsimony phylograms corresponding to the majority consensus trees from 1,000 bootstrap replicates. Bootstrap values are indicated in parentheses. (A) Phylogram rooted with cyanobacterial CesA sequences; (B) unrooted phylogram constructed with CesA fragments corresponding to the *Mc*2 PCR product. Scale bars indicate the number of changes. NpCesA1, *Nostoc punctiforme* CesA1; PtCesA1, *Populus tremuloides* CesA1; NtCesA1, *Nicotiana tabacum* CesA1.

corresponding to the seed plant CesAs (Fig. 6B). Some of the sequences included in Fig. 6A were omitted from Fig. 6B for clarity. When included, their positions were consistent with those shown in Fig. 6A. The topologies of trees created using distance methods (neighbor joining) were identical to those shown except for the position of *At*CesA7 in the rooted tree (data not shown).

DISCUSSION

CesA domains and terminal complex structure. The deduced amino acid sequence encoded by *McCesA1* includes the D-D-D-QXXRW motif and putative TMH that characterize all known CesAs and shows remarkable similarity to seed plant CesAs, with amino acid identities up to 59% (*Zm*CesA7 [accession number AF200531]). *Mc*CesA1 has a zinc-binding domain, a highly conserved CR-P between U1 and U2 that is 87% identical to the most similar seed plant CR-P (that of *At*CesA1 [AF027172]), and a variable region or CSR between U2 and U3. These features were previously known only in seed plant CesAs (11); however, our results indicate that these domains arose prior to the evolution of land plants. Of the three CesA domains unique to *M. caldariorum* and seed plants, the CR-P may have had the earliest origin, since CesAs from two strains of cyanobacteria contain an insertion with limited similarity to the CR-P (31) and a CesA from the cellular slime mold *D. discoideum* contains a highly divergent insertion between U1 and U2 (3).

Although we have no direct evidence that *Mc*CesAs function in cellulose synthesis, their strong similarity to seed plant CesAs is consistent with this hypothesis. In *Arabidopsis* and other seed plants, the *CesA* family is part of a larger superfamily that includes six families of cellulose synthase-like (*Csl*) genes of unknown function resembling bacterial *CesA*s in domain structure (37). The *McCesA*s most closely resemble members of the *CesA* family, based on the presence of the characteristic zinc-binding, CR-P, and CSR domains. Although tobacco *Csl* genes were amplified with primers similar to those used in this study (12), no *Csl* sequences were amplified from *M. caldariorum* DNA.

Rosette terminal complexes appear to have arisen among the charophytes, the group of green algae thought to be most closely related to land plants (23, 29). Whereas the linear terminal complexes of the chlorophyte green algae are similar to those of *Acetobacter* and *Dictyostelium* (8, 17, 44), land plants and two orders of charophyte green algae (Charales and Zygnematales) have rosette terminal complexes (44). The striking similarity between *Mc*CesAs and those of seed plants is consistent with the hypotheses that acquisition of their shared zinc-binding and CSR domains and specific features of the CR-P accompanied the origin of rosette terminal complexes and that these domains and features may be involved in the assembly or function of the rosette. The putative role of the zinc-binding domain in protein-protein interaction is consistent with this interpretation (26). In addition, *Mc*CesA1's deduced amino acid sequence is identical to that of wild-type *At*CesA1 in the 20 amino acids upstream of TMH3. In the *Arabidopsis rsw1* mutant, this region includes a V-for-A substitution that results in dissociation of rosettes, indicating that the A and surrounding amino acids may be involved in maintaining rosette integrity (2).

The terminal complexes of the two earlier-divergent orders of charophyte green algae (Klebsormidiales and Chlorokybales) have not been examined. Analysis of the terminal complexes and *CesA* genes in these groups and the *CesA*s of chlorophyte green algae with linear terminal complexes may further reveal the relationship between *CesA* domains and terminal complex organization. Additional insight might be

gained from examining the *CesA* genes of *C. scutata* in the order Coleochaetales, which appears to have diverged from the land plant lineage after the Zygnematales but before the Charales (23). The taxonomic position of *C. scutata* (23) is consistent with the hypothesis that its unique eight-particle terminal complex (32) was derived from a rosette.

Multiple *McCesA* genes. Our identification of several genomic clones very similar to *McCesA1* parallels results obtained in a screen of the same genomic library for phytochrome genes. To determine whether the similar phytochrome clones were alleles or distinct genes, Lagarias et al. (27) sequenced phytochrome fragments amplified from populations grown from single cells. Each of these clonal populations contained two or more very similar phytochrome genes, indicating that these similar genes are present within the genome of a single individual. *M. caldariorum* is expected to be haploid like other members of the Desmidiaceae, but polyploidy cannot be ruled out (27). By analogy, these data indicate that clones very similar to *McCesA1* represent separate genes, unless the population from which the library was derived is polyploid. The *McCesA*s represented by *Mc*2 to *Mc*4 are distinct from *McCesA1* based on substantial sequence divergence. Although we identified numerous genomic clones similar to *McCesA1* and the *Mc*1 PCR product, we found none corresponding to *Mc*2 to *Mc*4, even when only probes based on *Mc*2 were used for screening. The genes corresponding to *Mc*2 to *Mc*4 may be poorly represented in the genomic library. However, the cloning of three PCR products that are similar to each other but distinct from *Mc*1 and *McCesA1* indicates that *M. caldariorum* has at least two *CesA* genes and possibly six or more.

Seed plants have moderately large families of *CesA* genes. In *Arabidopsis*, expression analysis and genetic complementation studies have revealed that many of the 10 members of the *CesA* gene family serve distinct functions. For example, *AtCesA4* (20), *AtCesA7*, and *AtCesA8* (41) are expressed during secondary cell wall synthesis in vascular tissue whereas *AtCesA6* (13), *AtCesA1*, and *At*CesA3 (40) are expressed in expanding cells. However, even *CesA*s with identical expression patterns, such as *AtCesA7* and *AtCesA8* or *AtCesA1* and *AtCesA3*, cannot complement mutations in the coexpressing paralog (40, 41). These observations led to the hypothesis that rosette terminal complexes are assembled from at least two different CesA isoforms (see reference 35 for a review). Although *M. caldariorum* is unicellular, its cell division cycle includes distinct phases of wall deposition, such as cell plate formation during cytokinesis and primary and secondary cell wall deposition. The multiple *Mc*CesAs may be required for the different phases of cell wall deposition or for the assembly of rosettes composed of multiple CesA isoforms.

Phylogenetic analysis of *McCesA* genes. Phylogenetic analysis confirmed the high similarity between *M. caldariorum* and seed plant *CesA*s inferred from the alignments of deduced amino acid sequences. The possibility of the divergence of *McCesA1* before the diversification of seed plant *CesA*s was weakly supported. Our analysis is consistent with previously reported topologies for seed plant CesAs with the exception of the position of *At*CesA7, which varies between our parsimony and neighbor-joining trees and also among published phylogenies (20, 37, 46).

The *AtCesA* genes have been classified according to their expression patterns; i.e., type I is expressed in developing vascular tissue (*AtCesA4*, *AtCesA7*, and *AtCesA8*), type II is expressed in expanding tissue (*AtCesA1,* to *AtCesA,3*, *AtCesA5*, and *AtCesA6*), and type III finds limited expression in floral organs and leaf-stem junctions (*AtCesA9,* and *AtCesA10*) (D. P. Delmer, R. Eshed, P. Hogan, M. Doblin, D. Jacob-Wilk, L. Peng, A. Roberts, and D. Holland, Abstr. Quadr. Joint Annu. Meet. Am. Soc. Plant Biol. Can. Soc. Plant Physiol., abstr. 510, 2001). CesA phylogenies show that type I *At*CesAs form a clade with other CesAs expressed during secondary cell wall deposition and that type II and type III CesAs are distributed between two major clades (20, 37, 46). In our analysis, the two *Mc*CesAs occupy a separate clade with strong bootstrap support, indicating that *CesA*s specialized for primary and secondary cell wall deposition diverged after the origin of land plants.

When alignments (Fig. 4) and gene phylogenies (Fig. 6) are considered together, some interesting patterns emerge. *Mc*CesA1 has the longest N terminus of any of the CesAs analyzed and includes a unique sequence block and sequence blocks found in the other CesAs (Fig. 4). Thus, it appears that diversification of seed plant CesAs has involved the deletion of several sequence blocks. The *Arabidopsis* type I and related seed plant CesAs have lost the greatest number of N-terminal sequence blocks. The structure of the CSR region has been used previously to group the *CesA* genes of rice into different classes (46). Although the *Mc*CesA1 CSR is similar to those of seed plants, it cannot be assigned to one of the previously described classes based on the organization of the CSR.

Work is under way to examine the CesAs of algae with diverse types of terminal complexes in an effort to identify domains that may be involved in terminal complex structure and assembly. Analysis of expression patterns of CesAs in algae may also provide insight into the evolutionary origin of nonidentical pairs of CesAs and their roles in terminal complex assembly and microfibril synthesis.

ACKNOWLEDGMENTS

This work was support by grant DBI 9872627 from the National Science Foundation and the University of Rhode Island Foundation.

We gratefully acknowledge the assistance of members of the Delmer lab, especially Monica Doblin, Pat Hogan, and Yasushi Kawagoe. We also acknowledge Monika Doblin for sharing, prior to publication, the primer sequences upon which primers 3F and 3R were based, and Yasushi Kawagoe for the use of primers 1F, 2F, 1R, and 2R. We thank Donna and Clark Lagarias for use of the *M. caldariorum* genomic library and much helpful advice.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Arioli, T., L. Peng, A. S. Betzner, J. Burn, W. Wittke, W. Herth, C. Camilleri, H. Hofte, J. Plazinski, R. Birch, A. Cork, J. Glover, J. Redmond, and R. E. Williamson.** 1998. Molecular analysis of cellulose biosynthesis in *Arabidopsis*. Science **279:**717–720.
3. **Blanton, R. L., D. Fuller, N. Iranfar, M. J. Grimson, and W. F. Loomis.** 2000. The cellulose synthase gene of *Dictyostelium*. Proc. Natl. Acad. Sci. USA **97:**2391–2396.
4. **Brett, C. T.** 2000. Cellulose microfibrils in plants: biosynthesis, deposition, and integration into the cell wall. Int. Rev. Cytol. **199:**161–199.
5. **Brown, R. M., Jr.** 1985. Cellulose microfibril assembly and orientation: recent developments. J. Cell Sci. Suppl. **2:**13–32.
6. **Brown, R. M., Jr., C. H. Haigler, J. Suttie, A. R. White, E. Roberts, C. Smith, T. Itoh, and K. Cooper.** 1983. The biosynthesis and degradation of cellulose. J. Appl. Polymer Sci. **37:** 33–78.
7. **Brown, R. M., Jr., and D. Montezinos.** 1976. Cellulose microfibrils: visual-

ization of biosynthetic and orienting complexes in association with the plasma membrane. Proc. Natl. Acad. Sci. USA **73:**143–147.
8. **Brown, R. M., Jr., J. H. M. Willison, and C. L. Richardson.** 1976. Cellulose biosynthesis in *Acetobacter xylinum*: visualization of the site of synthesis and direct measurement of the *in vivo* process. Proc. Natl. Acad. Sci. USA **73:**4565–4569.
9. **Burge, C., and S. Karlin.** 1997. Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. **268:**78–94.
10. **Delmer, D. P.** 1987. Cellulose biosynthesis. Annu. Rev. Plant Physiol. **38:** 259–290.
11. **Delmer, D. P.** 1999. Cellulose biosynthesis: exciting times for a difficult field of study. Annu. Rev. Plant Physiol. Plant Mol. Biol. **50:**245–276.
12. **Doblin, M. S., L. De Melis, E. Newbigin, A. Bacic, and S. M. Read.** 2001. Pollen tubes of *Nicotiana alata* express two genes from different β-glucan synthase families. Plant Physiol. **125:**2040–2052.
13. **Fagard, M., T. Desnos, T. Desprez, F. Goubet, G. Refregier, G. Mouille, M. McCann, C. Rayon, S. Vernhettes, and H. Hofte.** 2000. *PROCUSTE1* encodes a cellulose synthase required for normal cell elongation specifically in roots and dark-grown hypocotyls of *Arabidopsis*. Plant Cell **12:**2409–2423.
14. **Giddings, T. H., Jr., D. L. Brower, and L. A. Staehelin.** 1980. Visualization of particle complexes in the plasma membrane of *Micrasterias denticulata* associated with the formation of cellulose fibrils in primary and secondary cell walls. J. Cell Biol. **84:**327–339.
15. **Giddings, T. H., Jr., and L. A. Staehelin.** 1991. Microtubule-mediated control of microfibril deposition: a re-examination of the hypothesis, p. 85–99. *In* C. W. Lloyd (ed.), The cytoskeletal basis of plant growth and form. Academic Press, New York, N.Y.
16. **Graham, L. E., M. E. Cook, and J. S. Busse.** 2000. The origin of plants: body plan changes contributing to a major evolutionary radiation. Proc. Natl. Acad. Sci. USA **97:**4535–4540.
17. **Grimson, M. J., C. H. Haigler, and R. L. Blanton.** 1996. Cellulose microfibrils, cell motility, and plasma membrane protein organization change in parallel during culmination in *Dictyostelium discoideum*. J. Cell Sci. **109:** 3079–3087.
18. **Hebsgaard, S. M., P. G. Korning, N. Tolstrup, J. Engelbrecht, P. Rouze, and S. Brunak.** 1996. Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. Nucleic Acids Res. **24:** 3439–3452.
19. **Herth, W.** 1983. Arrays of plasma-membrane "rosettes" involved in cellulose microfibril formation of *Spirogyra*. Planta **159:**347–356.
20. **Holland, N., D. Holland, T. Helentjaris, K. S. Dhugga, B. Xoconostle-Cazares, and D. P. Delmer.** 2000. A comparative analysis of the plant cellulose synthase (CesA) gene family. Plant Physiol. **123:**1313–1323.
21. **Hotchkiss, A. T., Jr.** 1989. Cellulose biosynthesis: the terminal complex hypothesis and its relationship to other contemporary research issues, p. 232–247. *In* N. G. Lewis and M. G. Paice (ed.), Plant cell wall polymers. Biogenesis and biodegradation. American Chemical Society, Washington, D.C.
22. **Itoh, T., R. M. O'Neil, and R. M. Brown, Jr.** 1984. Interference of cell wall regeneration of *Boergesenia forbesii* protoplasts by Tinopal LPW, a fluorescent brightening agent. Protoplasma **123:**174–183.
23. **Karol, K. G., R. M. McCourt, M. T. Cimino, and C. F. Delwiche.** 2001. The closest living relatives of land plants. Science **294:**2351–2353.
24. **Kimura, S., W. Laosinchai, T. Itoh, X. Cui, C. R. Linder, and R. M. Brown, Jr.** 1999. Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant *Vigna angularis*. Plant Cell **11:**2075–2086.
25. **Koyama, M., W. Helbert, T. Imai, J. Sugiyama, and B. Henrissat.** 1997. Parallel-up structure evidences the molecular directionality during biosynthesis of bacterial cellulose. Proc. Natl. Acad. Sci. USA **94:**9091–9095.
26. **Kurek, I., Y. Kawagoe, D. Jacob-Wilk, M. Doblin, and D. P. Delmer.** 2002. Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domains. Proc. Natl. Acad. Sci. USA **99:**11109–11114.
27. **Lagarias, D. M., S.-H. Wu, and J. C. Lagarias.** 1995. Atypical phytochrome gene structure in the green alga *Mesotaenium caldariorum*. Plant Mol. Biol. **29:**1127–1142.
28. **Lukowitz, W., U. Mayer, and G. Jurgens.** 1996. Cytokinesis in the Arabidopsis embryo involves the syntaxin-related KNOLLE gene product. Cell **84:**61–71.
29. **McCourt, R. M.** 1995. Green algal phylogeny. Tree **10:**159–163.
30. **Mueller, S. C., and R. M. Brown, Jr.** 1980. Evidence for an intramembrane component associated with a cellulose microfibril-synthesizing complex in higher plants. J. Cell Biol. **84:**315–326.
31. **Nobles, D. R., D. K. Romanovicz, and R. M. Brown, Jr.** 2001. Cellulose in cyanobacteria. Origin of vascular plant cellulose synthase? Plant Physiol. **127:**529–542.
32. **Okuda, K., and R. M. Brown, Jr.** 1992. A new putative cellulose-synthesizing complex of *Coleochaete scutata*. Protoplasma **168:**51–63.
33. **Page, R. D. M.** 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12:**357–358.
34. **Pear, J. R., Y. Kawagoe, W. E. Schreckengost, D. P. Delmer, and D. M. Stalker.** 1996. Higher plants contain homologs of the bacterial *celA* genes encoding the catalytic subunit of cellulose synthase. Proc. Natl. Acad. Sci. USA **93:**12637–12642.
35. **Perrin, R. M.** 2001. Cellulose: how many cellulose synthases to make a plant? Curr. Biol. **11:**R213–R216.
36. **Richmond, T.** 2000. Higher plant cellulose synthases. Genome Biol. **1:**3001.1–3001.6.
37. **Richmond, T. A., and C. R. Somerville.** 2000. The cellulose synthase superfamily. Plant Physiol. **124:**495–498.
38. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
39. **Saxena, I. M., F. C. Lin, and R. M. Brown, Jr.** 1990. Cloning and sequencing of the cellulose synthase catalytic subunit gene of *Acetobacter xylinum*. Plant Mol. Biol. **15:**673–683.
40. **Scheible, W.-R., R. Eshed, T. Richmond, D. Delmer, and C. Somerville.** 2001. Modifications of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in *Arabidopsis Ixr1* mutants. Proc. Natl. Acad. Sci. USA **98:**10079–10084.
41. **Taylor, N. G., S. Laurie, and S. R. Turner.** 2000. Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in *Arabidopsis*. Plant Cell **12:**2529–2539.
42. **Taylor, N. G., W.-R. Scheible, S. Cutler, C. R. Somerville, and S. R. Turner.** 1999. The *irregular xylem3* locus of Arabidopsis encodes a cellulose synthase required for secondary cell wall synthesis. Plant Cell **11:**769–779.
43. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
44. **Tsekos, I.** 1999. The sites of cellulose synthesis in algae: diversity and evolution of cellulose-synthesizing enzyme complexes. J. Phycol. **35:**635–655.
45. **Tusnády, G. E., and I. Simon.** 1998. Principles governing amino acid composition of integral membrane proteins: applications to topology prediction. J. Mol. Biol. **283:**489–506.
46. **Vergara, C. E., and N. C. Carpita.** 2001. β-D-Glycan synthases and the *CesA* gene family: lessons to be learned from the mixed-linkage (1→3),(1→4)β-D-glucan synthase. Plant Mol. Biol. **47:**145–160.
47. **Wong, H. C., A. L. Fear, R. D. Calhoon, G. H. Eichinger, R. Mayer, D. Amikam, M. Benziman, D. H. Gelfand, J. H. Meade, A. W. Emerick, R. Bruner, A. Ben-Bassat, and R. Tal.** 1990. Genetic organization of the cellulose synthase operon in *Acetobacter xylinum*. Proc. Natl. Acad. Sci. USA **87:**8130–8134.
48. **Wu, L., C. P. Joshi, and V. L. Chiang.** 2000. A xylem-specific cellulose synthase gene from aspen (*Populus tremuloides*) is responsive to mechanical stress. Plant J. **22:**495–502.