IDENTIFICATION AND CHARACTERIZATION OF A CELLULOSE SYNTHASE FROM THE CALCIFYING RED ALGA *CALLIARTHRON TUBERCULOSUM*

by

Jan Y Xue

B.Sc., The University of British Columbia, 2018

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Botany)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

January 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis entitled:

Identification and Characterization of a Cellulose Synthase from the Calcifying Red Alga *Calliarthron tuberculosum*

Submitted by **Jan Y Xue** in partial fulfillment of the requirements for the degree of **Master of Science** in Botany.

Examining Committee:

Anne Lacey Samuels, Botany
Co-supervisor

Patrick Martone, Botany
Co-supervisor

George Haughn, Botany
Supervisory Committee Member

Patrick Keeling, Botany
Supervisory Committee Member

Harry Brumer, Chemistry, Biochemistry, Botany
Additional Examiner

Abstract

In land plants and algae, cellulose, a glucose polymer, is important for strengthening tissues and preventing breakage in the face of physical forces. Cellulose synthase enzymes (CESA) are responsible for producing cellulose. While our understanding of land plant CESAs has advanced for several species, it is unclear whether the systems surrounding cellulose synthesis are the same in all plant lineages. For example, no red algal CESAs have been functionally demonstrated. The objective of this thesis is to discover and characterize putative CESA encoding genes from the calcifying red alga *Calliarthron tuberculosum* and compare their function to those from the land plant eudicot *Arabidopsis thaliana*.

Using a bioinformatics approach, I identified three candidate *CESA*s from *Calliarthron tuberculosum*'s transcriptome dataset (*CtCESA1*, *CtCESA2*, and *CtCESA3*). I explore the evolution of *CESA*s in gene tree analysis and find that while *CtCESA1* was closely related to other red algal *CESA* sequences, *CtCESA2* and *3* were more closely related to bacterial cellulose synthases (Ch 2). Using yeast and insect cell expression systems, I heterologously express, purify, and test the CtCESA1 protein in glucose tracer assays to look for polymer formation. CtCESA1 showed evidence of glucan synthase activity that was comparatively lower than plant (PttCESA8) and bacterial (BCSA/BCSB) cellulose synthases (Ch 3).

Finally, I test for functional compatibilities between the land plant (*A. thaliana)* and red algal (*C. tuberculosum*) CESAs. *A. thaliana* encodes multiple non-redundant CESAs that function in primary cell wall and secondary cell wall regions as well as several other accessory proteins

critical to cellulose synthesis. However, only some accessory proteins were recovered from *C tuberculosum*'s transcriptome in bioinformatics analyses (Ch 2). To ultimately test for functional differences, I introduced the *CtCESA1* gene into *A. thaliana cesa* mutants deficient in cellulose production (Ch 4). The red algal *CtCESA1* partially rescued the *A. thaliana* primary cell wall *cesa6* mutant but not *cesa3* or the secondary cell wall *cesa7* mutant. This thesis collectively presents the first functional evidence of a red algal CESA and demonstrates a combination of both deeply conserved and largely distinct aspects of cellulose production between the red algal and land plant lineages.

Lay Summary

Calliarthron tuberculosum is a calcified red alga that thrives in coastal habitats, surviving incredibly strong forces associated with wave impacts. This strength comes in part from cellulose, a biopolymer made of glucose chains, found in the walls surrounding each cell. Cellulose is incredibly common and found in virtually all plants that grow on land. Enzymes called cellulose synthases (CESA) are responsible for making cellulose in land plants, but this hadn't been confirmed in red algae. Additionally, it was also unclear if the system surrounding cellulose production was similar between land plants and red algae. Using several approaches, I identify a CESA from *Calliarthron tuberculosum* and show that the process of cellulose synthesis has major differences between the red alga *Calliarthron tuberculosum* and land plants. This research widens our understanding of the basic machinery needed to make cellulose, and provides a deeper understanding of this globally important process.

Preface

In Chapter 2, Dr. Cheong Xin Chan (University of Queensland, Australia) provided both annotations of *C. tuberculosum*'s transcriptome with KEGG pathways and transcriptome support with genome data. I designed and conducted all subsequent bioinformatics analyses including homolog searches, sequence characterization, and phylogenetic analysis.

Chapter 3 is based on work conducted in Dr. Jochen Zimmer's lab (University of Virginia, USA) that also gifted the vectors and reagents used in this chapter. Dr. Purushotham Pallinti and Dr. Justin Acheson provided advice in the design of the *CtCESA1*-pACEBac1 construct. Dr. Jochen Zimmer provided advice during CtCESA1 protein purification and assay. Dr. Ruoya Ho provided training for CtCESA1 expression in insect cell systems and helped with cell culturing. Dr. Purushotham Pallinti provided hands on training throughout CtCESA1 protein expression, purification, and enzyme assays. I conducted the experiments presented in this chapter.

In Chapter 4 the pDONR-P4P1R and pDEST-501 vectors were gifted from Dr. George Haughn's lab (University of British Columbia, Canada). I built the vector constructs and conducted all experiments in this chapter.

Dr. Patrick Martone and Dr. Lacey Samuels provided advice throughout the design and analysis stages of this thesis. I conducted all data analysis in this thesis.

Table of Contents

List of Tables

List of Figures

List of Symbols

^{3}H	Tritium
~	Approximately
°C	Degree celsius
∑	Summation
µ	Micro
m	Milli
α	Alpha
β	Beta
Ø	Nothing
cm	Centimeter
Ci	Curie
g	Gravity force
h	Hour
k	Thousand
kDa	Kilodalton
L	Liter
nm	Nanometer
m	Meter
M	Mole
x	Times
V	Volt

List of Abbreviations

At	*Arabidopsis thaliana*
ANOVA	Analysis of variance
AOX1	promoter of *alcohol oxidase 1*
BIC	Bayesian information criterion
BLAST	Basic local alignment search tool
BLASTP	Basic local alignment search tool (amino acid sequence)
BSA	Bovine serum albumen
CC	Companion of cellulose synthase protein
CHS	Chitin synthase protein
CMU	Cellulose synthase-microtubule uncoupling protein
COB	COBRA protein
COBL	COBRA-like protein
CSI	Cellulose synthase interacting 1/ POM-POM2 protein
CTL	Chitinase-like1/ POM-POM1 protein
BCSA	Bacterial cellulose synthesis subunit A protein
BCSB	Bacterial cellulose synthesis subunit B protein
BS	Ultrafast bootstrap
BSA	Bovine serum albumen
CBM	Carbohydrate-binding domain
CHS	Cholesteryl hemisuccinate
Col	Colony

Col	Columbia *Arabidopsis thaliana* ecotype
CSC	Cellulose synthase complex
Ct	*Calliarthron tuberculosum*
DDM	Dodecyl maltoside
DH10	*E. coli* cells that harbor the acceptor baculoviral DNA, the bacmid, and the transposase required for gene integration into the bacmid.
DPM	Disintegrations per minute
FF-EM	Freeze fracture electron microscopy
FPLC	Fast protein liquid chromatography
FT	Flow through
GDN	Glyco-diosgenin
GFB	Gel filtration buffer
GM	Germination media
GmR	Gentamycin resistance
GPI	Glycosyl-phosphatidyl inositol
GT	Glucosyltransferase
HAS	Hyaluronan synthase protein
HIS	Histidine
HMMER	Hidden markov model
irx	Irregular xylem
KEGG	Kyoto encyclopedia of genes and genomes
KOR	KORRIGAN protein
LacZ	β-galactosidase protein used in blue/white screening

LDAO	Lauryldimethylamine-N-oxide
LMNG	Lauryl maltose neopentyl glycol
L+C	Lauryl maltose neopentyl glycol + cholesteryl hemisuccinate
MES	2-N-morpholinoethanesulfonic acid
ML	Maximum likelihood
MRB	Membrane resuspension buffer
MS	Murashige-Skoog media
NaC	Sodium cholate
NCBI	National center for biotechnology information
NOD	N-acetylglucosaminyltransferase protein
OD	Optical density
Ori	Bacterial origin of replication
P0, P1…	Pass number 0, 1, 2 etc. that refers to the number of overall viral propagations
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
P-CR	Plant conserved region
PSI	Pound-force per square inch, measurement of pressure
Put	Putative
PBS	Protein buffer solution
PDB	Protein data bank
PCW	Primary cell wall
Pel	Pellet
Polh	Promoter of the baculoviral polyhedron protein-encoding gene

prc1	*procuste* allele 1
ProCESA7	Promoter of *CESA7*
Ptt	*Populus tremula* × *tremuloides hybrid*
Py	*Pyropia yezoensis*
RPM	Rotations per minute
SCW	Secondary cell wall
SEC	Size exclusion chromatography
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SF9	Clonal isolate from *Spodoptera frugiperda* Sf21 cells
STL	STELLO
Sup	Supernatant
SV40	Viral terminator sequence
T0, T1, T2…	Generation of transformed plant, T0 refers to originally transformed plant, T1 onward refers to subsequent propagated generations
TBS-T	Tris-buffered saline and TWEEN detergent
TEV	Tobacco etch virus
TLC	Thin layer chromatography
TM	Transmembrane
Tn7	Transposon site named Tn7
T-DNA	Transfer-deoxyribonucleic acid
UDP	Uridine diphosphate
ProUBQ10	Promoter of *Ubiquitin 10*

W	Wash
WT	Wild type
YEPD	Yeast extract peptone dextrose
YNBM	Yeast nitrogen base with ammonium sulfate and 5% methanol
Znf	Zinc finger domain
35S	Promoter of the cauliflower mosaic virus sequence

Acknowledgements

I am grateful for all the mentors who have supported my academic endeavors. In particular I would like to thank my supervisors Dr. Patrick Martone and Dr. Lacey Samuels and committee members Dr. George Haughn and Dr. Patrick Keeling and additional examiner Dr. Harry Brumer. I would also like to thank Dr. Jochen Zimmer for the opportunity to conduct experiments in their lab, and Dr. Purushotham Palinti, Dr. Ruoya Ho, and Dr. Justin Acheson for their hands on guidance and support.

I would like to thank members of the Samuels Lab, Martone Lab, and Zimmer Lab for their support in this project. I would also like to thank Dr. Gill Dean, Mendel Perkins, Dr. Yoshi Watanabe, and Dr. Miranda Meents for their advice and guidance.

Finally, I would like to thank my Chun, Bao, Dan, and Rowan, you have allowed me to endure and envision beyond what I thought I was capable of.

Funding for this research was provided from NSERC and UBC Affiliated Fellowship scholarships.

Chapter 1: Introduction

Cellulose is a major component in the cell walls of many land plants and red algae. The presence of cellulose reinforces the cell wall, increasing strength and structural integrity. The process of cellulose synthesis is relatively well characterized in land plants and bacteria, with significant biochemical advances within the past two decades. However, cellulose synthesis within red algae remains largely unexplored. In this thesis, I focus on the calcifying red alga *Calliarthron tuberculosum.* I identify and characterize their cellulose synthases (CESAs) and briefly examine accessory proteins involved in red algal cellulose production.

1.1 Coralline red algae

Rhodophyta, or red algae, is a photosynthetic group originating from an ancient endosymbiotic event with a cyanobacterium, resulting in a divergent lineage from green algae and land plants (Figure 1.1) (Qiu et al., 2015; Burki et al., 2016). There are two major taxonomic classes of red algae, the Florideophyceae and the Bangiophyceae (Yoon et al., 2006). Within the Florideophyceae, one distinctive group of red algae is the Corallinaceae (Borowitzka and Vesk, 1978), whose members calcify by depositing calcium carbonate into their cell walls. Because of their ostensible similarity to calcifying corals, we call them “coralline algae.”

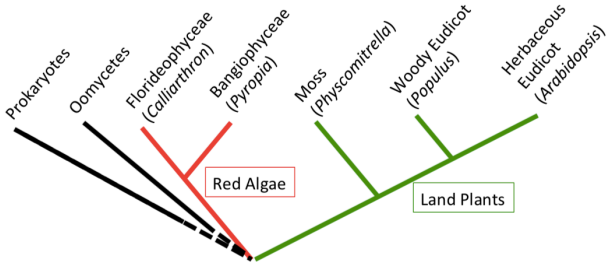

Figure 1.1 Relationships of major lineages described in this thesis, as shown in a cladogram. The major group is indicated with representative organisms in brackets.

Coralline red algae are major contributors to marine communities. They cement together reefs from the tropics to the pacific (for a review see Steneck & Martone, 2007). The presence of coralline red algae can also promote the recruitment of other invertebrates such as corals or molluscs (Kitamura et al., 2007; Williams et al., 2008). Research into coralline red algae is becoming more accessible with increased genomic and transcriptomic resources available for fleshy red algae such as *Pyropia yezoensis* (Brawley et al., 2017), *Chondrus chrispus* (Collén et al., 2013), and the calcifying red algae *Calliarthron tuberculosum* (Bi, Liu, Zhao, & Du, 2016; Chan et al., 2011; and an unpublished transcriptome available in the Martone Lab). These large datasets provide a starting point for a reverse genetics approach to identifying genes of interest.

1.1.1 Genus: *Calliarthron*

Calcifying red algae in the genus *Calliarthron* have garnered increased attention as a common coralline species found on rocky intertidal shores and subtidal regions along the pacific coast, from Mexico to Alaska (Konar and Foster, 1992; Gabrielson et al., 2011). *Calliarthron* are

upright "articulated" coralline algae that have a basal crust with fronds that extend upward. The calcified body of *Calliarthron* is divided into segments, termed intergenicula, separated by joints, or genicula. This system of segments and joints helps these organisms bend, and not break, in response to intertidal wave impacts (Martone and Denny, 2008; Denny et al., 2013). *Calliarthron* are long-lived algae that cover rocks in the wave-swept intertidal zone for approximately 3-10 years (Martone, 2010; Fisher and Martone, 2014). This continued presence in enduringly harsh conditions has led researchers to investigate organisms in the genus *Calliarthron* and their tissue properties, which is, in part, conferred by their cell walls.

1.2 The cell wall components: a brief overview in embryophyte land plants and red algae

Cell membranes can be strengthened and supported by the cell wall, a framework of sugars, proteins, and other macromolecules that encase the cell. Cell walls are found almost universally in various major lineages within the Archaeplastida (photosynthetic lineages that arose from endosymbiotic events with cyanobacteria), including seed-bearing, embryophyte, land plants and red algae. The number of cell walls and their composition can vary between organisms, cell types, developmental stages, environments, and species to fulfill different physiological and structural roles (Popper, 2008). Below I discuss the cell walls of land plants and red algae, focusing on *Arabidopsis thaliana* and *Calliarthron tuberculosum*.

1.2.1 Embryophyte cell walls

Most land plants produce two developmentally distinct cell walls, the primary and secondary cell wall. The primary cell wall (PCW) is deposited first as the cell is expanding and is found more or less ubiquitously in plants. In the eudicot *A. thaliana*, the PCW is composed of a cellulose microfibril skeleton embedded in a space-filling matrix of glycoproteins, pectic polysaccharides, and hemicelluloses such as xyloglucans (Lampugnani et al., 2018; Polko and Kieber, 2019). These PCWs are relatively thin and flexible with randomly orientated cellulose (Cosgrove, 2005; Cosgrove, 2018). Their polysaccharide composition and cellulosic characteristics allow the PCW to be malleable for both isotropic and anisotropic growth.

Later in development, specialized cell types lay down an additional thicker secondary cell wall (SCW) (for a review see Meents, Watanabe, & Samuels, 2018). Thickening of the SCW is one cellular strategy used to impart added strength. In most land plants, SCWs are present in a variety of cell types, some of which include fibers and the cells of water-conducting xylem tissues. The SCWs in these specific cell types necessarily include increased cellulose content, the hemicellulose xylan, and the polyphenolic polymer lignin (reviewed in Meents, Watanabe, and Samuels 2018). Deficiencies in SCW cellulose (Turner and Somerville, 1997), lignin (Jones et al., 2001), or xylan (Persson et al., 2007a) are accompanied by deformations of SCWs, collapsed or irregularly shaped cells, and weakened plant tissue (reviewed in Turner, Taylor, and Jones 2001), highlighting the importance of SCWs in tissue integrity.

1.2.2 Red algal cell walls

Many macroscopic red algae's cell walls are also made of skeletal polysaccharides filled with a matrix containing polysaccharides and proteins. These structural polysaccharides tend to be cellulose (Martone et al., 2019) or hemicelluloses (Lechat et al., 2000) such as xylans (reviewed in Hsieh and Harris 2019; Lahaye et al. 2003) or mannans, with matrix polysaccharides such as the sulfated galactans, carrageenans (Usov et al., 1997; Martone et al., 2010) and agarans (Flores et al., 2000). In some red algal groups, such as the Corallinophycidae (Peña et al., 2020), to which *C. tuberculosum* belongs, calcium carbonate is also deposited into their cell walls (Borowitzka, 1977; Nash et al., 2019). Although the mechanism of calcification has not been elucidated in red algae, the initial formation, or nucleation, of calcium carbonate crystals is hypothesized to occur on the cell wall polysaccharides. Observational electron microscopy surveys of calcifying red algae propose an intertwining role of cellulose extrusion into the cell wall and calcification (Nash et al., 2019). *In vitro* measurements of calcium carbonate formation showed that polysaccharides with a more neutral charge had a lower thermodynamic barrier to calcium crystal formation (Giuffre et al., 2013). For example, they found that alginate, with an overall negative charge, was worse at forming calcium carbonate. Considering the polysaccharide composition of many coralline red algal cell walls, we would expect nucleation to occur on the less charged structural polysaccharides such as cellulose. Indeed, *in vitro* studies have shown that calcium crystals can nucleate on cellulose (Granja et al., 2001; Matahwa et al., 2008), supporting the hypothesis that cellulose may act as the site of calcium carbonate formation.

Currently, no clear distinction is made between PCWs and SCWs of most algae. However, in recent years both PCWs and distinct SCWs have been identified in some calcifying red algae such as *Calliarthron cheilosporioides* (Martone, 2007; Martone et al., 2009), the sister species to *Calliarthron tuberculosum* (Gabrielson et al., 2011). Organisms in the genus *Calliarthron* initially deposit a thin PCW with higher levels of amorphous cellulose (Martone, 2007; Martone et al., 2019). As the alga develops, the cell walls continuously thicken with the deposition of a SCW (Martone, 2007). Unlike the PCW, this SCW contains G units of lignin, which was previously thought to only be in land plants, and increased levels of crystalline cellulose (Martone et al., 2009; Martone et al., 2019).

1.3 The native presence of cellulose in land plants and red algae

Cellulose, a linear homopolymer of β-1,4-linked glucose, is an important functional component in land plants, such as *A. thaliana*, and the red alga *C. tuberculosum*. Despite developmental and compositional differences between the PCW and SCW in land plants, a common feature of both wall types in land plants is the predominant presence of cellulose functioning as the major load-bearing component (McFarlane et al., 2014). PCWs are generally composed of 15 % – 40 % cellulose (reviewed in Cosgrove and Jarvis 2012; McNeil et al. 1984), while the SCWs are typically composed of 40 % – 50 % cellulose in woody plants (Timell, 1967), and *A. thaliana* whole stem tissues have upwards of 40 % cellulose (Kumar and Turner, 2015). This trend of increasing cellulose content is also present in *C. tuberculosum*'s cell walls, where cellulose is continuously deposited in random orientations along the cell wall and comprises ~ 8 % in primary cell walls and ~ 22 % in secondary cell walls (Martone et al., 2019).

While cellulose from *A. thaliana* and *C. tuberculosum* may be made of the same glucan chains, the interactions among these glucan chains and their macromolecular characteristics can differ. Cellulose can be thought of as a rope composed of multiple individual glucan chains tied together through hydrogen bonding and van der Waals forces. Cellulose forms a crystalline material and its crystalline properties are determined by this patterning of hydrogen bonding (Nishiyama et al., 2003). Glucose chains that associate parallel to each other are called Cellulose I, which can be further categorized into two forms or allomorphs, Iα and Iβ, that differ by their hydrogen bonding pattern and subsequent cellulose crystallinity (Nishiyama et al., 2003). Land plant cellulose, such as cotton, have high levels of Iβ (60% – 70%) (Atalla and VanderHart, 2011). X-ray diffraction and crystalline cellulose labelling with gold particle conjugated cellobiohydrolase I has also demonstrated that cellulose I exists in the red alga *Erythrocladia subintegra* (Okuda et al., 1994). In *C. tuberculosum*, antibody labelling of cell walls with carbohydrate binding module 3 (CBM3) and CBM28, that preferentially bind flat crystalline cellulose and single chains of amorphous cellulose respectively, showed a mix of non-crystalline and crystalline cellulose throughout their cell walls (Martone et al., 2019).

Changes to cellulose quantities or crystallinity can severely impact the cell wall's mechanical properties (for review of these properties in land plants, see Wang et al., 2016). Between land plants such as *A. thaliana* and red algae such as *C. tuberculosum,* there are clear differences in the amount of cellulose present and differences in the crystallinity of their cellulose.

1.4 Cellulose biosynthesis

1.4.1 Characteristics of a cellulose synthase (CESA) enzyme

The majority of what we know about cellulose production comes from work done in bacteria and land plants. Cellulose is produced at the plasma membrane (PM) by a type two glycosyl-transferase protein, the cellulose synthase enzymes (CESAs) (for a review see Anderson and Kieber 2020; McFarlane, Döring, and Persson 2014). CESAs add UDP-glucose substrate (Omadjela et al., 2013) from the cytosol onto the growing β-1,4-linked glucan chain polymer extruded into the apoplastic space (Morgan et al., 2013). This addition of glucose molecules is thought to push the CESA through the plasma membrane, guided along microtubule tracks. This phenomenon has been visualized by fluorescently tagged CESAs traversing the plasma membrane (Paredez et al., 2006). These growing cellulose chains then aggregate into crystalline or para-crystalline cellulose through regularly patterned inter and intra-molecular hydrogen bonding facilitated by the hydroxyl groups of the glucose constituents (Nishiyama et al., 2003).

Each individual CESA protein produces one glucan chain as demonstrated in protein modelling of the Cotton *GhCESA* (Sethaphong et al., 2013), structural studies of the active bacterial cellulose synthase A/B complex (BCSA/BCSB) (Morgan et al., 2013; Morgan et al., 2016), biochemical studies of BCSA/BCSB (Omadjela et al., 2013), structural studies of the land plant *Populus tremula x tremuloides* PttCESA8 (Purushotham et al., 2020), and biochemical studies of the PttCESA8 and *Physcomitrella patens* PpCESA5 (Purushotham et al., 2016; Cho et al., 2017).

The aforementioned structural studies have provided a clear idea of the overall protein structure and important sequence characteristics of a given *CESA* sequence. Each *CESA* has a central active site, the glucosyltransferase (GT) domain, which is flanked by seven transmembrane domains. High sequence conservation exists between the active sites of distantly related *CESA*s as demonstrated by the early identification of cotton *CESA*s from sequence library screens using the bacterial *BCSA* sequence as a probe (Pear et al., 1996). This *CESA* GT domain has several conserved motifs. The QXXRW and D,D,D, motifs are required for co-ordination of the uridine diphosphate (UDP) glucose substrate and the growing glucan chain (Morgan et al., 2013). In addition to the functionally required GT and transmembrane domains, there are major lineage-specific sequence variations. The bacterial *CESA*s have a unique C-terminal PiL-Z domain that senses the signalling molecule cyclic-di-GMP (Amikam and Galperin, 2006) to trigger cellulose production. Land plant *CESA*s have unique central hypervariable regions, and plant conserved regions that facilitate CESA-CESA interaction (Purushotham et al., 2020). Land plants CESAs also uniquely have a N-terminal zinc finger (Znf) domain that likely facilities CESA-CESA interactions (Purushotham et al., 2016) as well as CESA-microtubule interactions via an intermediary tethering protein, the cellulose synthase interacting protein (CSI) (discussed further in Ch 1 section 1.4.5) (Bringmann et al., 2012; Purushotham et al., 2020). Oomycete *CESA*s have a predicted pleckstrin homology domain at their N-terminus, but its functional role has not been determined (Fugelstad et al., 2009). In other proteins, pleckstrin homology domains have been demonstrated to play important roles in protein-protein interactions and protein interactions with phospholipid head groups (reviewed in Scheffzek and Welti 2012). Perhaps this pleckstrin homology domain could function in oomycete CESA complex formation. Red algal *CESA*s have a unique predicted carbohydrate-binding family 48 (CBM48) domain at its N-terminus that is

also of underdetermined function in the red algal CESA context (Matthews et al., 2010). This domain is predicted to facilitate cytosolic starch-binding interactions in red algae, though this has not been tested. Clearly, these variations bring into question what functional roles these unique domains may play in cellulose production within the various lineages.

1.4.2 CESA specialization in primary cell wall and secondary cell wall synthesis

All land plants identified to date encode multiple *CESA*s in their genome. For example, cotton has 32 *CESA* genes (Zhang et al., 2015), *Populus* has 17 *CESA* genes (Suzuki et al., 2006), *Arabidopsis* has 10 *CESA* genes (Carroll and Specht, 2011; Little et al., 2018), and *Physcomitrella* has 6 full-length *CESA* genes (Roberts and Bushoven, 2007). Mutational studies in *A. thaliana* have demonstrated that their *CESA* sequences can be categorized into six distinct groups that function non-redundantly in either the primary cell wall (PCW; *CESA1, 3, 6/2/5/9*) (Persson et al., 2007b) or secondary cell wall (SCW; *CESA4, 7, 8*), but not both. Moreover, these *CESA*s have distinctly controlled expression during either PCW or SCW synthesis (Watanabe et al., 2018).

Three distinct groups of CESAs (CESA4, CESA7, and CESA8) function in SCW cellulose synthesis. *CESA4*, *7*, and *8* are tightly co-expressed with a suite of SCW genes (Brown et al., 2005). When any of these three are mutated, there are severe defects in the SCW accompanied by major decreases in cellulose, implicating these proteins in SCW specific cellulose synthesis (N. G. Taylor, Laurie, and Turner 2000; Neil G. Taylor et al. 1999, 2003) . Moreover, these mutant phenotypes can only be fully rescued by the introduction of their wild type gene, and not

a different wild type SCW *CESA* sequence, suggesting each class of SCW CESAs are required and not functionally redundant (Kumar et al., 2017).

The other three distinct groups of *CESA*s (*CESA 1* (Arioli et al., 1998)*, CESA3,* and the functionally redundant group of *CESA6/2/5/9* (Desnos et al., 1996; Fagard et al., 2000; Desprez et al., 2007)) function in PCW cellulose synthesis. PCW CESAs deposit cellulose throughout the cell wall (Paredez et al., 2006; Desprez et al., 2007; Fujita et al., 2013). Impaired cellulose deposition due to mutations in PCW *cesa*s can cause severe cellular swelling and loss of directional cellular growth (Fagard et al., 2000; Hu et al., 2018), demonstrating the key roles PCW CESAs have in shaping developing plant cells and tissues. Each type of PCW CESA is necessary, and mutants fully defective in either of these three classes of CESAs are gametophyte lethal (Persson et al., 2007b). Mutations in *cesa6* (Desprez et al., 2002) and *cesa3* (Desprez et al., 2007) cause increased resistance to the cellulose synthesis inhibitor, isoxaben, adding evidence to their function in cellulose synthesis.

These classifications as either PCW CESAs or SCW CESAs are consistently identified in phylogenetic studies of vascular land plant *CESA*s (Carroll and Specht, 2011; Little et al., 2018). This PCW and SCW classification has also independently evolved in non-vascular plants (Roberts and Bushoven, 2007; Carroll and Specht, 2011). However, functional or categorical specialization has not been explored in other lineages outside of the land plants. Intriguingly, red algae are predicted to have only one CESA sequence (Roberts and Roberts, 2009; Matthews et al., 2010). However, the identification of both primary cell walls and secondary cell walls in

Calliarthron (Martone et al., 2009; Martone et al., 2019) raises the question of whether one CESA functions in both primary and secondary cell wall synthesis in the red alga.

1.4.3 CESAs oligomerize and associate into cellulose synthase complexes

CESAs are thought to associate into oligomers and organize into larger multimeric cellulose synthase complexes (CSC), also referred to as terminal complexes (TC). The currently most supported model is that CSCs are composed of hexamers of trimers of CESAs (Figure 1.2) (Nixon et al., 2016; Purushotham et al., 2020; Ramírez-Rodríguez and McFarlane, 2020). In land plants, CESA proteins categorized as either PCW or SCW CESAs associate into trimers within their class group. This class-specific interaction between non-redundant CESAs is supported by co-purification of CESA4, 7, and 8 from plant extracts in *Arabidopsis* (Taylor et al., 2003). Formation of oligomeric trimers is also seen during *in vitro* structural studies of the *Arabidopsis thaliana* CESA1 catalytic domain (Vandavasi et al. 2016) and full-length *Populus tremula x tremuloides* CESA8 (Purushotham et al., 2020).

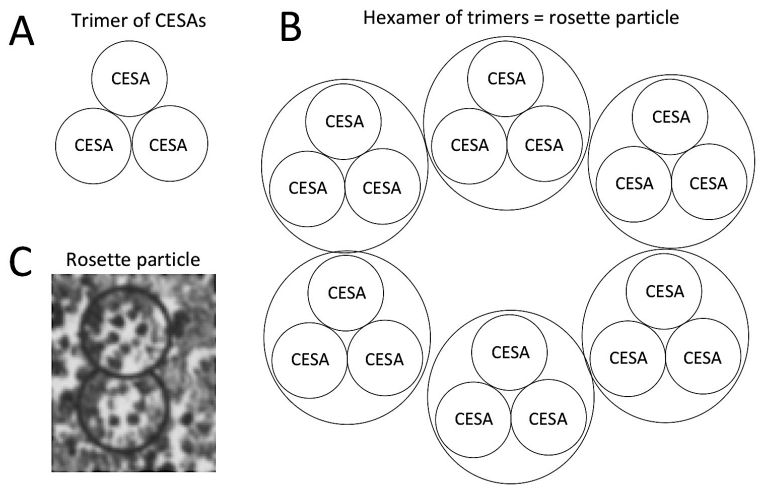

Figure 1.2 Diagram of CESA complex formation. (A) Individual CESAs oligomerize into trimmers. (B) Trimers for the basis of a "lobe" in the rosette shaped hexamer. This hexemer is the land plant cellulose synthase complex (CSC). (C) Two rosette CSCs seen in freeze fracture electron microscopy images (circled; adapted from Figure 3 in Mueller and Brown 1980).

Freeze fracture electron microscopy (FF-EM), has been a critical technique in both the early observations of the CSC complexes and producing a survey of, presumably, CSC complex formation across a variety of cellulose producing lineages. During FF-EM, samples are frozen and fractured to reveal the membrane embedded proteins. Terminal enzyme complex (TC) was coined in reference to the complexes found at the terminus of the cellulose microfibils (Brown and Montezinos, 1976). For this reason, TC and CSC are used interchangeably in the FF-EM related literature.

TC complexes are seen as "rosette" shapes in FF-EM images from land plants such as moss (Rudolph and Schnepf, 1988; Nixon et al., 2016), corn (Mueller and Brown, 1980), and red

mung bean (Kimura et al., 1999) (Figure 1.2). Anti-CESA antibody labelling in FF-EM samples of red mung bean showed that these TC/CSCs are composed of CESAs (Kimura et al., 1999).

Unlike land plant rosette TCs, FF-EM images of Florideophyceae red algae show TCs that form linear arrays and produce cylindrical cellulose (Tsekos, 1993). FF-EM images of red algal TCs are likely associated with cellulose fibrils as shown with x-ray diffraction, discussed earlier in section 1.3 (Okuda et al., 1994). If these linear TCs truly are composed of CESAs in red algae, presumably the red algal CESAs oligomerize to form the larger TCs seen. However, prior to this thesis, no red algal CESAs have been experimentally verified, and biochemical evidence is required to demonstrate oligomerization in this group. Oligomerization is likely essential to the formation of crystalline cellulose. CESAs that lack their N-terminal domain and, therefore likely not able to associate, form amorphous and not fibrillar cellulose (Purushotham et al., 2016). Moreover, the variable TC formation within these distinct organismal lineages raises interesting questions of cellulose formation and quality as terminal complex arrangement is associated with differing shapes of cellulose (Tsekos et al., 1993; Tsekos, 1999; Huang et al., 2020). A more comprehensive array of information on the diversity of CESA sequences will allow us to compare their effects on varying cellulose production and quality.

1.4.4 Cellulose trafficking to the plasma membrane

As a general overview, CESAs are trafficked from the site of synthesis at the ER to the Golgi to their functional location at the plasma membrane (Figure 1.3) (for a review see Meents, Watanabe, and Samuels 2018; Wightman and Turner 2010). CESA trafficking between the Golgi (trans Golgi network/early endosome) and plasma membrane occurs in vesicles termed small cellulose synthase compartments (SmaCCs) (Gutierrez et al., 2009) or microtubule associated SmaCCs (MASCs) (Crowell et al., 2011), depending on their microtubule association. SmaCCs/MASCs delivery at the plasma membrane is thought to localize with microtubules. Fluorescently tagged CESAs consistently show SmaCCs/MASCs co-localized with fluorescently tagged microtubules at the plasma membrane and disruption of these microtubules result in randomized insertion of CESAs (Crowell et al., 2009; Gutierrez et al., 2009). This outlined CESA trafficking pathway (Figure 1.3) is supported by fluorescent imaging of tagged PCW CESAs (Paredez et al., 2006; Crowell et al., 2009; Gutierrez et al., 2009) and SCW CESAs (Watanabe et al., 2015) that consistently shows bright localizations in doughnut-shaped Golgi bodies, SmaCCs/MASCs, and at the plasma membrane. Once at the plasma membrane, CESAs can also be recycled back into the cell via clathrin mediated endocytosis (CME) (Figure 1.3), as visualized by co-localized fluorescently tagged CESAs and clathrin large subunits during endocytosis (Miart et al. 2014) and CESA interaction with additional protein machinery associated with CME (Bashline et al., 2013; Bashline et al., 2015; Sánchez-Rodríguez et al., 2018).

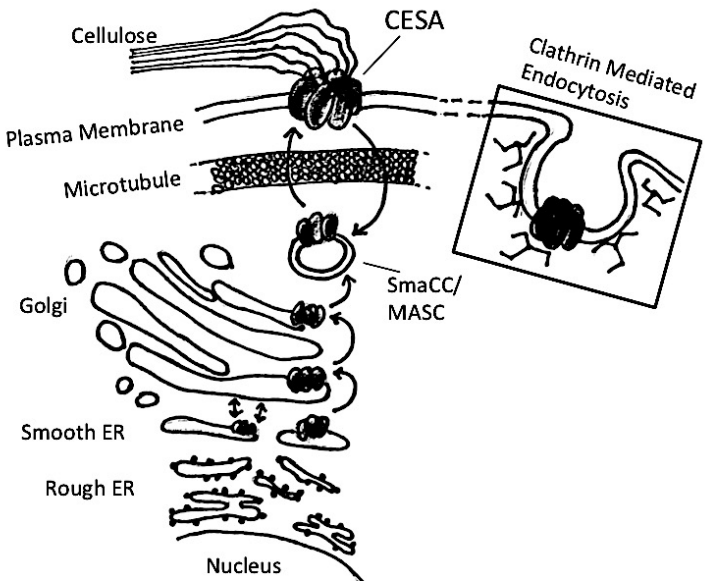

Figure 1.3 Basic schematic of cellulose synthase complex (CSC) trafficking in land plants. CSCs are trafficked from the ER to the Golgi. CSC trafficking from the Golgi to their functional locations at the microtubule-localized domains on the plasma membrane is mediated by SmaCCs/MASCs. Inset shows clathrin mediated CSC endocytosis during recycling into the cell. This diagram is brief for simplicity and does not show all known components of CSC trafficking.

This elucidated CESA trafficking pathway is solely rooted in studies from the land plant *A. thaliana* while the trafficking of CESAs in red algae is entirely unexplored. Perhaps overall similarities exist in their trafficking of cell wall components. Treatment of *A. thaliana* seedlings (Gutierrez et al., 2009) or tobacco pollen tubes (Cai et al., 2011) with the protein secretion inhibitor Brefeldin A (BFA) (Langhans et al., 2011) causes intracellular CESA accumulation in intracellular aggregations termed BFA bodies. Similarly, in an indirect measurement, BFA treatment of monospores from the red alga *P. yezoesis* resulted in decreased deposition of cell

wall material as visualized with calcofluor white, which stains indiscriminately against cellulose/chitin (Li et al., 2008). Perhaps red algal cell wall components are trafficked in a similar manner with land plants, where the lack of cellulose observed in the red alga may be due to a BFA induced inhibition of CESA trafficking to the plasma membrane.

1.4.5 Non-CESA accessory proteins during cellulose synthesis

Apart from the CESA machinery, there are multiple additional, or accessory, proteins that facilitate cellulose synthesis in land plant systems (Figure 1.4; Table 1.1). Here I describe in detail the cellulose synthesis accessory proteins that I address in this thesis.

Before insertion into the plasma membrane, CESAs are organized into CSC complexes. Golgi-localized STELLO proteins are thought to affect this CSCs assembly (Zhang et al., 2016). *A. thaliana* mutated in *stello* family proteins had altered localization of CESA3 in the Golgi and reduced delivery of CESAs to the plasma membrane. Moreover, *stello* mutants have reduced cellulose and show defects broadly across both PCW and SCW phenotypes. These data suggest that STELLO proteins directly interact with both PCW and SCW CESAs in Golgi localized CSC formation and secretion to the plasma membrane (Figure 1.4; Table 1.1). However, the direct mechanism of STELLO's action on CESAs is currently unknown.

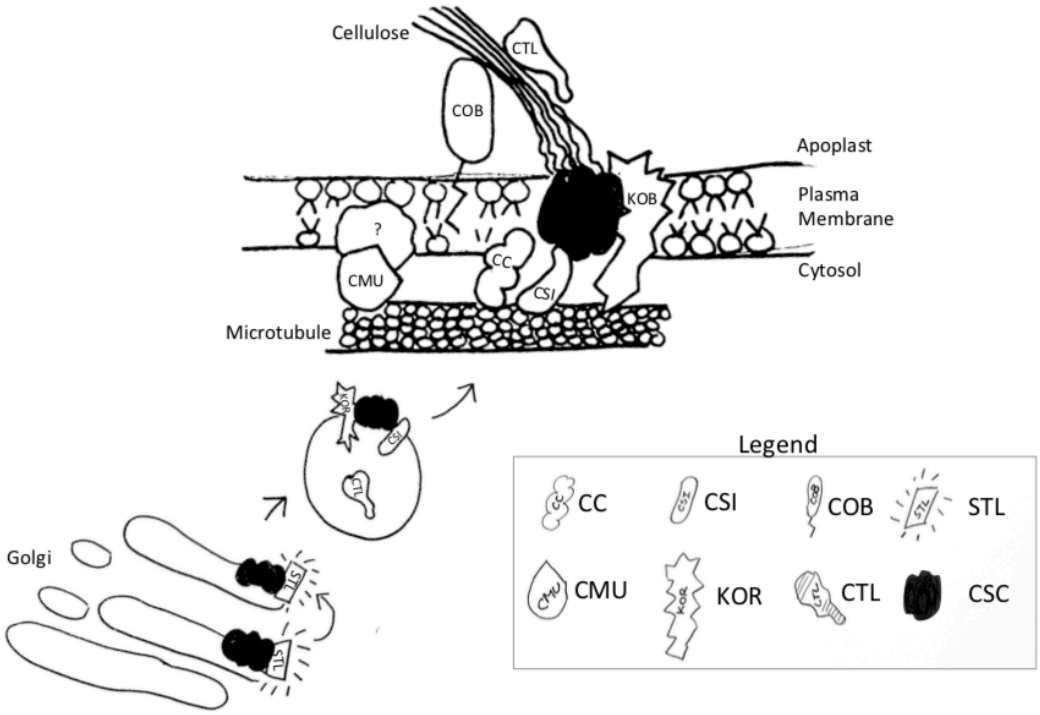

Figure 1.4 Diagrammatic summary of cellulose synthase accessory proteins listed in Table 1.1. (Acronyms are as follows: CC - Companion Of Cellulose Synthase; CSI - Cellulose Synthase Interacting 1/ POM-POM2; COB – COBRA; STL – STELLO; CMU - Cellulose Synthase-Microtubule Uncoupling Protein; KOR – KORRIGAN; CTL - Chitinase-Like1/ POM-POM1; CSC – Cellulose Synthase Complex).

Once at the plasma membrane, CESAs track along microtubules in *A. thaliana* cells as they produce cellulose. The Cellulose Synthase Interacting 1/ POM-POM2 (CSI) protein directly mediates this CESA-microtubule interaction. CSI binds to CESAs in pull-down assays (Purushotham et al., 2020) and co-localizes with microtubules (Gu et al., 2010). In *csi*/*pompom2* mutants, CESAs no longer co-localize with microtubules (Bringmann et al., 2012), supporting their role in mediating the CESA-microtubule interaction (Figure 1.4; Table 1.1).

As CSCs synthesize cellulose, the addition of each UDP-glucose monomer propels the complex along the plasma membrane. This CSC movement likely imposes a destabilizing force against the microtubules to which they are tethered. The Cellulose Synthase-Microtubule Uncoupling Protein (CMU) is thought to stabilize the microtubule and resist the force generated by the CSC movement by "pinning" the microtubule in place to the plasma membrane. Pull-down assays and fluorescent co-localization experiments show immobile CMUs bind microtubules and indirectly tether them to the plasma membrane (Liu et al., 2016) through an unknown association (Figure 1.4; Table 1.1). In *cmu* mutants, the microtubules are less stable and bend in shape. This is likely due to transient associations with CESAs and their imposed forces, resulting in the slightly more erratic movement of the CESA proteins, supporting the model that CMU proteins have a stabilizing influence in microtubule-mediated CESA movement.

Cellular stress can cause temporary microtubule depolymerization and endocytosis of CSCs from the plasma membrane (Endler et al., 2015). Companion Of Cellulose Synthase (CC) proteins are thought to stabilize microtubules under salt stress to promote CESA recovery to the plasma

membrane and resume cellulose synthesis in these conditions (Figure 1.4; Table 1.1) (Endler et al., 2015). *cc* mutants under salt stress cannot permanently stabilize microtubules, and CESAs show low levels of reinsertion into the plasma membrane. CC and SmaCCs/MASCs co-localize after stress exposure, where the CC's N-terminus binds and stabilizes microtubules (Kesten et al., 2019), and CSCs return to the plasma membrane (Endler et al., 2015).

Several other proteins are thought to stabilize cellulose shortly after its production, but their direct mechanisms are less clear. These proteins include KORRIGAN (KOR), Chitinase-Like1/ POM-POM1 (CTL), and COBRA (COB). KOR is an endo-1,4-β-glucanase located in the plasma membrane (Nicol et al., 1998) that co-localizes with both PCW CESAs (Vain et al., 2014) and the SCW CESA4 and CESA8 (Mansoori et al., 2014). In *kor* mutants, cell wall structure is compromised and there is a reduction in cellulose content (Szyjanowicz et al., 2004). Though there is no direct evidence elucidating KOR's role, this protein is hypothesized to release stress created by cellulose synthesis by selectively hydrolyzing glucose bonds (Figure 1.4; Table 1.1). CSC velocities are decreased and individual CSC movement gradually slows down over time in *kor* mutants (Paredez et al., 2008). Perhaps the absence of KOR causes increased rotational strain during cellulose production that compounds to hamper cellulose production over time, though this hypothesis requires more functional evaluation.

After cellulose is produced and extruded into the cell, it is unclear what additional factors contribute to cellulose crystallization. CTL is proposed to bind and perhaps temporarily stabilize the emerging cellulose to facilitate its crystallization shortly after being produced by the CSC (Figure 1.4; Table 1.1). In *ctl* double mutants, there is a reduction in the percentage of crystalline

cellulose, supporting a role in facilitating cellulose crystallinity (Sánchez-Rodríguez et al., 2012). The wild type CTL can bind cell wall glucans, such as cellulose and the hemicellulose xyloglucan, to varying degrees but does not show hydrolytic activity, supporting its ability to bind cellulose without cutting it. Moreover, CTL co-localizes with CESAs in Golgi and SmaCCs/MASCs during secretion to the plasma membrane but is then exocytosed into the apoplast (Sánchez-Rodríguez et al., 2012). In this fashion, the CTL is likely released close to the CSC machinery and its emerging cellulose chain.

COB is an extracellular facing protein anchored to the plasma membrane by a post-translational lipid modification, a glycosyl-phosphatidyl inositol (GPI) moiety (Roudier et al., 2005) (Figure 1.4; Table 1.1). COB is hypothesized to bind and stabilize cellulose in the extracellular, or apoplastic, space. *cob* mutants were first discovered by their abnormal root cell expansion (Benfey et al., 1993), likely due to its decreased levels of crystalline cellulose in root cells (Schindelman et al., 2001; Roudier et al., 2005). COB-like (COBL) proteins, which share high sequence similarity with COB (Roudier et al., 2002), have a CBM2 domain that binds crystalline cellulose in carbohydrate microarray analysis (Liu et al., 2013). COB-like4 (COBL4) is thought to function during secondary cell wall cellulose synthesis specifically (Sato et al., 2010b). Abnormal secondary cell wall phenotypes caused by mutant *cobl4* are normally complemented by the introduction of its wild type gene. However, when the predicted cellulose binding CBM2 domain of the COBL4 sequence is mutated, it no longer complements the *cobl4* mutant in rice (Sato et al., 2010a). This suggests that COBL4 likely binds cellulose via its CBM2 domain to properly facilitate cellulose production. The binding properties of the COBL4's CBM2 domain to crystalline cellulose was also confirmed *in vitro* (Liu et al., 2013). Perhaps COB's CBM

domain can also bind cellulose, though this has not been formally tested, and facilitates cellulose crystallinity through an unknown mechanism.

These accessory proteins clearly affect cellulose synthesis as defects in these proteins results in severe cell wall phenotypes, abnormal cellulose, or CESA behaviors. It is unknown whether these mechanisms exist in red algae. One bioinformatics survey argues against any similarities, suggesting that none of these cellulose synthesis accessory proteins exist in red algae (Lampugnani et al., 2019). However, this study surveyed only one red algal genome. The lack of queried sequence recovery raises several issues, including potential quality of the genome used, e.g. sequence fragmentation. Further bioinformatics surveys, with greater inclusion of red algal taxa, could help elucidate any possible parallels of mechanisms surrounding cellulose synthesis in land plants and red algae.

Table 1.1 Accessory proteins associated with cellulose synthesis from initial study that implicated protein involvement with cellulose production is indicated.

Acronym	Name	Function	Citation
STL	STELLO	Required for CSC assembly in the Golgi, as well as trafficking to the plasma membrane. Molecular mechanism unknown.	(Zhang et al., 2016)
CSI	Cellulose Synthase Interacting 1/ POM-POM2	Intermediary protein that tethers CESAs to microtubules.	(Gu et al., 2010)
CMU	Cellulose Synthase-Microtubule Uncoupling Protein	Hypothesized to stabilize microtubules by opposing forces generated by CSC.	(Liu et al., 2016)
CC	Companion Of Cellulose Synthase	Maintains cellulose production under salt stress by promoting microtubule polymerization and stabilization.	(Endler et al., 2015)
KOR	KORRIGAN	β-1,4-endo-glucanase hypothesized to release stress created by cellulose synthesis by selectively hydrolyzing glucose bonds.	(Nicol et al., 1998)
CTL	Chitinase-Like1/ POM-POM1	Hypothesized to bind and stabilize the emerging cellulose in the apoplastic space. Molecular mechanism unknown.	(Sánchez-Rodríguez et al. 2012)
COB	COBRA	Glycosylphosphatidylinositol (GPI)-anchored protein hypothesized to bind and stabilize cellulose.	(Benfey et al., 1993; Schindelman et al., 2001)

1.5 Research objectives

The aim of this thesis is to elucidate the mechanisms underlying cellulose synthesis in the red algae *Calliarthron tuberculosum*. The specific questions of this thesis are:

1. How does *C. tuberculosum* produce cellulose, i.e. what gene represents the functional cellulose synthase (CESA) in this red alga?
2. How does cellulose synthesis differ between red algae and other organismal lineages?

To address these two questions, the four main objectives of my thesis were:

1. Identify putative *CESA* sequence(s) from *C. tuberculosum* and determine their relationships with other known *CESA*s.
2. Biochemically characterize the putative *C. tuberculosum* CESA's glucan synthase activity.
3. Test if *C. tuberculosum*'s CESA can fulfill similar functional roles with some of the primary and secondary cell wall CESAs from the flowering land plant *Arabidopsis thaliana*.
4. Determine if any putative cellulose synthesis accessory proteins are found in *C. tuberculosum*.

Chapter 2 addresses objectives one and four by looking for homologous genes in *C. tuberculosum*'s transcriptome. *C. tuberculosum*'s putative *CESA* sequences were further characterized using various bioinformatics methods, and its relationship with other known *CESA*s was elucidated in gene tree analysis.

Chapter 3 addresses objective two by purifying *C. tuberculosum*'s CESA1 protein using heterologous expression in yeast and insect cell systems and testing the protein for glucan synthase activity in *in vitro* enzymatic assays. Additionally, the enzymatic activity of *C. tuberculosum*'s CESA1 was tested together with the bacterial *Rhodobactor sphaeroides* BCSA/BCSB complex and the land plant *Populus tremula x tremuloides* CESA8 to view potential broad differences in activity between these lineages.

Chapter 4 addresses objective three by heterologous expression of *C. tuberculosum*'s CESA1 protein in *A. thaliana* primary and secondary *cesa* mutants. I tested whether the algal gene could act in the land plant system by looking for functional rescue of the *cesa* mutant phenotypes.

The data presented in this thesis presents the first experimentally characterized *CESA* from a red alga. Using a variety of bioinformatics and heterologous systems, I highlight major differences in cellulose synthesis between the land plant and red algal lineages. In summary, I identify the best gene candidate for a *C. tuberculosum* CESA, characterize it as a glucan synthase, and present evidence that major aspects of cellulose synthesis have likely evolved divergently between the red algal and land plant lineages.

Chapter 2: Bioinformatic analysis identifies putative cellulose synthases from the red alga *Calliarthron tuberculosum*

2.1 Introduction

CESA sequences have been well characterized and functionally demonstrated in a variety of land plant species such as *Gossypium hirsutum* (Pear et al., 1996), *Arabidopsis thaliana* (reviewed in McFarlane, Döring, and Persson 2014), *Populus tremula x tremuloides* (Purushotham et al., 2016), *Physcomitrella patens* (Cho et al., 2017), and the bacterial BCSA/BCSB complex from *Rhodobacter sphaeroides* (Morgan et al., 2013). In contrast, only two candidate sequences have been identified in fleshy red algae, from *Pyropia yezoensis* (Bangiophyceae) (Roberts and Roberts, 2009; Brawley et al., 2017) and *Griffithsia monilis* (Florideophyceae) (Matthews et al., 2010). CESA encoding sequences have yet to be identified in any calcifying red alga.

In contrast to fleshy red algae, calcifying red algae deposit calcium carbonate into their cell walls (Nash et al., 2019). Some studies propose that cellulose deposition facilitates this calcium carbonate biomineralization in red algae (Nash et al., 2019). *In vitro* studies of cellulose soaked in calcium-rich solutions readily form crystallized calcium phosphate and calcium carbonates (Granja et al., 2001; Matahwa et al., 2008), suggesting that cellulose can act as a scaffold for this mineralization process. As is the case, understanding what proteins may be responsible for cellulose synthesis in calcifying red algae is imperative for future biological studies of calcification in these algae. However, there are limited resources available within this group. Of the calcifying red algae, *C. tuberculosum*'s cellulose content (up to 22 % cellulose) and quality

(a mix of amorphous and crystalline cellulose) have been characterized (Martone et al., 2010; Fisher and Martone, 2014; Martone et al., 2019). *C. tuberculosum* also has both transcriptomic and genomic resources (Chan et al. 2011; Xue et al. unpublished). This foundational research and resources make looking for calcifying red algal *CESA* sequences accessible in *C. tuberculosum*.

Reciprocal best hit search is a common bioinformatic approach to look for orthologous genes of interest (Tatusov et al., 1997). In this approach, two genes are considered orthologous if they can identify each other as top hits in searches against each other's sequence database. Profile hidden Markov model based (HMMER) searches (Finn et al., 2011) can be used to conduct reciprocal sequence searches and find homologous genes. HMMER searches provide a more sensitive approach for homolog detection by assigning probabilities of conserved vs. variable regions from an alignment of *CESA* sequences. However, the sequences used to curate the *CESA* profile may bias the modelled conservation or variation of an amino acid at a given site. As a control for this, the Kyoto Encyclopedia of Genes and Genomes (KEGG) also annotates genes using the HMM-based method (Kanehisa and Goto, 2000), but uses a different composition of sequences for their *CESA* profile.

Cellulose synthases belong to the glycosyl-transferase (GT) 2 family and have several characteristic motifs and domains (for an overview, see Figure 2.1). CESAs are generally around 800-1000 amino acids long (Richmond and Somerville, 2000). Although earlier predictions of most land plant CESA suggest eight TM domains, recent x-ray crystallography of the *Populus tremula x tremuloides* (PttCESA8) demonstrates that there are only seven true TM domains

(Purushotham et al., 2020). The previously annotated fifth TM domain is incorrect, and this domain is actually interfacial. There are two N-terminal and five C-terminal transmembrane domains (Morgan et al., 2013; Sethaphong et al., 2013; Purushotham et al., 2020). There is a centrally located cytosolic domain around 500 amino acids that contains the glucosyltransferase (GT) active site (yellow Figure 2.1B). In this GT domain, the four motifs D,D,D and QXXRW, responsible for binding the uridine diphosphate (UDP) glucose substrate and growing polymer chain, are necessary for catalytic function (Morgan et al., 2013; Kumar and Turner, 2015). Mutations in any of these D residues cause cellulose deficiency (Pear et al., 1996). Protein modelling against the bacterial catalytic cellulose synthase subunit (BCSA) (Morgan et al., 2013), the only available full crystal protein structure for a cellulose synthase, has been critical in predicting structures of the cotton CESA1 (Sethaphong et al., 2013), CESA complex (CSC) formation (Nixon et al., 2016), and guiding structural modelling of the cryo-electron microscopy determined PttCESA8 structure (Purushotham et al., 2020). Similarly, *CtCESA* sequences identified can be modelled against the BCSA crystal structure for structural comparison.

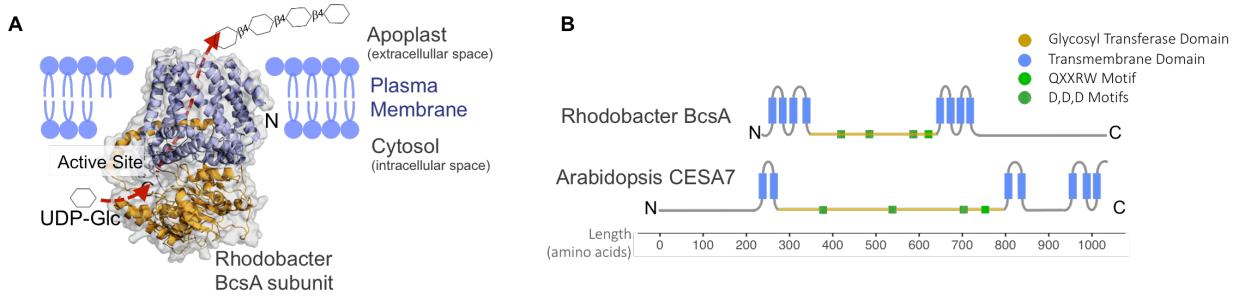

Figure 2.1 Bacterial cellulose synthase crystal structure and bacterial and plant cellulose synthase sequence topology. (A) Diagrammatic representation of cellulose synthesis by the known crystal structure of BCSA (modified from Morgan et al. 2013). Each CESA in the CSC complex takes up UDP-glucose from the cytosol and adds it onto the growing β-1,4-linked glucan chain extruded into the apoplastic space. (B) Topology of the known BCSA from *R. sphaeroides* and CESA7 from *A. thaliana*. Both start with an intracellular N-terminus. The transmembrane domains (blue), glucosyltransferase domain (yellow), and glucose coordinating motifs (green) are indicated. Images are drawn to scale in the x-axis.

*CESA*s have a complicated evolutionary history, with relationships that tend to be well resolved within major evolutionary lineages (i.e. within red algae or within land plants) but not between major lineages (Roberts and Roberts, 2009; Michel et al., 2010; Brawley et al., 2017). Red algal *CESA*s and oomycete *CESA*s, such as the potato pathogen *Phytophthora infestans*, share a uniquely close relationship (Roberts and Roberts, 2009; Michel et al., 2010). In contrast, the fleshy red algal *CESA* sequences and land plant sequences are generally distantly related (Roberts and Roberts, 2009; Brawley et al., 2017). In land plants, functional *CESA*s form one tight clade where both flowering land plant *CESA*s (Little et al., 2018) and moss *CESA*s (Roberts

and Bushoven, 2007; Li et al., 2019) further divide into two groups based on their primary and secondary cell wall classification. However, many of these studies do not include a comprehensive set of red algal *CESA*s or *CESA* sequences from diverse lineages. Understanding the relationship between the red algal *CtCESA*s and the land plant CESAs is necessary to inform later experiments (chapter four) that test their biological compatibilities. This will involve mutant rescue experiments by heterologously expressing the *CtCESA* sequences in *A. thaliana cesa* mutants. To visualize if *A. thaliana* could be a biologically compatible surrogate to study the algal *CtCESA*'s, I also examine the relationships between the *CtCESA* and other known *CESA*s in a gene tree analysis. With this, I can see whether the *CtCESA* clusters with *A. thaliana* PCW or SCW *CESAs* to inform further experiments. However, this may not be as straightforward as replacing one CESA machinery for the other.

In land plants, proper cellulose synthesis requires the coordinated action of multiple accessory proteins, including STELLO (STL), Cellulose Synthase Interacting 1/ POM-POM2 (CSI), Cellulose Synthase-Microtubule Uncoupling Protein (CMU), Companion Of Cellulose Synthase (CC), KORRIGAN (KOR), Chitinase-Like1/ POM-POM1 (CTL), and COBRA (COB) (as discussed in Ch 1). These proteins have been found to facilitate CESA trafficking, CESA stabilization, CESA-microtubule interactions, and cellulose crystallization in the land plant *A. thaliana*. It is unclear whether some or all of the mechanisms are universally required for proper cellulose synthesis or present in red algae, such as *C. tuberculosum.* Perhaps some modalities of cellulose synthesis have evolved independently or divergently between cellulose producing lineages.

The objective of this bioinformatic analysis was to identify putative cellulose synthase (*CESA*) encoding gene(s) from the *Calliarthron tuberculosum* transcriptomic dataset using sequence similarity to known *CESA*s from other species. Once *C. tuberculosum*'s *CESA* (*CtCESA*) candidates have been identified, additional sequence criteria and protein modelling can be applied to elucidate their identity as *CESA*s. The relationship of *CtCESA*s with *CESA*s from other lineages will be evaluated in gene tree analysis. Then to visualize potential parallels in the mechanisms surrounding cellulose synthesis and co-ordination, I search for homologs of *CESA* accessory proteins in *C. tuberculosum*'s transcriptomic dataset.

2.2 Results

2.2.1 Two distinct families of candidate *CESA* genes exist in *C. tuberculosum*

The functionally characterized *CESA*s to date, such as the bacterial *R. sphaeroides BCSA/BCSB* complex and land plant *A. thaliana*'s *CESA*s, share clear sequence similarity in their catalytic domains despite their large evolutionary distances. These regions of sequence homology amongst the *bona fide CESA*s were used as a profile or model (Appendix Table 1) and identified four candidate *CESA* sequences from the *C. tuberculosum* transcriptome (HMM searches, $E \leq 10^{-40}$). The four transcripts (sequence contigs) were designated *CtCESA1, 2, 3,* and *4* (Table 2.1). The candidate *CtCESA1*, *2*, and *3* have only one isoform, and *CtCESA4* has two isoforms. Each isoform represents a predicted splice variant detected during transcriptome assembly. These four candidate *CESA*s were supported by genomic data (reported in Table 2.1) when queried against *C. tuberculosum*'s genomic dataset (BLASTN $E \leq 10^{-5}$).

Table 2.1 *Calliarthron tuberculosum's* candidate *CESA* sequences identified from the HMMER and KEGG search using the profile from Appendix Table 1.

Contig name	Gene	Sequence length (aa)	HMMER E-value	Identified in HMMER search	Identified in KEGG search	Genome Support
c137075	*CtCESA1*	877	6.60E-159	Yes	Yes	Yes
c101319	*CtCESA2*	797	3.40E-53	Yes	Yes	Yes
c140545	*CtCESA4*	649	7.70E-48	Yes	Yes	Yes
c142521	*CtCESA3* isoform 1	770	1.80E-40	Yes	Yes	Yes
c142521	*CtCESA3* isoform 2	770	2.30E-40	Yes	Yes	Yes
c129642	Not included	68	Nothing	No	Yes	No

CtCESA 1-4 were also identified in KEGG based homology mapping (Table 2.1). The identification of the same sequences, despite using different *CESA* HMM models, demonstrated that the candidate *CtCESA*s presented are robust to variations in the model and likely contain core domains found in most *CESA* sequences. KEGG identified one additional sequence that was not considered due to the lack of genomic support and short sequence length.

To broadly consider what proteins these *CtCESA*s are related to, BLASTP showed that the *in silico* translated CtCESA1 shared close homology to other previously reported red algal putative CESAs, and that CtCESA2, 3, and 4 showed close homology to other red algal CESAs as well as to the bacterial cellulose synthase A subunit (Table 2.2)

Table 2.2 Top 5 blast hits of the candidate *CtCESA* sequences against the NCBI non-redundant (NR) database. The sequence was translated *in silico* and then used as a query in BLASTP.

Sequence	BLAST Description (At tax ID 3701)	Query cover	E value	Ident	Accession
CtCESA1	Cellulose synthase (UDP-forming), family GT2 [*Chondrus crispus*]	62 %	0	78 %	XP_005715589.1
	Cellulose synthase catalytic subunit A [UDP-forming] [*Gracilariopsis chorda*]	61 %	0	78 %	PXF48501.1
	cellulose synthase A [*Griffithsia monilis*]	61 %	0	78 %	ADK77974.1
	Cellulose synthase (UDP-forming), family GT2 [*Chondrus crispus*]	60 %	0	63 %	XP_005711952.1
	Cellulose synthase catalytic subunit A [UDP-forming] [*Gracilariopsis chorda*]	60 %	0	60 %	PXF48502.1
CtCESA2	Cellulose synthase catalytic subunit [UDP-forming] [*Gracilariopsis chorda*]	94 %	0	51 %	PXF43171.1
	Cellulose synthase catalytic subunit [UDP-forming] [*Gracilariopsis chorda*]	94 %	0	48 %	PXF44963.1
	CESA-like Cellulose synthase (UDP-forming), family GT2 [*Chondrus crispus*]	74 %	0	63 %	XP_005709948.1
	Cellulose synthase (UDP-forming) fragment, family GT2 [*Chondrus crispus*]	50 %	4.00E-101	56 %	XP_005710660.1
	Cellulose synthase catalytic subunit [*Bacillus cereus* AH1271]	47 %	8.00E-56	33 %	EEL82716.1

Sequence	BLAST Description (At tax ID 3701)	Query cover	E value	Ident	Accession
CtCESA4	Cellulose synthase catalytic subunit [UDP-forming] [*Gracilariopsis chorda*]	94 %	0	53 %	PXF43171.1
	Cellulose synthase catalytic subunit [UDP-forming] [*Gracilariopsis chorda*]	96 %	0	48 %	PXF44963.1
	CESA-like Cellulose synthase (UDP-forming), family GT2 [*Chondrus crispus*]	74 %	0	69 %	XP_005709948.1
	Cellulose synthase (UDP-forming) fragment, family GT2 [*Chondrus crispus*]	56 %	2.00E-94	53 %	XP_005710660.1
	Cellulose synthase catalytic subunit [*Bacillus cereus* AH1271]	52 %	1.00E-55	33 %	EEL82716.1
CtCESA3	Cellulose synthase catalytic subunit [UDP-forming] [*Gracilariopsis chorda*]	76 %	2.00E-176	41 %	PXF44963.1
	Cellulose synthase catalytic subunit [UDP-forming] [*Gracilariopsis chorda*]	76 %	1.00E-170	40 %	PXF43171.1
	CESA-like Cellulose synthase (UDP-forming), family GT2 [*Chondrus crispus*]	70 %	2.00E-154	42 %	XP_005709948.1
	Cellulose synthase (UDP-forming) fragment, family GT2 [*Chondrus crispus*]	29 %	2.00E-62	46 %	XP_005710660.1
	hypothetical protein AMAG_04432 [*Allomyces macrogynus* ATCC 38327]	36 %	1.00E-41	34 %	KNE58894.1

A maximum likelihood *CESA* phylogenetic tree was produced to further examine the relationship between the identified *CtCESAs 1-4* and other known *CESAs* (Figure 2.2). The *CESA*s generally grouped within their kingdoms of origin (Figure 2.2: labelled on outer circle, *C. tuberculosum's CESA* indicated with pink dots). *CtCESA2*, *3*, and *4* grouped together in a clade. This clade was sister to the bacterial *CESA*s and the cyanobacterial *CESA*s with moderate support (ultrafast bootstrap [BS] = 82 %). *CtCESA1* sequence grouped separately from the other *CtCESA*s, in the middle of the red algal *CESA* clade with high support (BS > 97 %). Interestingly, the red algal clade was sister to the oomycete *CESA* clade with high support (BS > 97 %). This relationship is maintained when increasing taxanomic sampling (Appendix Figure 1). However, there is one putative dinoflagellate CESA that separates the red algal and oomycete CESAs with high support (BS > 97%). The origin of this relationship is unclear but may be a unique acquisition from a red alga as the dinoflagellate CESA sequences have a complicated evolution and are spread throughout the CESA gene tree (Appendix Figure 1). The land plant *CESA*s formed a separate clade with short branch lengths and high support (BS > 97 %). CESAs from the land plants and other major organismal lineages were separated by the *CESA-like* (*CSL*) clades *D, B, G*, and *E*. This suggests that the four *CtCESA*s identified show two distinct evolutionary histories, evolving either together with the red algal lineage or more closely with the bacterial *CESA*s, and that their evolution is distinct from the *CESA*s of land plants.

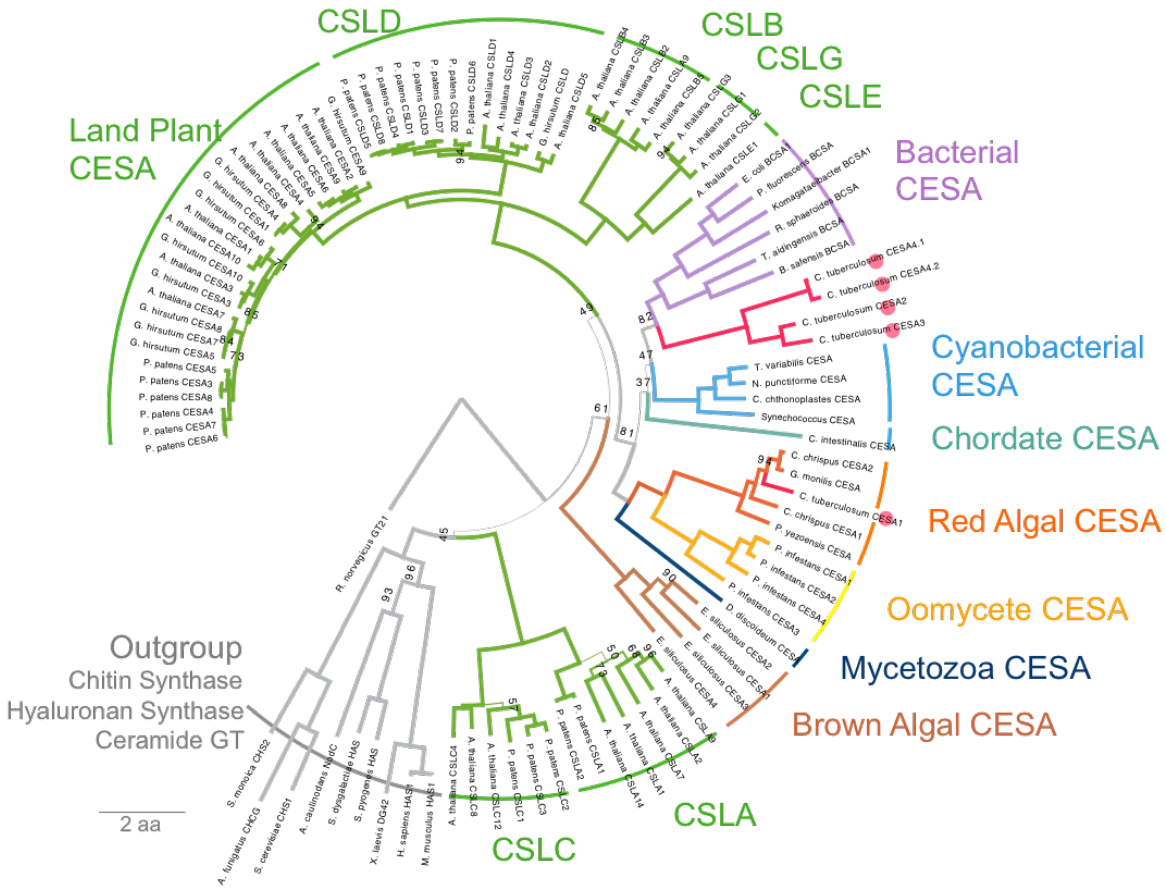

Figure 2.2 Rooted maximum likelihood gene tree of *CESAs* showing the position of the *Calliarthron tuberculosum* sequences (red dots). Increasing support is represented by branch thickness, branches support values < 97 % are shown. Clades are indicated in the outer circle. *C. tuberculosum*'s sequences are indicated with the pink dots. The tree is manually rooted at the *R. norvegicus* GT21 sequence that is outside the CESA GT2 family. Other proteins from the GT2 family are added as a secondary root. Scale indicates the predicted number of substitutions per site. Alignment sites with < 90 % coverage were removed. IQtree was used to search for the model then estimate the maximum likelihood

tree (20 search replicates; log likelihood = -116855). The model under a BIC criteria was VT + F + R6. For sequence identifiers see Appendix Table 2.

2.2.2 *C. tuberculosum CESA*s share similar structure and domains to true plant and bacterial *CESA*s in topological predictions

*CESA*s have several protein characteristics with a clear basis from the solved bacterial BCSA (Morgan et al., 2013; Morgan et al., 2016), eudicot *Populus tremula* × *tremuloides* CESA8 (*PttCESA8*) protein structure (Purushotham et al., 2020), and mutational studies in *A. thaliana*. As a positive control in subsequent comparisons, the bacterial *Rhodobactor sphaeroides BCSA* sequence and the *Arabidopsis CESA7* sequence are functional *CESA*s with expected structural motifs.

The solved bacterial BCSA protein structure, the unknown A. thaliana CESA7 structure, and the C. tuberculosum CESA structure were modeled to visualize the cellulose synthase domains using the DALI program. This DALI algorithm provides the best structural match for query amino acid sequences based on the Protein Data Bank (PDB) of amino acid sequences from known protein structures (Holm, 2020). The best match for the bacterial BCSA, AtCESA, and all three CtCESA sequences was the Rhodobacter cellulose synthase A subunit (c4hg6A in Table 2.3).

Table 2.3 Best hits of protein threading CESA amino acid sequences into the PHYRE database. The *Rhodobacter sphaeroides* sequence is included as a positive control.

Sequence Query	Group	Hit	Description	Alignment %
C. tuberculosum (CtCESA1)	Red Alga	c4hg6A	BCSA	63
C. tuberculosum (CtCESA2)	Red Alga	c4hg6A	BCSA	60
C. tuberculosum (CtCESA4)	Red Alga	c4hg6A	BCSA	54
C. tuberculosum (CtCESA3)	Red Alga	c4hg6A	BCSA	63
A. thaliana (AtCESA7)	Land Plant	c4hg6A	BCSA	71
P. yezoensis (PyCESA)	Red Alga	c4hg6A	BCSA	83
R. sphaeroides (BCSA)	Bacteria	c4hg6A	BCSA	95

Upon visualization of these models (Figure 2.3A), all proteins shared a similar overall shape with transmembrane domains (blue) and glucosyltransferase (GT) domains (yellow). Notably, the number of known transmembrane helixes was more conserved in CtCESA1 than CtCESA2, 3, and 4. CtCESA4, in particular, was missing several transmembrane alpha helixes.

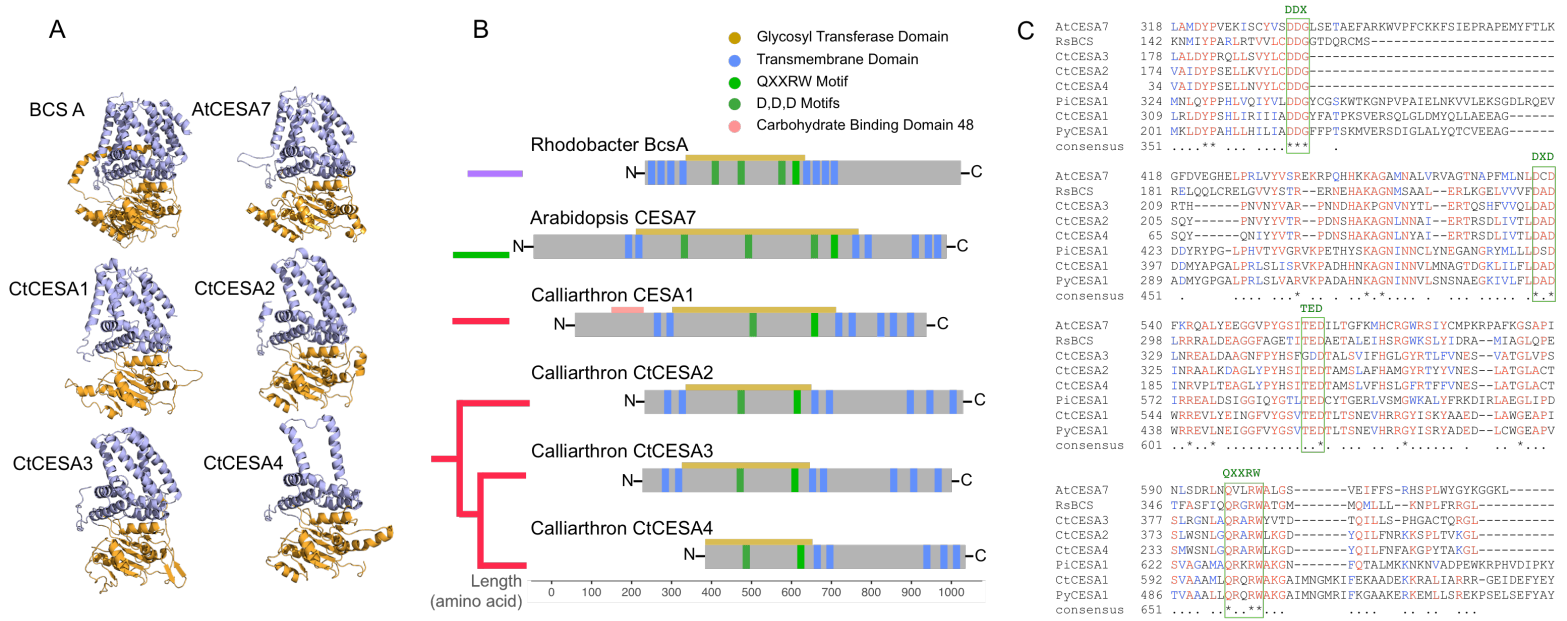

Figure 2.3 *C. tuberculosum CESA*s are used in (A) protein modelling and (B) gene topology maps indicating characteristic motifs. Transmembrane domains (blue) the cytosolic glucosyl-transferase domain (yellow) is indicated. Four characteristic motifs (D,D,D and QXXRW) involved in substrate binding are highlighted (green). All sequences start with intracellular N-terminal domain. Broad phylogenetic relationships are indicated (left). Sequences are to scale. (C) The substrate binding motifs (DDX, DXD, TED, and QXXRW) are highlighted in a multiple sequence alignment. At – *A. thaliana* (land plant eudicot), Rs – *R. sphaeroides* (bacterial), Ct – *C. tuberculosum* (red algal), Pi – *P. infestens* (oomycete), Py – *P. yezoensis* (red algal). Amino acid conservation is indicated in the consensus line (100% "*"; >50% ".").

These key CESA domains were characterized more precisely in the *in silico* translated open reading frames of the *A. thaliana CESA7*, *R. sphaeroides* bacterial cellulose synthase subunit A (*BCSA*), and the *C. tuberculosum CESA* candidates (Figure 2.3B). Topcons analysis of transmembrane domains (blue annotation Figure 2.3B) revealed that *CtCESA1*, *2*, *3* had seven transmembrane domains, and *CtCESA4* had five (Figure 2.3B). *CtCESA4* lacked the transmembrane domains that precede the GT domain in the other sequences. The absence of these transmembrane domains indicated that *CtCESA4* is likely a fragmented sequence and, therefore, a poor candidate for further experimental work. The GT domain (yellow annotation Figure 2.3B), identified in domain searches against the NCBI database, was found in the centrally located cytosolic loop and contained the glucose coordinating motifs D,D,D and the QXXRW motif (green annotation Figure 2.3B). Upon manual inspection in a multiple sequence alignment of the *R. sphaeroides BCSA*, *A. thaliana*, *C. tuberculosum*, and oomycete *CESA*s, all three aspartate residues (DDX, DXD, and TED) and the QXXRW motif were highly conserved (Figure 2.3C). Interestingly, a carbohydrate-binding domain 48 (CBM48) was uniquely identified in the N-terminus of the CtCESA1 sequence (Figure 2.3B pink bar). Bioinformatics analysis of these significant domains revealed that *CtCESA1*, *2*, and *3* were viable CESA candidates and not *CtCESA4*, which was missing key transmembrane domains.

2.2.3 Shared presence of some accessory proteins involved in cellulose synthesis between land plants and *C. tuberculosum*

From studies in land plants, we've learned that cellulose synthesis involves a complex set of accessory proteins to facilitate proper CESA co-ordination and cellulose crystallization. However, there are likely differences between red algal and land plant CESA machinery where land plant CSCs (rosette shape) do not resemble red algal CSCs (linear shaped). Moreover, land plant *CESA*s grouped separately from other lineages such as the red algal *CESA*s (Figure 2.2), suggesting that land plant *CESA*s have likely evolved independently from red algal CESAs. Given these differences in their cellulose complexes, it is unclear if red algae utilize the same cellulose synthesis accessory proteins.

Candidate sequences were identified for *CMU*, *CTL*, and *STELLO* but were not identified for *CC*, *COB*, *CSI,* and *KOR* (Table 1.4) (sequences used for HMM profiles can be found in Appendix Table 3). A comparative species map indicating the presence (green) or absence (red) of these accessory proteins for a variety of cellulose-synthesizing species as identified by Lampugnani et al. (2019) (Figure 1.4) was adapted to include the newly identified candidate accessory proteins from *C. tuberculosum*. Here I present candidates for several *CESA* accessory proteins contrary to Lampugnani et al.'s (2019) previous annotations where none were present in red algae.

Table 2.4 *C. tuberculosum*'s candidate *CESA* accessory protein sequences identified from the HMMER search using the profile from Appendix Table 3 and KEGG. Sequences supported by genomic scaffold are as indicated.

Contig name	Gene	Contig Sequence length (aa)	HMMER E-value	Identified in HMMER search	Genomic Support
-	*KOR*	-	-	No	-
-	*COBRA*	-	-	No	-
-	*CC*	-	-	No	-
-	*CSI*	-	-	No	-
c140937	*CtCMU1*	1898	4.50E-08	Yes	Yes
c141595	*CtCMU2* isoform 2	478	2.10E-07	Yes	Yes
c141595	*CtCMU2* isoform 1	381	4.40E-07	Yes	No
c141595	*CtCMU2* isoform 3	240	4.50E-07	Yes	Yes
c141292	*CtCMU3* isoform 3	525	1.30E-06	Yes	No
c139844	*CtCMU4*	402	1.30E-06	Yes	Yes
c95453	*CtCMU5*	385	2.70E-06	Yes	Yes
c142512	*CtSTELLO*	4176	1.20E-32	Yes	Yes
c142519	*CtCTL1* isoform 1	1421	4.50E-16	Yes	Yes
c142519	*CtCTL1* isoform 4	1397	4.50E-16	Yes	Yes
c142519	*CtCTL1* isoform 5	1436	1.00E-13	Yes	Yes
c142519	*CtCTL1* isoform 2	1552	1.40E-13	Yes	Yes
c140381	*CtCTL2*	953	1.40E-12	Yes	Yes
c142519	*CtCTL3*	1567	3.00E-12	Yes	Yes

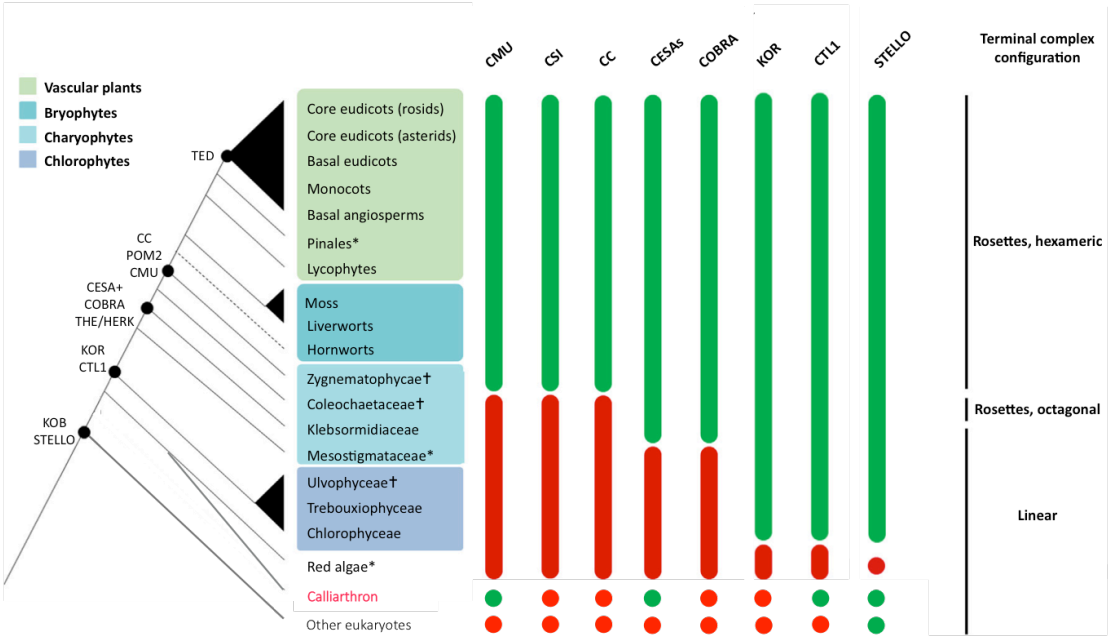

Figure 2.4 "Emergence of Core Components of the Cellulose Synthesis Machinery in the Archaeplastida" (Lampugnani et al., 2019, pp. 408) modified to include sequence search results from the *Calliarthron tuberculosum* transcriptome. A cladogram of Archaeplastida (left) is annotated with the proposed emergence of the cellulose synthesis related proteins at the nodes. Proteins, indicated above, are designated as present (green) or absent (red). Their terminal complex formation is further indicated on the right. Notably, candidate sequences for *CMU*, *CTL1*, *STELLO*, and *CESAs* are identified in this thesis that had not been identified previously in the red algae by Lampugnani et al. (2019).

2.3 Discussion

Here, I identify three promising *CESA* gene candidates from the calcifying red algae *C. tuberculosum*, designated *CtCESA1*, *CtCESA2*, and *CtCESA3*. *CtCESA4* is likely a fragmented sequence, missing the first two transmembrane domains, and not further considered. *CtCESA1*, *CtCESA2*, and *CtCESA3* contain all major domains expected of a functional CESA protein in protein modelling and amino acid sequence analysis. These *CtCESA* sequences are also supported with genomic data indicating they likely originated from *C. tuberculosum* and are not contaminant sequences within the transcriptome dataset.

The relationship of *C. tuberculosum*'s candidate *CESA* sequences with each other and *CESA* sequences from other organisms, as presented in their gene tree (Figure 2.2), is not straightforward. There is evidence for two distinct lineages of *CESA*s in *Calliarthron* where *CtCESA2*, *3*, and *4* were more closely related to bacterial cellulose synthesis subunits, and *CtCESA1* was most closely related to other red algal *CESA*s. None of the *CtCESA* candidates are closely related to the land plant *CESA*s. Below I examine these relationships and their implications in more detail.

2.3.1 *CtCESA2* and *CtCESA3* are closely related to other bacterial *BCSA*s pointing towards possible horizontal gene transfer between bacteria and *C. tuberculosum*

CtCESA2 and *3* are more closely related to each other, but form their next closest relationship with the bacterial cellulose synthase A subunits (*BCSA*) with moderate support (BS = 82). Additionally, BLASTP of *CtCESA2* returned a bacterial *BCSA* in their top 5 results (Table 2.2), while *CtCESA3* returned a bacterial *BCSA* in their top 20 results (not shown). Identifying candidate red algal *CESA*s that are more closely related to bacterial *CESA*s than other red algal *CESA*s is unusual and has not been previously reported. Perhaps *CtCESA2* and *3* were acquired from a bacterium through horizontal gene transfer (HGT). Previous bioinformatics studies of metabolic genes support signals of HGT between bacteria and *C. tuberculosum* (Labeeuw et al., 2015). However, there are clear differences between *CtCESA2*, *3*, and the bacterial cellulose synthase subunits to suggest this acquisition did not happen within their recent evolutionary history. First, the *CtCESA2* and *3* sequences are moderately related to and not embedded within the *BCSA* clade. Secondly, the bacterial cellulose synthesis machinery is a multiprotein complex specialized for cellulose synthesis across the bacterial inner and outer membrane (reviewed in McNamara, Morgan, and Zimmer 2015). *C. tuberculosum* has no additional sequence candidates for the bacterial cellulose synthase complex subunits (BCSB, BCSC, etc.) within the KEGG annotations, pointing towards the absence of a bacterial like multiprotein complex. For these two reasons, if *CtCESA2* and *3* were acquired through HGT, it would have likely occurred early, before the divergence of BCSA subunits within the bacterial lineage. Identifying and including more bacterial *CESA*s within this tree may better help map their relationship by identifying the sequence closest to the ancestrally acquired *CESA*.

2.3.2 *CtCESA1* is related to other putative red algal *CESA*s

CtCESA1 has a highly supported (BS > 97%) and close relationship with other annotated red algal *CESA*s (Figure 2.2 and Appendix Figure 1). Apart from *CtCESA2* and *3* discussed above, all red algal *CESA*s identified to date group tightly into this one clade (Roberts and Roberts, 2009; Matthews et al., 2010; Brawley et al., 2017). This indicates that the red algal *CESA* was likely acquired before the divergence of the two major red algal groups (Florideophyceae and Bangiophyceae). This is exemplified by the short branch lengths between the red algal CESAs (Figure 2.2) and the relationships between the red algal CESA sequences largely mirroring their species evolution (Yoon et al., 2006; Qiu et al., 2016). A unique feature of all red algal CESAs including the *CtCESA1* sequence, but not the *CtCESA2* or *3* sequences, is a cytosolic exposed N-terminal carbohydrate-binding module 48 (CBM48) domain (Figure 2.3) predicted to bind glycogen or starch carbohydrates (Matthews et al., 2010). Generally, land plants store their starch in the chloroplast (Lewis and McCourt, 2004), inaccessible to cytosolic enzymes. However, red algae uniquely store their photosynthetic products, Floridean starch, as free-floating granules in the cytoplasm (reviewed in Viola, Nyvall, and Pedersén 2001). In theory, the Floridean starch granules could directly interact with the N-terminal CBM48 on the cytoplasmic face of the plasma membrane. Similar CBM48 domains function as a sensory domain in some rat kinases (Polekhina et al., 2005), and share some structural similarities with the predicted red algal CBM48 domains based on protein modelling (Matthews et al., 2010). The CBM48 domain in red algal CESAs could be a unique regulatory domain that interacts with these cytoplasmic Floridian starch granules, though this hypothesis has yet to be tested.

There is currently no functional evidence within the red algal group to suggest that this clade has CESA activity. Instead, these red algal *CESA* sequences have been annotated solely using sequence similarity-based approaches such as bioinformatics screens of red algal genomic data in *P. yezoensis* and *C. chrispus* (Collén et al., 2013; Brawley et al., 2017) and sequence characterization by amplification from expression libraries in *P. yezoensis* and *G. monilis* (Roberts and Roberts, 2009; Matthews et al., 2010). Despite the lack of direct functional evidence, we can make inferences in the red algal CESA group from their close relationship with the Oomycete *CESA*s as observed here (Figure 2.2) and in previous phylogenetic studies of algal CESA sequences (Roberts and Roberts, 2009; Michel et al., 2010; Collén et al., 2013). Several species in the oomycete group have demonstrable CESA activity. When CESA expression levels are inhibited in RNAi experiments on appressoria tissue in *Phytophthora infestans*, cellulose production decreases over 50% (Grenville-Briggs et al., 2008). Additionally, detergent-solubilized CESAs from the oomycete *Phytophthora capsici* heterologously expressed in yeast strains engineered to have low endogenous glucan synthase activity showed activity *in vitro* with glucose incorporated into short cellulose products, similar to cellobiose (Pang et al., 2020). Perhaps these pathogenic oomycete *CESA*s have red algal origins. Signals of gene integration from an ancestral endosymbiotic relationship or horizontal gene transfer from red algae to oomycetes have been hypothesized (Wang et al., 2017). Given this hypothesized orthologous or shared evolution, the oomycete group's demonstrated activity strengthens our predictions of protein function in the red algal *CtCESAs*.

Despite these red algal CESA sequence annotations and functional prediction by association with oomycete CESA activity, there is a lack of biochemical evidence to support the identity of the

red algal CESAs. To build on previous work on the red algal CESA group identified, I will perform enzymatic characterization of the *CtCESA1* sequence in chapters three and four of this thesis.

2.3.3 Land plants and red algae: parallels and divergences in the mechanics surrounding cellulose synthesis

Our current paradigm of cellulose production involves a multilayered process that involves the trafficking of CESAs from the endoplasmic reticulum to the plasma membrane and a host of accessory proteins that support the cellulose synthase complex (CSC) machinery. As this understanding comes solely from studies in land plants, it is unclear if the processes surrounding cellulose synthesis are similar in non-land plant systems. Here I present strong evidence to the contrary. First, the monophyletic land plant CESA clade (Figure 2.2) is separate from red algal CESAs in phylogenetic analysis, suggesting CESA evolution and specialization likely occurred independently in these two lineages. Second, CESA accessory proteins are present in land plants but mostly absent in red algae (Figure 2.4). Accessory protein sequence candidates were identified for CMU, STELLO, and CTL proteins but not KOR, COBRA, CC, or CSI in *C. tuberculosum*. With these two findings, I propose that the evolved mechanism of cellulose synthesis likely differs between land plants and red algae.

The presence or absence of these accessory proteins has interesting implications for cellulose biosynthesis in the red algae. STELLO proteins have been suggested to function in proper CSC assembly and trafficking to the plasma membrane by an unknown mechanism (Zhang et al.,

2016). The presence of STELLO in *C. tuberculosum* and other eukaryotic lineages (Figure 2.4) suggests that STELLO mediated CSC complex formation in the Golgi is a deeply conserved process and may function similarly.

Microtubule-CESA interactions are critical to cellulose production in land plants, and while my analysis supports the general concept that these interactions are important in red algal cellulose biosynthesis, it also suggests unknown mechanisms for CtCESA interaction with microtubules. In *A. thaliana*, directional movement of the CSC at the plasma membrane tracks along microtubules (Paredez et al., 2006). There the CESA-microtubule interaction is physically mediated by the cellulose synthase interacting protein (CSI) (Gu et al., 2010; Purushotham et al., 2020) while cellulose synthase-microtubule uncoupling (CMU) and companion of cellulose synthase (CC) proteins may mediate microtubule stability. CMU proteins are thought to oppose the forces generated by CESA complexes as they move through the plasma membrane and push against the microtubules (Liu et al., 2016) and CC is thought to maintain cellulose production under salt stress by promoting microtubule polymerization (Endler et al., 2015). In this thesis I found sequence candidates for CMU, indicating that there could be microtubule-CESA interactions in red algae (Figure 1.4). However, the absence of CC or CSI protein candidates raises the question of how the red algal CESAs might interact with the microtubules and what role microtubules play in cellulose production under stress responses. The CSI has been shown to interact with the N-terminal domain of PttCESA8 in *in vitro* pull-down assays (Purushotham et al., 2020). However, this N-terminal domain has low similarity between land plant and red algal CESAs, perhaps suggesting that the CSI-CESA-microtubule interaction may be specific to land plant CESAs. This is supported by CSI's corresponding emergence in charophytes and

early-diverging land plants (Figure 2.4). Instead, there may be an unidentified novel protein that aids in this interaction in the red algae.

It is unclear how cellulose crystallization is facilitated in red algae. In land plants, KORRIGAN (KOR), COBRA, and chitinase-like (CTL)/pom-pom1 have glucan binding motifs and are hypothesized to bind cellulose as stabilizing factors. KOR, a β-1,4-endo-glucanase associated with the CESA complex (Mansoori et al., 2014; Vain et al., 2014) in *A. thaliana*, is hypothesized to release stress created by cellulose synthesis by selectively hydrolyzing glucose bonds. Though COBRA and CTL's mechanistic function remains unknown, defects in these proteins directly affect cellulose. However, among these proteins, only sequence candidates for CTL were identified in *C. tuberculosum* suggesting major differences in cellulose stabilizing factors.

In summary, processes surrounding cellulose production have likely evolved divergently between the major lineages of Chlorophytes (green algae) and Rhodophytes (red algae), with many accessory proteins arising in the green plant lineage (encompassing the green algae to vascular plants) and not present in the red alga (Lampugnani et al., 2019) *C. tuberculosum* (Figure 2.4). Evidently, our current paradigm of cellulose production rooted in studies from land plants requires re-evaluation in more distant lineages such as the red algae.

2.4 Methods

2.4.1 *C. tuberculosum CESA* and accessory protein sequence identification

To create a database for sequence searches, the *C. tuberculosum* transcriptome dataset (Xue et al. in prep; European Nucleotide Archive project number PRJEB39919) was translated into all 6 reading frames using EMBOSS Transeq (Rice et al., 2000). Amino acid sequences were used for the subsequent analyses as they incorporate more information regarding functional conservation.

CESA, and *CESA* accessory protein sequences were identified from *C. tuberculosum* using sequence similarity with previously annotated sequences in a HMMER search. "Profiles", HMM queries, were produced for CESA (Appendix Table 1) and CESA accessory proteins (Appendix Table 3) by obtaining *in silico* translated amino acid sequences from a variety of species and aligning them using muscle (Edgar, 2004). No manual adjustment of the alignment was made. This profile was used to search against the translated *C. tuberculosum* transcriptome for candidate CESAs. The transcriptome was also mapped onto the KEGG database (Kanehisa and Goto, 2000) and candidate *C. tuberculosum* sequences were recovered. Sequences shorter than 700bp were not considered for further analysis. To predict homology of these *CtCESA* sequences, they were searched against the *Arabidopsis* genomes under default settings (Mahram and Herbordt, 2010) (BLASTP, $E \leq 10^{-5}$). Transcriptome sequences supported by previously produced genomic scaffolds were identified (BLASTN, $E \leq 10^{-5}$) to increase confidence as non-contaminant sequences (Xue et al. in prep; supporting information can be found at https://github.com/martonelab/geneAnnotCalliarthronTranscriptome).

2.4.2 CESA motif analysis

NCBI's ORF finder was used to find the longest open reading frames for each of the *CtCESA* sequences (Rombel, Sykes, Rayner, & Johnston, 2002). The nucleotide sequences were imputed and searched under a standard genetic code with "ATG only" as the start codon.

The topology conservation program (TOPCONS) was used to find transmembrane domains (Tsirigos, Peters, Shu, Kall, & Elofsson, 2015). TOPCONS calculates several topology predictions and quantifies the consensus prediction, the TOPCONS prediction, based on the agreement of the individual underlying programs. The TOPCONS prediction was used for topology annotations.

NCBI's conserved domain search tool based on the conserved domain database (Marchler-Bauer et al., 2014) and the PROSITE scanning tool (de Castro et al., 2006) was used to search for general domains in the open reading frame. Default search settings were used. The Find Individual Motif Occurrences (FIMO version 5.05) was used to search for the QXXRW motifs in the *in silico* translated CtCESA amino acid sequences (Grant, Bailey, & Noble, 2011).

Specific motifs were also manually identified in a multiple sequence alignment (MSA). Sequences were compiled and aligned with MUSCLE under the default settings (Edgar, 2004). Sites with < 90 % coverage were removed using trimAl (Capella-Gutiérrez et al., 2009). The MSA was visualized using boxshade (https://embnet.vital-it.ch/software/BOX_form.html).

2.4.3 Protein modelling of CESA sequences

Protein modelling was conducted by querying the amino acid sequences against the DALI webserver (Reviewed in Holm 2020). DALI identified the closest sequence match from the PDB. These protein matches were also confirmed using the Protein Homology/analogY Recognition Engine (PHYRE version 2.0) (Kelley, Mezulis, Yates, Wass, & Sternberg, 2015). The proteins were modelled against the *BCSA* catalytic subunit. Pymol was used to visualize the subsequent protein structures (The PyMOL Molecular Graphics System, Version 2.3, Schrödinger, LLC.).

2.4.4 Gene tree analysis

Maximum likelihood (ML) phylogenetic methods were applied to the CESA sequences from *C. tuberculosum* and other species (sequence identifier in Appendix Table 2). The candidate sequences from the HMMER search and the curated *CESA* sequences were compiled and realigned using MUSCLE under the default settings (Edgar, 2004). Alignments were visually examined using SeaView (Gouy et al., 2009). Sites with < 90 % coverage were removed using trimAl (Capella-Gutiérrez et al., 2009). IQtree was used to search for the evolutionary model of the alignment under a Baysian information criterion (BIC) (Luo et al., 2010; Nguyen et al., 2015). A maximum likelihood gene tree search with 20 search replicates and branch support calculated under 1000 replicates of ultrafast-bootstrap approximation was conducted using IQtree (Hoang et al., 2018). Interpretation of clade reliability is high when support is ≥ 95 %

under ultrafast bootstrap approximation. FigTree was used to edit branch width and colors (Rambaut and Drummond, 2010).

Outgroup sequences to the *CESA* family are as indicated (Appendix Table 2). Outgroup allow a clearer interpretation of the relationships that fall within the clade of interest. It should be a related sequence or group of sequences that fall outside the clade of interest. I use two outgroups listed here. First, hyaluronan synthase (*HAS*), N-acetylglucosaminyltransferase (*NOD*), and chitin synthase (*CHS*) are part of the glycosyltransferase 2 (GT2) family that CESAs belong to and thought to be closely related to CESAs. They have conserved motifs such as the QXXRW; D,D,D motifs. Second, ceramide glycosyltransferase also has the QXXRW; D,D,D motifs and is able to bind UDP-glucose, but this protein is part of the GT21 family and not the GT2 family. Assuming these shared motifs are due to homology (sequence similarity through shared ancestry) and not homoplasy (sequence similarity acquired through other means such as convergence or reversal) this makes the ceramide glycosyltransferases a good candidate to further establish the identity of the outgroup.

Chapter 3: A heterologously expressed CESA from the red alga *C. tuberculosum* shows glucan synthase activity *in vitro*

3.1 Introduction

Previously, I identified three supported cellulose synthase (*CESA*) candidates from the *C. tuberculosum* transcriptome (*CtCESA 1-3*). *CtCESA1* was most closely related to putative red algal *CESA*s (Roberts and Roberts, 2009; Matthews et al., 2010; Brawley et al., 2017), while *CtCESA2* and *CtCESA3* were more closely related to bacterial BCSA subunits (chapter two). However, previous studies have provided no experimental data to support the annotation of any red algal *CESA*s. This chapter tested whether *C. tuberculosum*'s putative cellulose synthase 1 (*CtCESA1*) encodes a functional protein with glucan synthase activity.

The purification and study of land plant cellulose synthases (CESAs) *in vitro* has been a long sought after goal. Previous researchers' attempts to purify CESAs from their native membranes have been unsuccessful. This was largely due to the inability to extract CESA complexes from their native membranes using common detergents (Taylor et al., 2004) and the high levels of endogenous glucosyltransferase activity from callose synthases, which have obscured even crude studies of CESAs in native isolated microsomes (Lai-Kee-Him et al., 2002; Cifuentes et al., 2010). The first breakthrough was the solved crystal structure and enzymatic characteristics of cellulose synthases from the bacteria *Rhodobacter sphaeroides* (BCSA/BCSB) (Morgan et al., 2013; Omadjela et al., 2013; Morgan et al., 2016), followed by the successful characterization of

the eukaryotic CESAs from *Populus tremula x tremuloides* and *Physcomitrella patens* (Purushotham et al., 2016; Cho et al., 2017; Purushotham et al., 2020).

Within the last decade, there have also been major advances in heterologous eukaryotic gene expression in the yeast *Pichia pastoris* (Pichia) (Cereghino and Cregg, 2000) and insect cell host systems (Kost et al., 2005), which have facilitated breakthroughs in the purification and characterization of eukaryotic CESAs *in vitro*. Insect cells and yeast cells are common host systems suitable for expressing eukaryotic proteins in large scales. They share a eukaryotic codon usage and post-translational modifications with other eukaryotes that allow for successful expression and translation of full-length genes (reviewed by Kost, Condreay, and Jarvis 2005; Bernaudat et al., 2011).

Once protein expression has been established, further assays must be performed to optimize their purification. CESAs are transmembrane proteins, embedded in the plasma membrane. To isolate and purify CtCESA1, detergents are required to extract and separate these membrane bound proteins from their host membrane. Addition of detergent in optimal conditions disrupts the membrane bilayer and replaces the native membrane lipids surrounding the transmembrane domain of the protein (Figure 3.1A). Because the isolated CtCESA1 protein will ultimately be used in enzymatic assays, the detergent chosen must both isolate (solubilize) the membrane bound CtCESA1 protein into solution as well as maintain its native and functional shape. Detergents can differ in their ability to perform these two functions and a variety of common detergents (Figure 3.1B) should be tested to find the optimal purification solution (reviewed by Privé 2007 and Seddon, Curnow, and Booth 2004).

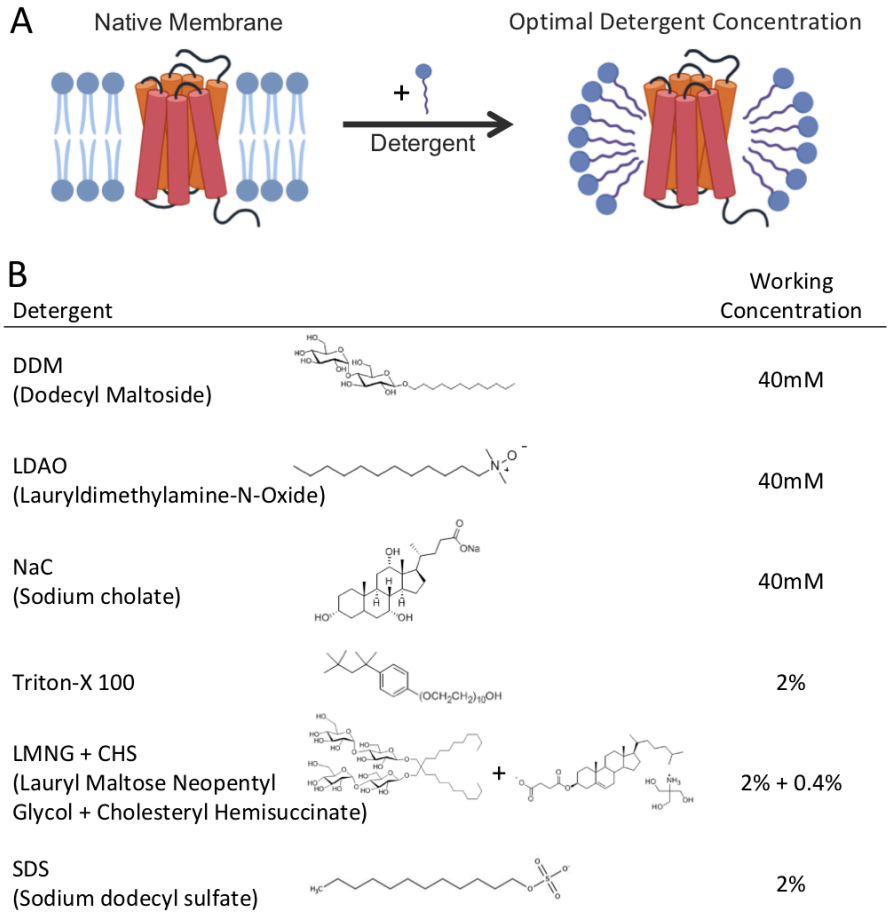

Figure 3.1 Isolation of membrane embedded proteins using common detergents.

(A) Cartoon depiction of the membrane protein isolation process. (Image produced from BioRender.com) (B) Detergents used to screen for CtCESA1 membrane protein isolation. The commonly used acronym, structure, and working concentrations are shown for each detergent.

The final step in creating an optimal environment for CtCESA1 enzymatic function is to modify the detergent/lipid into either a glyco-diosgenin (GDN) detergent environment or to reconstitute the CtCESA1 proteins into *E. coli* proteoliposomes. GDN has been shown to have superior protein stabilizing effects over a wide range of integral membrane proteins (Kotov et al., 2019). Another common method for protein stabilization is reconstituting the purified protein into proteoliposomes (for a detailed description of the method see Rigaud and Lévy 2003). In this process, the environment surrounding a portion of the CtCESA1, purified from insect cells or Pichia cells, was then switched from a detergent to total *E. coli* lipids in order to better mimic native conditions and facilitate protein folding into its functional shape.

Cellulose synthases (CESAs) are glucosyltransferase enzymes that take up activated UDP-glucose monomers and add them onto a growing β-1,4-polymer chain. To function, CESAs also require a cationic co-factor such as magnesium (Mg^{+2}) or manganese (Mn^{+2}) (Pear et al., 1996; Omadjela et al., 2013; Purushotham et al., 2016). Calcium (Ca^{+2}) is used as a control, as callose synthase activity, but not cellulose synthase activity is stimulated. The Ca^{+2} cation is larger in size and less electronegative than Mg^{+2} or Mn^{+2}, and therefore cannot act as a cofactor for CESA activity, and in this condition, little to no glucan transferase activity occurs.

Previously, glucan synthase activity of individual land plant CESAs, expressed in Pichia or insect cells, could produce individual β-1,4-glucan chains (Purushotham et al., 2016; Cho et al., 2017; Purushotham et al., 2020). These studies were instrumental in establishing a methodology for heterologous CESA expression, purification, and *in vitro* experimentation. Given the success demonstrated in Purushotham *et al*. (2016, 2020), I established a collaboration with the Zimmer

lab at the University of Virginia to apply these methodologies for the first characterization of a *CESA* from the red algal lineage (*CtCESA1*).

3.2 Results

3.2.1 CtCESA1 expression and purification

3.2.1.1 An engineered CtCESA1 expressed in SF9 insect cells

CtCESA1 tagged with a 12x HIS tag was engineered into a baculovirus used to infect *Spodoptera frugiperda* (SF9) insect cell cultures (Figure 3.2A). Before viral infection, cells were more uniform in diameter and, on average, smaller in size than cells following viral infection. Infection also resulted in a higher number of dead cells, assessed with tryptan blue staining, demonstrating successful infection (Figure 3.2A). CtCESA1 protein expression levels from four independently transformed small-scale cultures (Figure 3.2B; called P0 stage for the purpose of propagating the virus) had successfully expressed the full-length CtCESA1 polypeptide at the predicted size of 101.7 kDa (Figure 3.2C black arrow) as shown in a western blot against the 12x HIS tag. In the absence of infection, no such polypeptide was detected (Figure 3.2C), indicating that the presence of the strong band at the predicted CtCESA1 size in the infected cultures is specifically due to the introduction of the *CtCESA1* sequence. Thus, the insect cell culture system was successfully used to produce the *CtCESA1* for further characterization.

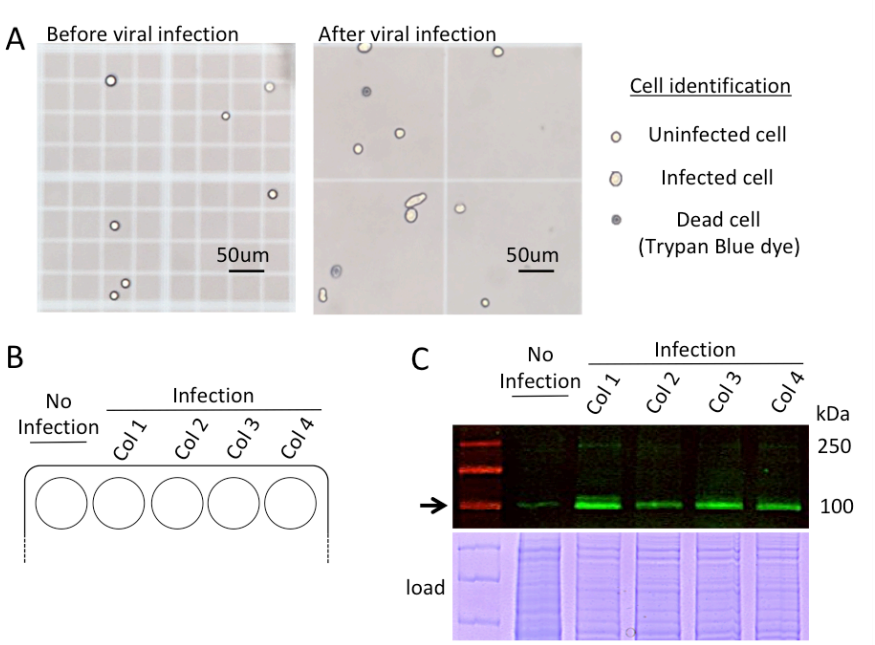

Figure 3.2 Infected SF9 insect cells with CtCESA1-containing virus for heterologous protein expression. (A) SF9 cells before and after viral infection. (B) Cartoon diagram of initial, P0, infection of insect cells with the bacmid viral DNA containing CtCESA1. Four independent bacmid insertion events were used for infection. No infection represents SF9 cells with no virus added. (C) CtCESA1 expression levels in P0 SF9 insect cells western blotted with primary anti-HIS. CtCESA1 is indicated with an arrow. Protein load is indicated.

3.2.1.2 Optimization of CtCESA1 purification from both insect cell membranes and yeast cell membranes

A detergent screen was conducted to find the optimal detergent for isolating heterologously expressed CtCESA1 from the membranes of insect or yeast host cells. Lysed insect cells or yeast

cells expressing CtCESA1 protein were separately incubated with several common membrane solubilization detergents: DDM, LDAO, NaC, Triton-X 100, LMNG+CHS (listed in Figure 3.1B). The membrane-detergent solutions were centrifuged at high speed to separate detergent-isolated CtCESA1, found in the supernatant fraction, from the insoluble CtCESA1 in the host cell membrane, found in the pelleted fraction.

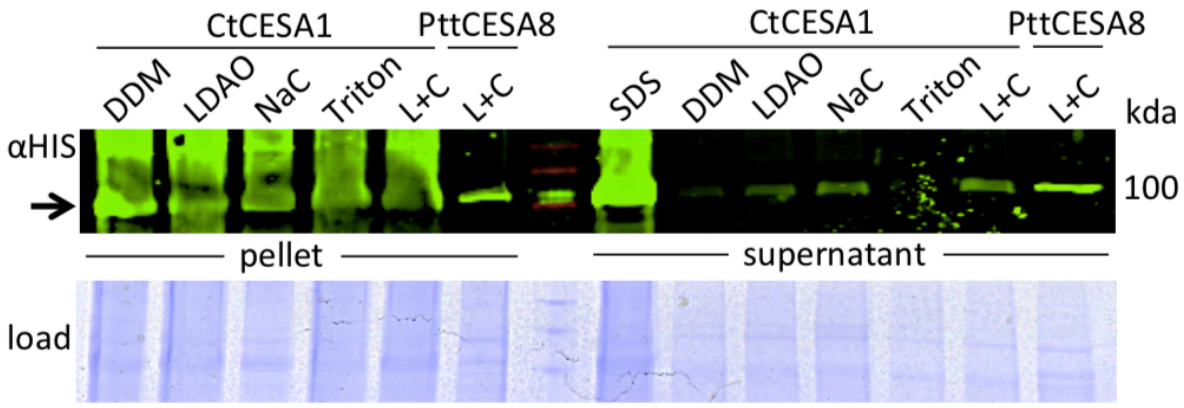

Figure 3.3 Detergent screen of HIS-tagged CtCESA1 embedded in insect cell membranes shows 1 % LMNG + 0.2 % CHS was an acceptable detergent for CtCESA1 isolation. Detergent used is indicated directly above each lane. Insect cells expressing PttCESA8 (Purushotham et al., 2020) **were exposed to L+C as a positive control. Pellet fraction represents unsolubilized proteins. Supernatant represents solubilized proteins. The predicted CtCESA1 size is indicated with an arrow (left). Protein load is shown. Western blot with primary anti-HIS.**

SDS was a harsh ionic detergent used as a positive control to demonstrate near complete protein solubilization. It was not appropriate for experimental CtCESA1 isolation followed by enzyme activity assays, as it would cause complete protein denaturation.

In insect cell samples, as expected, SDS was an effective positive control for solubilization of the HIS-tagged *CtCESA1*, with the strongest band in the supernatant fraction (Figure 3.3). To test less harsh detergents, insect cell lysate with the membrane bound CtCESA1 was subjected to a detergent screen with DDM, LDAO, NaC, Triton-X 100, and LMNG+CHS (Figure 3.3). The majority of CtCESA1 proteins, indicated at the predicted 101 kDa protein size (black arrow), were insoluble and lost in the pellet fraction (bright bands) relative to the soluble protein in the supernatant fraction (faint bands). LMNG+CHS (listed as L+C in Figure 3.3) and NaC perform the best in the detergent screen on CtCESA1 in insect cell membranes as indicated by the brightest bands in the supernatant fraction (Figure 3.3). Protein concentrations in each condition were not due to the amount of total protein loaded, shown by the equal loading in the Coomassie stained gel (Figure 3.3, load). The incomplete solubilization of CtCESA1 was consistent with previously optimized published results for *Populus tremula x tremuloides* CESA8 (PttCESA8) (Purushotham et al., 2020). PttCESA8, simultaneously processed with CtCESA1 samples, showed approximately equal brightness between pellet and supernatant bands, which suggested that ~ 50 % of the PttCESA8 was isolated into the supernatant (Figure 3.3 PttCESA8 at 110.5kDa). Although much of the CtCESA1 was lost in the insolubilized form in the pellet, the LMNG+CHS was the best detergent for isolating sufficient CtCESA1 in the supernatant for further testing.

A similar detergent screen was performed on the Strep-12xHIS-tagged CtCESA1 containing Pichia cells. Again, much of the CtCESA1 proteins (arrow) were insoluble and lost in the pellet fraction (bright bands) relative to the soluble protein in the supernatant fraction (faint bands) (Figure 3.4). DDM, Triton-X 100, and LMNG+CHS performed equally well in the detergent screen on CtCESA1 in Pichia cell membranes, indicated by the brightest bands in the supernatant fraction of the anti-Strep panel (Figure 3.4). These same bands were not seen in the anti-HIS western panel, potentially due to protein degradation. This particular gel was run after the anti-Strep gel. Complete solubilization of CtCESA1 was possible in the detergent screen, seen in the SDS supernatant positive control. A blank was loaded between the supernatant of the SDS and other detergent samples to prevent cross-contamination. The equal load in the Coomassie stained gel (Figure 3.4, load) indicated that the relative brightness of the bands in the western blot is qualitatively comparable.

The LMNG+CHS was the detergent that performed most effectively at solubilizing CtCESA1 into solution from both the insect cell membranes and the Pichia cell membranes, and for PttCESA8 purification from insect cell membranes. To increase comparability between proteins isolated from these sources, all further protein solubilization used LMNG+CHS.

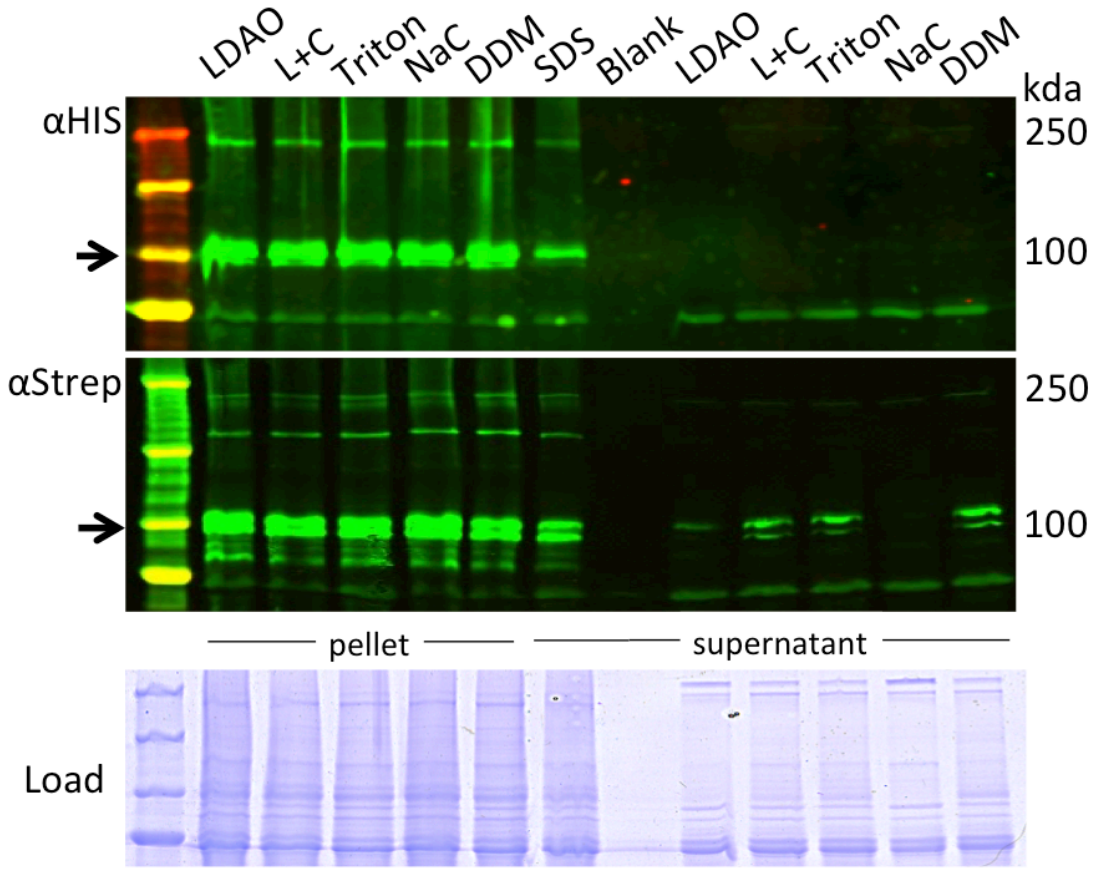

Figure 3.4 Detergent screen of Strep-HIS-tagged CtCESA1 embedded in Pichia cell membranes shows 1% LMNG + 0.2% CHS was an acceptable detergent for CtCESA1 isolation. Detergent used is indicated directly above each lane. Pellet fraction represents unsolubilized proteins. Supernatant represents solubilized proteins. The predicted size for CtCESA1 is indicated with an arrow on the left. The protein amount loaded can be seen below. Western blots with primary anti-HIS and primary anti-Strep are indicated.

3.2.1.3 Two-step purification of CtCESA1: gravity flow anti-HIS column purification and size exclusion chromatography

The solubilized CtCESA1 protein was purified in two steps before enzymatic analysis. For CtCESA1 expressed from both insect and Pichia, gravity flow column purification against CtCESA1's HIS-tag roughly purified the CtCESA1 out of solution. Size exclusion chromatography (SEC) further refined the CtCESA1 purification by isolating a small fraction of proteins that corresponded to the CtCESA1 size range (~100kDa – 150kDa).

Before column purification, the CtCESA1 was found in both the insoluble (pellet) and soluble (supernatant) as expected (Figure 3.5A). After HIS-tag purification, CtCESA1 was purified in the elution fraction at the correct predicted size (arrow, Figure 3.5A). As expected, the purification process was not completely efficient as some CtCESA1 was lost in the flow through and wash (bright band in lanes FT and W, Figure 3.5A). The elution fraction that contained the roughly purified CtCESA1, was further purified in SEC (Figure 3.5B).

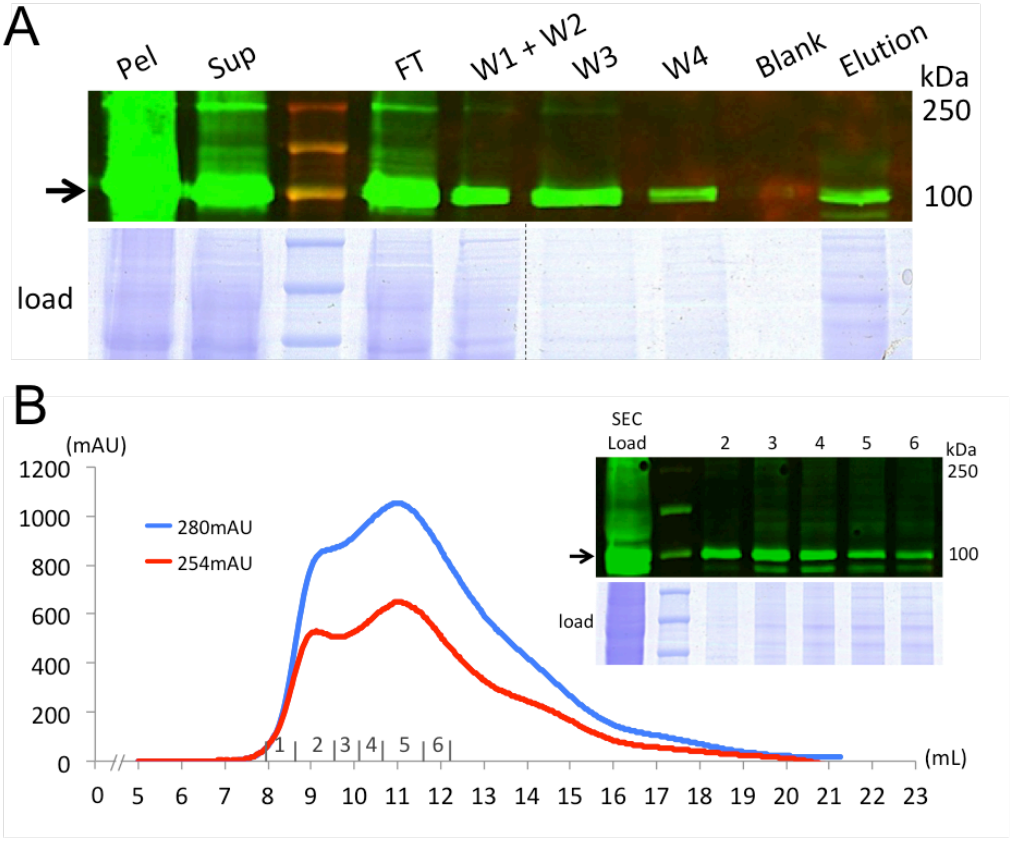

Figure 3.5 CtCESA1 expressed in SF9 insect cell lines was isolated from column purification and SEC. (A) CtCESA1 isolated from gravity flow column purification against the HIS tag as shown in elution (black arrow). Each lane shows a purification step. (Pel = pellet; Sup = supernatant; FT = flow through: W = wash). The protein load is indicated. The dotted line in the load panel represents a cropped out section containing a blank lane. (B) CtCESA1 isolated from SEC. Fractions of protein taken from SEC shown in grey numbers above the x-axis and above the western blot panel. Beginning volumes in the x-axis has been truncated for brevity. Protein loaded onto the column at time 0, from the concentrated elution fraction in (A) is indicated. Western blots used primary anti-HIS antibody.

When the sample passed through the SEC and the measured protein levels (280mAu blue Figure 3.5B; 254mAu is nucleic acid levels as a control) began to increase, multiple fractions were collected (numbered grey intervals directly above the x-axis, Figure 3.5B). Initially, higher molecular weight proteins in the "void" range eluted, followed by smaller protein sizes corresponding to the CtCESA1 size. Fractions one to six from both the "void" fractions and molecular size fractions corresponding to CtCESA1 showed that all contained CtCESA1 (arrow Figure 3.5B). Void fractions two and three were discarded (between 8mL and 9.5mL in the SEC) as they likely contained other aggregated proteins and proteases that could exacerbate protein degradation. Fractions four, five, and six were collected for further enzymatic assays.

The same process described above was used to purify CtCESA1 from heterologous expression in Pichia. CtCESA1 had both a 12xHIS-tag and Step-tag added for the purpose of protein purification after heterologous expression in Pichia. After sequential column purification against the HIS-tag and then the Strep-tag, CtCESA1 (indicated with an arrow) was present in the elution of the HIS-tag purification (Figure 3.6A) but not the Strep-tag (Figure 3.6B). The majority of the protein was lost in the flow through (FT, Figure 3.6B), which indicated CtCESA1's Strep tag was unable to bind the anti-Strep antibody on the purification column, and therefore an ineffective purification step. For this reason, further Pichia expressed CtCESA1 was purified using the HIS-tag only.

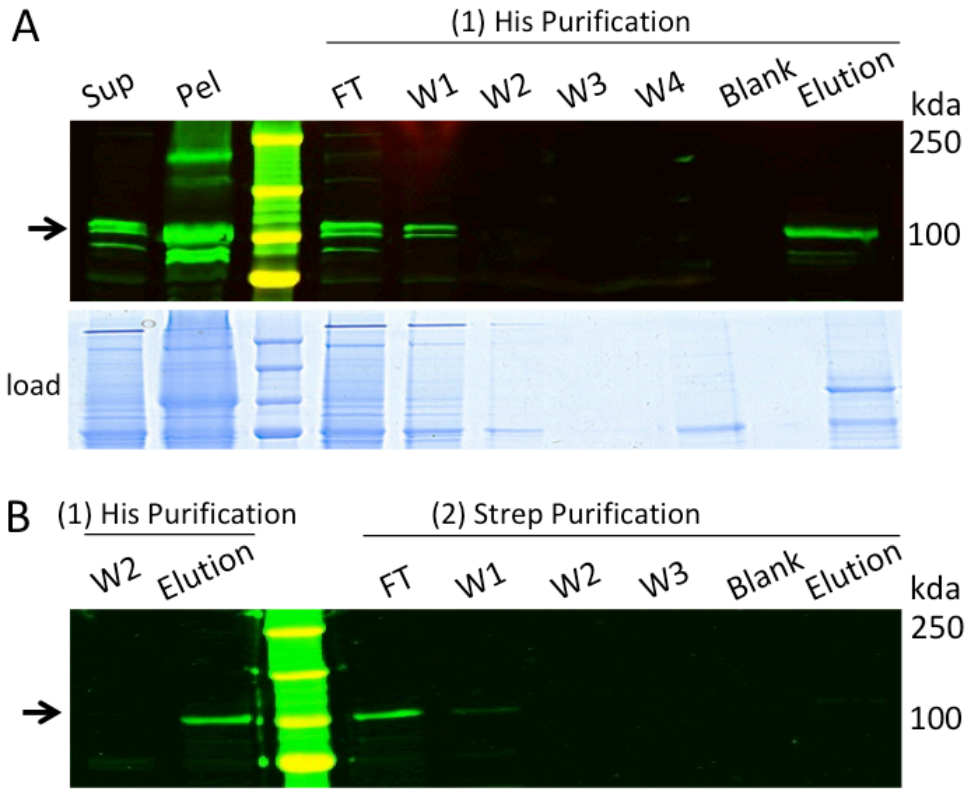

Figure 3.6 Two-step Strep-HIS-tag CtCESA isolation was unsuccessful. (A) First step in CtCESA1 (black arrow) isolation from gravity flow column purification against the HIS tag. Supernatant, soluble CtCESA1, was purified. The elution was applied to the anti-Strep column for the second purification step. (B) Second step in CtCESA1 (black arrow) isolation from gravity flow column purification against the Strep tag. Samples in the anti-Strep column were lost in the flow through. The W2 and the elution from the HIS purification (indicated) are shown with the anti-Strep column purification steps (indicated). (Pel = pellet; Sup = supernatant; FT = flow through: W = wash). Primary anti-HIS used in panel A. Primary anti-Strep used in panel B.

Full-length detergent solubilized CtCESA1 from Pichia lysate was separated into insoluble (pellet) and soluble fractions (solution) (arrow Figure 3.7A). Full length HIS-tagged CtCESA1 was purified from gravity flow column purification (Elution; Figure 3.7A). As expected, some HIS-tagged CtCESA1 was lost in flow through (FT) and wash steps (W1 + W2) (Figure 3.7A). Eluted CtCESA1 was further purified in SEC (SEC load Figure 3.7B). Protein size fractions corresponding to the "void" volume (fractions one and two) and CtCESA1 size fractions (fractions three, four, and five) were collected (indicated above x-axis in grey Figure 3.7B). CtCESA1 was most prominent in fractions two to four as seen in western blot (arrow in Figure 3.7B). Proteins from the "void" fractions were discarded. Fraction three and four containing CtCESA1 were collected for further enzymatic assays.

CtCESA1 was then exchanged into a GDN detergent environment or an *E. coli* lipid environment for further functional enzyme activity analysis.

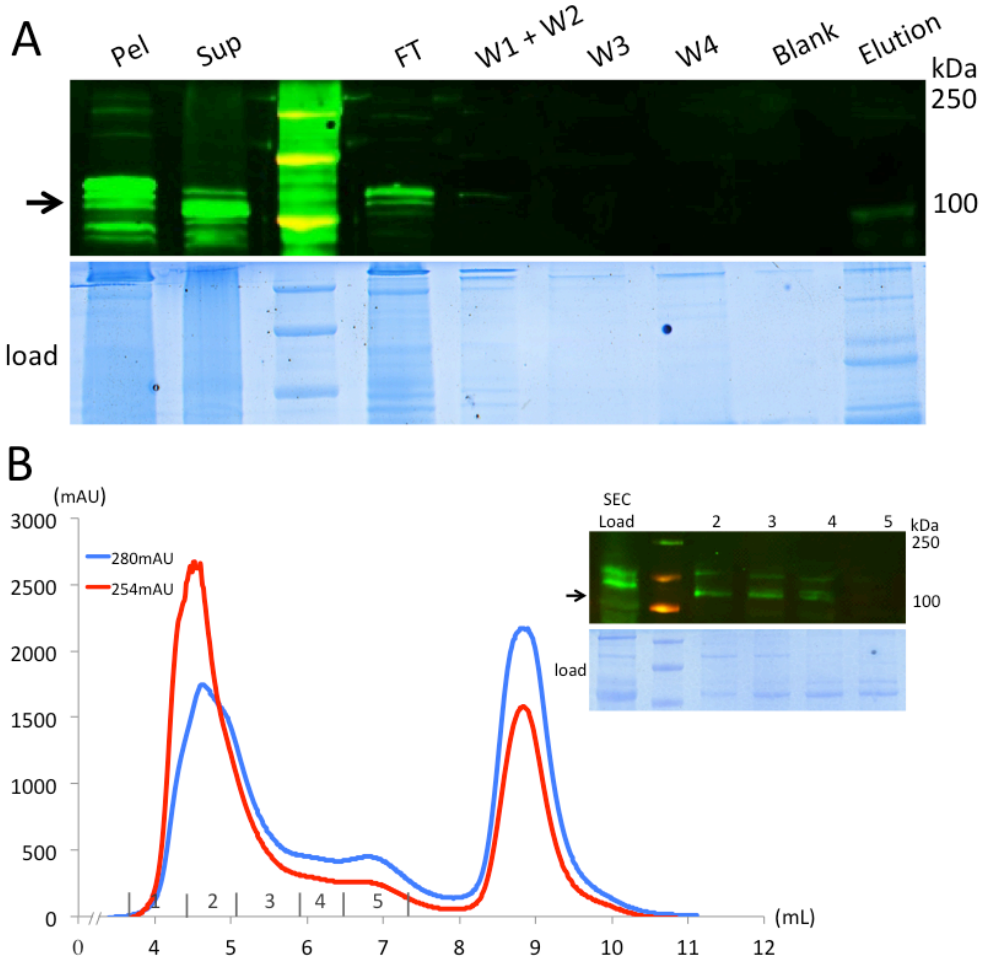

Figure 3.7 CtCESA1 expressed in Pichia yeast cell lines isolated from column purification and SEC. (A) Gravity flow column purification against the HIS tag (black arrow). Purification steps shown (Pel = pellet; Sup = supernatant; FT = flow through: W = wash). Protein load is indicated. During column purification the detergent is switched from LMNG+CHS to GDN. (B) CtCESA1 solubilized in GDN detergent is isolated from SEC. Fractions of protein taken from SEC shown above the X-axis and the western blot panel. Beginning volumes in the x-axis have been truncated for brevity. The protein load at time 0, from the elution fraction in (A), is indicated. Bottom panel represents protein load in each lane.

3.2.2 CtCESA1 shows glucosyltransferase activity

Glucosyltransferase activity of the purified CtCESA1 proteins in either a GDN detergent or an *E. coli* lipid environment was tested with ^{3}H-labelled UDP-glucose and one of the cations Mg^{+2}, Mn^{+2}, or Ca^{+2}, or with no cation added (Ø). Because no enzyme activity was expected in the presence of nothing or Ca^{+2}, radioactivity measured in these conditions was considered background level radiation for the assay.

ANOVA and Tukey post hoc tests were used to assess significant differences between the cation treatments within each purification group. In each case, assumptions for ANOVA were met based on tests for normality (Shapiro-Wilk $p > 0.25$) and homogeneity of variance (Levene's Test $p > 0.18$).

When exposing CtCESA1 isolated from insect cell membranes encapsulated in GDN detergent or in *E. coli* proteoliposomes, the treatments containing no cofactor (Ø) or Ca^{+2} showed no significant difference, and represented background radioactivity (~ $0.9 \pm 0.1 \times 10^2$ DPM) (mean ± standard deviation) where no enzyme activity is present (Figure 3.8). In the presence of Mg^{+2} there was no significant difference in radioactivity from background for CtCESA1 in GDN ($0.87 \pm 0.01 \times 10^2$ DPM) or CtCESA1 in proteoliposomes ($1.1 \pm 0.16 \times 10^2$ DPM). In the presence of Mn^{+2} there was a significant increase of ~ 0.5 x radioactivity above background for both CtCESA1 in GDN ($1.7 \pm 0.4 \times 10^2$ DPM; $F_{3,8} = 20.4$, $p \leq 0.01$) and CtCESA1 in proteoliposomes ($1.65 \pm 0.19 \times 10^2$ DPM; $F_{3,7} = 8.25$, $p \leq 0.01$) (Figure 3.8).

When exposing CtCESA1 isolated from Pichia cell membranes, nothing added and Ca^{+2} added showed no significant difference (Figure 3.8). This background radiation measured for CtCESA1 that is encapsulated in GDN detergent (~ 2 ± 0.2 x 10^2 DPM) and in *E. coli* proteoliposomes (~ 1.5 ± 0.2 x 10^2 DPM) is representative of no enzyme activity. In the presence of Mn^{+2}, there was no significant difference in radioactivity from background for CtCESA1 in GDN (2.38 ± 0.4 x 10^2 DPM) or CtCESA1 in proteoliposomes (1.56 ± 0.06 x 10^2 DPM). In the presence of Mg^{+2} there was a significant difference in radioactivity ($F_{3,7} = 33.81$, $p \leq 0.01$), ~ 3.5 x above background, for Pichia-expressed CtCESA1 in GDN (7.70 ± 1.3 x 10^2 DPM) (Figure 3.8), indicative of glucan synthase activity. This trend was seen again when Mg^{+2} is added to Pichia expressed CtCESA1 in *E. coli* proteoliposomes and there was a significant difference ($F_{3,8} = 63.96$, $p \leq 0.01$) in radioactivity (3.14 ± 0.26 x 10^2 DPM) ~ 2 x above background (Figure 3.8).

CtCESA1 shows measured radioactivity above background radiation, indicative of glucan synthase activity. This activity was found at lower levels for insect cell expressed CtCESA1 with the Mn^{+2} cofactor and higher levels for Pichia cell expressed CtCESA1 with the Mg^{+2} cofactor. Overall, the highest levels of glucan synthase activity were seen with CtCESA1 encapsulated from GDN, purified from Pichia.

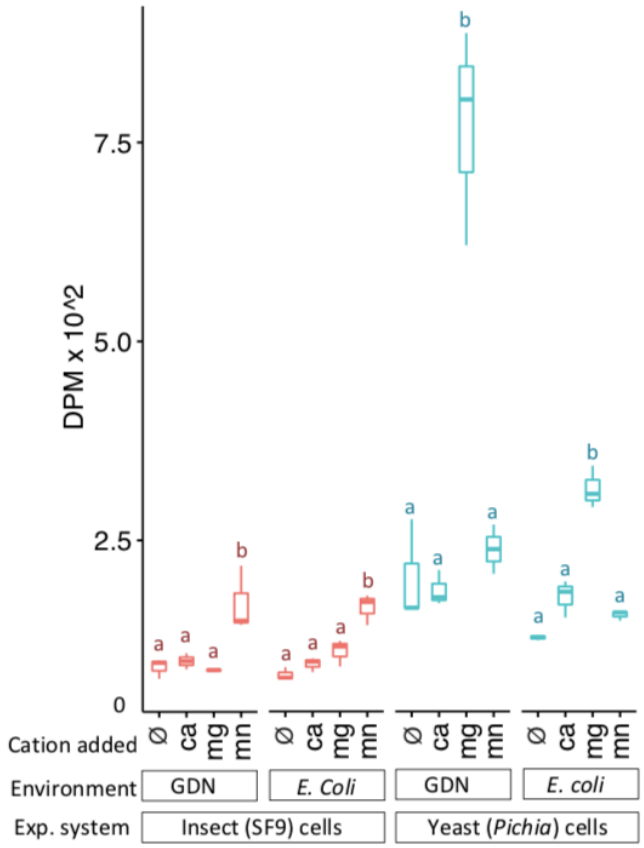

Figure 3.8 Glucosyltransferase activity of the red algal *C. tuberculosum* CESA1 is highest in yeast purified GDN. UDP-glucose ^{3}H-labeled tracer activity assays was measured with either nothing (Ø), Ca^{+2} (ca), Mg^{+2} (mg), or Mn^{+2} (mn). Glucan polymer formation measured hundreds of disintegrations per minute (DPM x 10^2). Purified CtCESA1 produced in either the insect (SF9) or yeast (Pichia*)* heterologous expression system, reconstituted in GDN detergent or proteoliposomes from total *E. coli* lipids. Values shown above each boxplot are the average of n = 3 measurements. Within each group statistically different means are indicated with letters.

Activities between the red algal *C. tuberculosum* CtCESA1 in GDN detergent isolated from a Pichia host, land plant eudicot *Populus tremula x tremuloides* PttCESA8 in GDN detergent isolated from an insect cell host, and bacterial *Rhodobacter sphaeroides* cellulose synthesizing BCSA/BCSB complex embedded in *E. coli* proteoliposomes were compared (Figure 3.9). Relative to background radiation (Ø = 2.01 ± 0.6 x 10^2 DPM), CtCESA1 showed ~ 3.5 x increase in activity in the presence of Mg^{+2} (7.70 ± 1.3 x 10^2 DPM) (Figure 3.9). Relative to background radiation (Ø = 1.1 ± 0.1 x 10^2 DPM), PttCESA8 showed ~ 50 x increase in activity in the presence of Mg^{+2} (115.18 ± 7.9 x 10^2 DPM) and ~ 100 x increase in activity in the presence of Mn^{+2} (220.81 ± 76.3 x 10^2 DPM) (Figure 3.9). Relative to background radiation (Ø = 2.80 x ± 0.2 x 10^2 DPM), the BCSA/BCSB complex showed an ~ 230 x increase in activity in the presence Mg^{+2} (699.77 ± 136.1 x 10^2 DPM). Here, clear differences were seen in the glucosyltransferase activities of cellulose synthases from different organismal origins presented with the least to most active being CtCESA1, PttCESA8, and BCSA/BCSB complex.

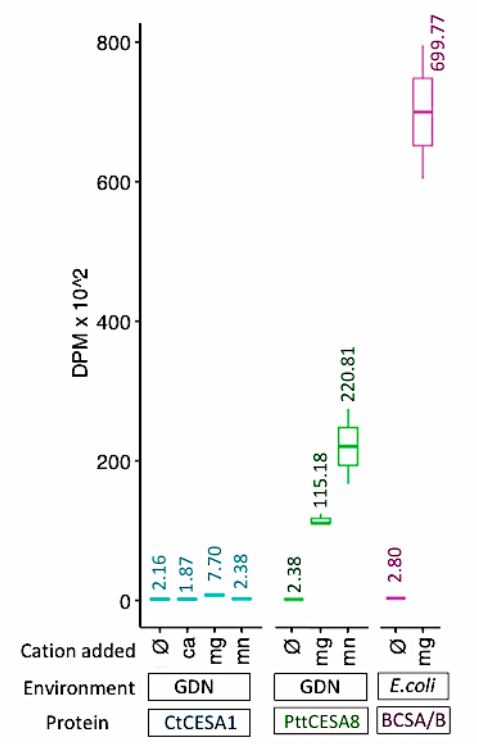

Figure 3.9 Glucosyltransferase activity of the red algal *C. tuberculosum* CESA1 is significantly lower than the woody eudicot *P. tremula x tremuloides* CESA8, and the bacterial *R. sphaeroides* BCSA/BCSB. UDP-glucose ^{3}H-labeled tracer activity assays with measured with either nothing (Ø), Ca^{+2} (ca), Mg^{+2} (mg), or Mn^{+2} (mn). Glucan polymer formation was measured as hundred disintegrations per minute ($DPMx10^{2}$). CtCESA1 and PttCESA8 were purified in GDN detergent and the BCSA/BCSB was reconstituted in *E. coli* lipid proteoliposomes. Values shown above each boxplot were the average of n = 3 for CtCESA1 measurements and n = 2 for PttCESA8 and BCSA/BCSB measurements.

3.3 Discussion

3.3.1 Variations in CtCESA1 activity

Cellulose synthases are the functional unit that makes glucan chains from UDP-activated glucose monomers. In this thesis, I present functional evidence that the nucleotide sequence that encodes the putative cellulose synthase protein from the red alga *C. tuberculosum* (CtCESA1) shows evidence of glucan synthase activity (Figure 1.11). This CtCESA1 activity shows relatively little variation between independent measurements. True cellulose polymers are composed solely of β-1,4-linked glucose. Though not conducted in this thesis, further carbohydrate linkage analysis should be conducted to confirm that glucose molecules are connected via β-1,4-linkage in the CtCESA1 samples. Previous samples of Poplar CESA8 heterologously expressed in Pichia (Purushotham et al., 2016) have variable batch-to-batch contamination of a putative *Pichia pastoris* 1,3-glucan synthase (GSL2). *P. patens* CESA5 samples heterologously expressed under the same conditions also reported batch-to-batch variation of *P. pastoris* GSL2 protein presence. Therefore, we cannot exclude that some activity seen may be due to the presence of this protein without linkage analysis data for samples derived from *P. pastoris* host cells. However, the fact that background activity was not observed under Ca^{+2} conditions, which usually stimulates callose synthases (Li et al., 1993), argues against callose synthase being responsible for the glucan synthase activity observed. In addition, such endogenous callose synthase activity would not explain the above-background activity of the CtCESA1 produced in insect cells, which are not known to contain callose synthases (Figure 3.8).

CtCESA1 activity occurs in the presence of either magnesium (Mg^{+2}) or manganese (Mn^{+2}), dependent on the recombinant expression host. CtCESA1 purified from expression in insect cells only showed activity in the presence of Mn^{+2}, whereas CtCESA1 purified from expression in yeast cells only showed activity in the presence of Mg^{+2}. The cation usage is coordinated by the DXD motif in the highly conserved catalytic cleft of the CESA protein. There is partiality for Mg^{+2} vs Mn^{+2} amongst CESAs, with bacterial CESAs preferring Mg^{+2} to Mn^{+2} (Omadjela et al., 2013) while land plant CESAs prefer Mn^{+2} to Mg^{+2} (Purushotham et al., 2016; Cho et al., 2017). Due to the high conservation between the catalytic cleft of all CESAs, likely slight electrostatic variations in proximity to the nitrogen and oxygen atoms of the D residues cause these minor differences in preference. Variations in protein folding may occur between CtCESA1 expressed from these two systems resulting in changes in the electrostatic interactions (Bock et al., 1999), which affect the preference for Mn^{+2} and Mg^{+2} seen.

The reconstitution of membrane proteins from detergent into a lipid environment is a common technique used to improve enzyme activity (Rigaud and Lévy, 2003). However, CtCESA1 activity was invariably lower when reconstituted in an *E. coli* lipid environment and higher for CtCESA1 encapsulated in GDN detergent (Figure 3.8). This is likely due to an experimental artifact and not biologically relevant, but it does suggest that currently unknown lipid requirements may exist that could increase the CtCESA1 glucan synthase activity. Activity assays of CtCESA1 in GDN detergent were performed directly after purification, whereas proteoliposome-reconstituted CtCESA1 activities were measured 24 h after the original purification. This delay, due to the time required for CtCESA1 incorporation into *E. coli* liposomes, may have caused both an increased time exposed to proteases lingering in solution

and, depending on the protein stability, loss of CtCESA1 functional protein shape resulting in lower activity levels.

3.3.2 *In vitro* rate of activity varies among CESAs from different organisms

The measured rate of glucosyltransferase activity differs among the *in vitro* purified CESAs, with the fastest to slowest activity being bacterial (BCSA), eudicot plant (PttCESA8), and red algal (CtCESA1) respectively. Granted, these differences in activity cannot be untangled from protein-specific experimental caveats that need to be considered, such as relative protein stability, degradation over time, or the absence of additional secondary modifications or co-factors for protein function. Though plant CESAs are demonstrably able to function without a secondary activator molecule or protein (Purushotham et al., 2016), this is not the case for all CESAs. In the case of the more distantly related bacterial cellulose synthase, the B subunit (BCSB) requires the allosteric activator molecule cyclic di-GMP for the BCSA/BCSB complex function, and negligible amounts of activity occur in its absence (Omadjela et al., 2013). The land plant, bacterial, and red algal CESAs have divergent protein evolutions (demonstrated in the CESA gene tree from chapter 2), likely originating at their prokaryotic endosymbiotic ancestor (Nobles et al., 2001). Therefore, it is not inconceivable that their requirements, or lack thereof, for allosteric protein activators could differ. The observed CtCESA1 activity is markedly lower than the bacterial and land plant CESAs. Perhaps the CtCESA1 protein could be regulated by an additional factor and have lower activity in its absence. This additional factor could be the unique domain at their N-terminus, the CBM48 domain discussed in chapter 2, predicted to bind

red algal starch (Matthews et al., 2010). Though this is speculative, ongoing work is being conducted to elucidate the function of this unique domain (Xue et al., in progress).

Under the assumption that the measured differences in CESA activity reflect true native activities *in vivo*, the differences in activity seen may be due to factors such as (1) the crystallinity of the cellulose produced and (2) the physiological demand for cellulose required by their growth pattern. Both explanations will be examined below.

The land plant and red algal CESAs may have lower activity as an evolutionary adaptation that results in the production of crystalline cellulose, whereas bacterial CESAs may have higher activity reflecting their production of amorphous cellulose. Cellulose exists in two general forms, crystalline cellulose that has ordered hydrogen bonding between neighboring glucan chains, and amorphous cellulose where glucan chains are bonded by irregularly spaced hydrogen bonds (Kondo and Sawatari, 1996). Many bacterial CESAs, such as the *Rhodobacter sphaeroides* BCSA/BCSB complex (Figure 3.9), produce short fibers of cellulose or amorphous cellulose that does not form higher-ordered structures or fibrils (Jonas and Farah, 1998; Omadjela et al., 2013). However there are Conversely, eukaryotic CESAs predominantly produce crystalline cellulose, exemplified by the protofilaments produced by PttCESA8 *in vitro* (Purushotham et al., 2016) and the cellulose microfibrils observed *in vivo* (Ding and Himmel, 2006). Moreover, *C. tuberculosum* and *Populus trichocarpa* have been demonstrated to have crystalline cellulose in antibody-labelling and nuclear magnetic resonance experiments on whole tissue and isolated cellulose respectively (Foston et al., 2009; Martone et al., 2019). Given the type of cellulose produced, slower rates of production may facilitate bonds forming at regular intervals, and higher rates of

production may not allow enough time for ordered hydrogen bonding to occur, resulting in amorphous cellulose. By correlation, we would expect that the bacterial BCSA/BCSB complex would have a drastically higher activity rate than their eukaryotic CESA counterparts as seen in Figure 3.9. However, there are some exceptions of bacteria, such as *Acetobacter xylinum*, that produce highly crystalline cellulose (Benziman et al., 1980). For this reason, the production of crystalline cellulose is likely one of multiple different rate-limiting factors.

Given the more comparable cellulose crystallinity of the eukaryotic CESAs, we still see that the land plant PttCESA8 is considerably more active than the red algal CtCESA1. In this case, the enzymatic rate of activity may reflect their physiological requirement for cellulose to maintain their cell growth and biomass production. Cellulose is a structural component in almost all cell walls for both *C. tuberculosum* and *P. tremula x tremuloides*. However, these organisms have drastically different cellulose deposition patterns and growth rates. In *C. tuberculosum*, cellulose is continuously deposited throughout the lifetime of the algae, comprising ~ 8 % of the primary walls and ~ 22 % of the secondary cell wall (Martone et al., 2019) with growth rates of ~ 2 cm/year in the field (Fisher and Martone 2014) to ~ 4.9 cm/year in the lab (Smith 1972). In contrast, trees in the genus *Populus* are comprised of ~ 50 % cellulose (reviewed in Balatinecz and Kretschmann 2001) with growth rates of 1.2 m/year to 1.8 m/year (Demeritt, 1981). Compared to *C. tuberculosum, Populus* species grow much faster, are composed of more cellulose, and would therefore require higher rates of cellulose production to maintain their biomass. The more active PttCESA8 relative to CtCESA1 could reflect these differences in cellulose demand, provided that the activity levels are determined at the protein level.

Here I propose a hypothesis where the activity levels of eukaryotic CESAs are an evolutionary adaptation of crystalline cellulose production and their cellulose output. This hypothesis could be tested by a more extensive survey of putative CESA enzyme activities within the Rhodophyta group.

3.3.3 CtCESA1 may form homo-oligomers

Eukaryotic CESAs are thought to associate in oligomeric complexes. Recent structural evidence from cryo-EM images of *Populus tremula x tremuloides* CESA8 indicates that CESAs homo-oligomerize into trimers (Purushotham et al., 2020). This oligomeric complex, or "lobe", in land plant CESAs measures 10.6 nm on average (Kimura et al., 1999; Nixon et al., 2016). Freeze fracture images of red algal protein complexes, presumed to be made up of CESAs, also identify "lobes" averaging 8.6 nm (Tsekos and Reiss, 1992) and 9.2 nm (Tsekos and Reiss, 1994). Considering that the land plant CESAs and red algal CESAs are relatively similar in protein size, ~ 110 kDa for both, and that the 10.6 nm particle in the CSC complex found in land plants is a trimer, it stands to reason that the ~ 9 nm particles found in red algae are likely also oligomers of either CESA dimers or trimers. In the western blots of the CtCESA1 particles, there are consistently higher molecular weight bands that are double the predicted ~ 110kDa CtCESA1 size (~ 240kDa for example in Figure 3.2C, Figure 3.4, and Figure 3.5A). This is perhaps evidence that, like the PttCESA8 that is able to homo-trimerize (Purushotham et al., 2020), the CtCESA1 may be capable of forming homo-oligomers. This would likely explain the size of the CESA containing "lobes" seen in previous freeze-fracture images of red algal CESAs. Indeed, a higher molecular weight band likely indicative of a PttCESA8 dimer is also occasionally seen in

western blots of PttCESA8 purifications (seen in Figure 1A from Purushotham et al. 2016). However, the homo-oligomerization of CtCESA1 would need to be formally tested. This could be tested by expressing two forms of CtCESA1, each tagged with a different epitope, and seeing if the higher molecular weight band would correspond to the presence of both epitopes, therefore indicating homodimerization. As the distinct evolution of CESAs from land plants and red algae likely converge at their prokaryotic ancestor and that the bacterial BCSA/BCSB monomeric complex best represents that prokaryotic state, it is likely that CESA oligomerization evolved in parallel between the two lineages.

3.4 Methods

3.4.1 CtCESA1 protein sequence

The CtCESA1 protein sequence can be found in the transcriptome dataset, which has been deposited in the European Nucleotide Archive (project number PRJEB39919).

>CtCESA1

MANIGPSGGPIQGADMPYVPEPADPRMLADNMSDAGSSTSSGMLSDVRSRRTGSIDTTP
SAGPRGSLRSRHTRASMHSKAPTNYEEVPNALSLFEWNCGGRNVYLTGSWDNYTEKIP
MESIQPGNFRCTVKVPQERLEFKFIVDGVEKFNPDYPTIYTETGERVNVKHVDPDGKKN
SAGTVRKIVSKISGLDMYSPFHMAETLSMIIFRVFYILTIPAAFYYFYWLSWVGGNRNDA
PVSWIVFIIAEILSFLSAMIGLFGMWKPVKRKWRSLDSLKPPLPEADWPSVDICIAHYKEP
PEQLRDTIRAALRLDYPSHLIRIIIADDGYFATPKSVERSQLGLDMYQLLAEEAGYDPLLE

EVMNDQGLVEHYTVLADDEILRPDAAKECHVFDFGPFGDDMYAPGALPRLSLISRVKP
ADHHNKAGNINNVLMNAGTDGKLILFLDADMRPTENMLLRMVPLLLEEMRDDAVENT
LMLDDDPEIGRGVNTSWRVNRDVAFIQAPQRFHNVDNADIMAHRNALFYDGIIRGRDG
FGMTPFVGTNALWRREVLYEINGFVYGSVTEDTLTSNEVHRRGYISKYAAEDLAWGEA
PISVAAAMLQRQRWAKGAIMNGMKIFEKAADEKKRALIARRRGEIDEFYEYRRHGRRP
NNGFVRAMFWLDSTLYPLLGVAAYMYMFVAMYYLIKAQPPIAPDNIYDLASAFITYYII
RYFAFFAAYSGVAPIDVLRGQETWYGYNICHVVGMWDALMGAKMSWVANTGQRSRR
NWMEWVNILICGLMVFGIIFRLVAFLYFEKGCQPWQTFGAVGFGFYIFGHMWPMAAISL
NERLNPSQDDDTIGEPYQLPTPLIYAALTILVVLVLSRWAETACGRNSETVEGRRF

This protein sequence was determined by *in silico* translation of the nucleotide sequence obtained from sequencing of the *CtCESA1* cDNA. This same amino acid sequence was used for both insect cell expression and Pichia cell expression. The nucleotide sequence used in insect cell expression had a red algal based codon whereas the nucleotide sequence used in Pichia cell expression had a Pichia based codon.

3.4.2 Expression of CtCESA1 in SF9 insect cells

3.4.2.1 *CtCESA1*-pACEBAC1 construct design and bacmid preparation

First, a modified *CtCESA1* sequence was isolated for later engineering in a baculovirus. The *CtCESA1* sequence was produced by PCR amplification from a *CtCESA1* containing plasmid (pdonrZeo:*CtCESA1*), originally isolated from cDNA of environmentally collected *C. tuberculosum* as described in the chapter 4 methods. CtCESA1 was engineered with additional 12xHIS tags and cut sites by primer addition in PCR (Figure 3.10). PCR was performed in two separate amplification cycles using Phusion High-Fidelity DNA polymerase in Phusion GC buffer (New England Biolabs) and 5 % ethylene glycol with an annealing temperature of 65 °C. This first round of amplification that used the forward primer 1 and reverse primer was PCR cleaned (QIAquick PCR purification kit) and used for the second round of PCR amplification with forward primer 2 and the reverse primer (Figure 3.10), which was also PCR cleaned. At the 5' end of *CtCESA1*, a 12x histidine (HIS) tag was added to facilitate protein purification. A Tobacco Etch Virus (TEV) protease cut site was added as a linker between the CtCESA1 protein and the histidine tag, to both prevent the 12xHIS tag from interfering with later CtCESA1 protein folding, and to improve protein purification by increasing the HIS tag's mobility to bind its affinity partner during column purification. A Kozac sequence, a common eukaryotic translation initiation site, was added directly upstream of the 12xHIS coding sequence to improve mRNA translation rates. SacI and HINDIII restriction cut sites were added to flank the entire sequence for restriction digest based cloning into the shuttle vector. The strong viral polyhedron promoter, provided by the shuttle vector, drives the entire coding sequence (Figure 3.11A).

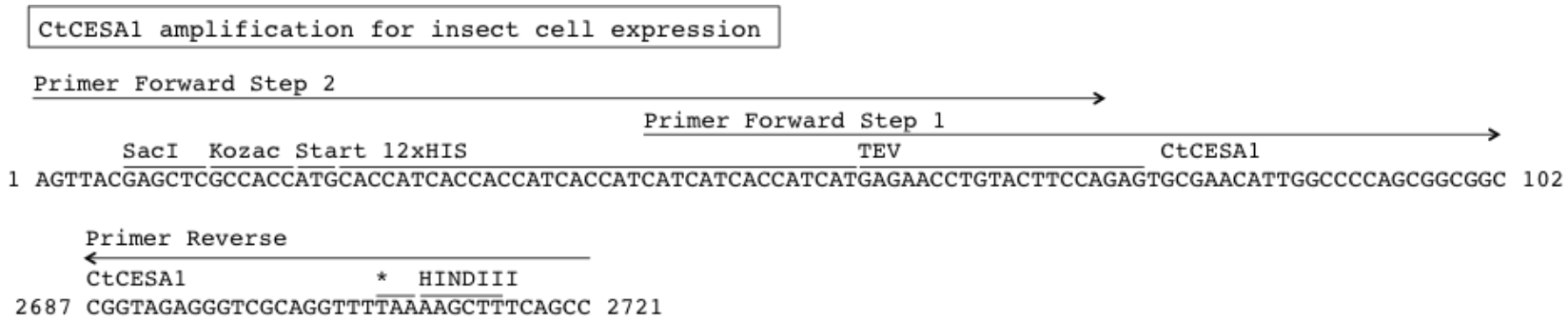

Figure 3.10 Nucleotide sequences flanking the 5' (1 - 102) and 3' (2687 - 2721) *CtCESA1* coding sequence. The positions of key features are indicated directly above the nucleotide sequence: two forward primers and the reverse primer were used to produce the amplicon in a two-step PCR amplification.

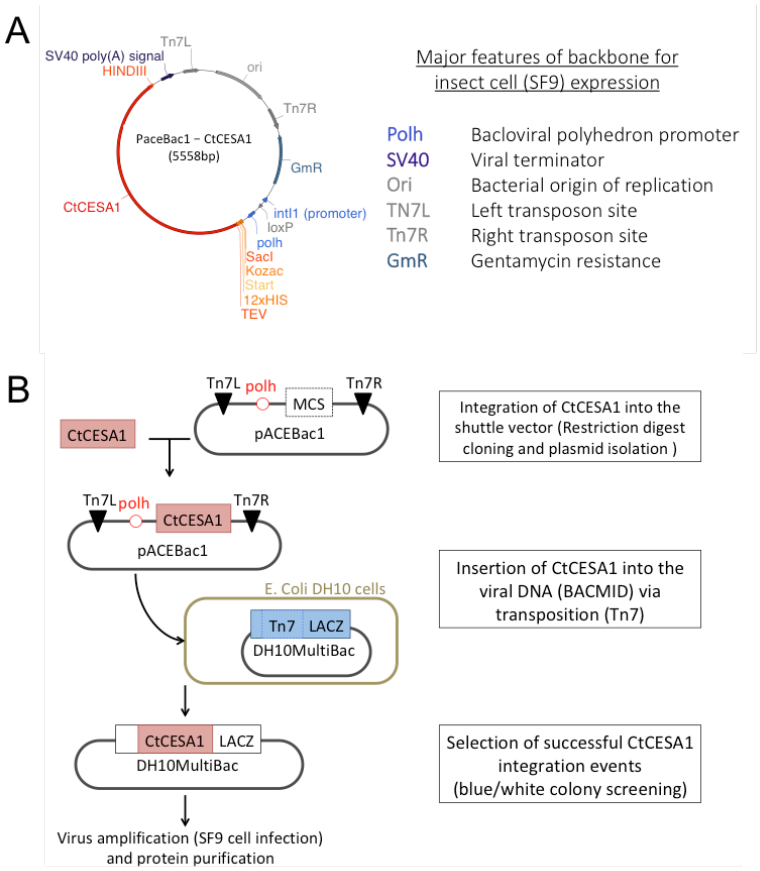

Figure 3.11 Overview of *C. tuberculosum CESA1* (*CtCESA1*) viral bacmid preparation for protein expression in *Spodoptera frugiperda* (SF9) insect cells. (A) Vector map of *CtCESA1* in the PaceBac1 backbone used to shuttle the *CtCESA1* gene into the viral bacmid for insect cell transfection. (B) Overview of viral DNA preparation. *CtCESA1* inserted into the pACEBac1 vector through the multiple cloning site (MCS). Verified *CtCESA1*-pACEBac1 constructs were introduced to the DH10 cells that harbor the acceptor baculoviral DNA, the bacmid, and the transposase required for gene integration into the bacmid. In DH10 cells, *CtCESA1* integrated from pACEBac1 into the DH10MultiBac bacmid via Tn7 transposition. Tn7 integration site was located within a lacZ gene. As a result, CtCESA1 successfully integrated into the bacmid were selected with blue/white screening, as blue

colonies contain no successful integration events, while white colonies containing at least one successful integration event due to disruption of the *lacZ* gene. Bacmids with the successfully integrated *CtCESA1* were harvested to later transfect and produce protein in insect cell lines.

Several steps were required to insert the modified *CtCESA1* into a virus for insect cell infection. The full-length amplicon was cloned into the multiple cloning site (MCS) of the PaceBac1 shuttle vector (Figure 3.11). PaceBac1 and the tagged *CtCESA1* amplicon were digested with SacI and HindIII (New England Biolabs) for 1 h then run on a 1 % gel and correct fragments were gel purified and ligated together (T4 DNA Ligase New England Biolabs) for 1 h. The ligated reaction was heat inactivated at 65 °C and 1 µL was added per 25 µL Top10 *E. coli* competent cells for heat shock transformation. Colonies were selected using gentamycin antibiotic selection and verified with sequencing (GENEWIZ). To obtain enough plasmid, these colonies were cultured in LB and gentamycin and their plasmids were subsequently harvested (Qiagen miniprep kit). The *CtCESA1* that was then integrated from the pACEBac1 vector into an engineered viral DNA, called the bacmid, via transposition sites, in DH10 *E. coli* host cells (Figure 3.11B). To insert the *CtCESA1* gene into the viral DNA (bacmid) for later transfection of the insect cells, 1000 ng/µL of the pACEBac1 acceptor vector carrying the tagged *CtCESA1* gene was inserted into the chemically competent DH10MultiBac *E. coli* cells (Fig 2B) using heat shock transformation. The pACEBac1 shuttle vector uses Tn7 sites to integrate its content into the viral bacmid using Tn7 transposition (Figure 3.11). The DH10 *E. coli* cells have the bacmid and transposase enzyme required for integration of DNA transfer between Tn7 sites (Figure 3.11B). The transformed DH10MultiBac cells with the pACEBac1 vector containing the tagged

CtCESA1 was selected on dark grown LB plates with the antibiotics kanamycin/gentamycin/tetracyclin blue/white selection media blu-gal and induction chemical IPTG and grown overnight at 37 °C. After one day, blue and white colonies were restreaked onto a new plate and grown overnight at 37 °C to ensure stable integration of the *CtCESA1* insert. The bacmid DNA containing the tagged *CtCESA1* was harvested using the Qiagen plasmid prep kit without the column, precipitating with isopropanol, and resuspending in 50 µL of elution buffer. To identify successful integrations, the Tn7 transposon integration site on the bacmid is located in a lacZ-alpha site which, when disrupted, causes the cells to become white (Figure 3.11B). To identify bacmids containing at least one successfully integrated *CtCESA1* gene, DH10 *E. coli* colonies were blue/white screened. Each independent Tn7 transposition can contain a variable number of *CtCESA1* genes transposed, which may produce more variable amounts of protein. For this reason, bacmids from four independent integrations of *CtCESA1*, four white colonies, were isolated for further protein production in insect cells.

3.4.2.2 SF9 insect cell culturing and viability

To maintain a supply of uninfected insect cell culture for continued use, the cells were split into 1E10 three times per week quantified by cell counting. An active culture of uninfected insect cells was counted using a hemocytometer counting grid and diluted to a 1E10 cell count concentration with the SF9 media (Figure 3.2A). Only live cells were counted in this process. Dead cells were distinguished by adding trypan blue that permeates the membrane of dead cells dying them blue (Figure 3.2). Cell viability was also counted ($\sum$ cells alive / $\sum$ cells dead) with a high viability indicating a healthy culture. A minimum of 100 cells were counted for confident

predictions. Maintained insect cell cultures were grown in vented Erlenmeyer flasks at 27 °C and 120 RPM and re-cultured a maximum of 40 times.

3.4.2.3 Virus propagation for SF9 insect cell transfection

To produce a working amount virus to infect the insect cells for protein purification, the virus must first be propagated to large enough quantities by infecting smaller amounts of insect cells (Roest et al., 2016). Isolated bacmid DNA containing the tagged *CtCESA1* was used to infect insect cell cultures to create transfected cell lines (P0 stage). DNA from four separate colonies of bacmid DNA was used for transfection. Per one sample of bacmid DNA, 50 µL of SF9 transfection media and 3 µL of Fugene transfection reagent (Promega) was incubated for 5 mins then added to 5 µL of the bacmid and 50 µL of SF9 transfection media. The combined mixture was incubated at room temperature for 30 mins. This mixture was later added to insect cells at 1E10 cell count concentration.

The insect cells were then added to a deep 24 well plate (for plate layout see Figure 3B). The bacmid carrying *CtCESA1* and transfection media mixture were then added to the insect cells drop by drop. A control well with media but no bacmid infection was also grown on the same plate. The mixture was covered with adhesive foil and gown in a shaking incubator at 27 °C and 300 RPM for 4 days. A separate tray with water was also added into the incubator to increase the humidity.

The virus containing CtCESA1 was further propagated to produce more virions (called the P1 stage). 170 µL of P0 was added to 4 mL of freshly diluted insect cells at 1E10 concentration in a deep 24 well plate and grown at 27 °C and 300 RPM for 4 days. The remaining unused portion of P0 was used to check and visualize CtCESA1 expression in the cell cultures by diluting the samples to an OD of 3, resuspending in protein buffer solution (PBS), adding 1 x loading buffer and western blotted.

The virus containing *CtCESA1* was further propagated to produce more virions (called the P2 stage). This P2 had a critical amount of virions and was used for all subsequent insect cell infections for protein purification. Visual queues of infection include increasing number of dead (blue) cells, and enlargement of infected live cells (Figure 3.2A). P2 virions were harvested by passing through a 0.2 micron filter (VWR 97066-200) under vacuum filtration and then stored at 4 °C.

3.4.2.4 Large scale protein production from SF9 insect cells

CtCESA1 was grown in large scale insect cell cultures for protein purification. 10 mL of *CtCESA1* containing P2 virus was added per 450 mL of SF9 cells at 2E6 density with a total of 2 L of cells. Infected cells were grown in ventilated PET bottles shaking 220 RPM at 30 °C. Cells were harvested at 50 % - 70 % viability indicating high infection and protein production but prior to complications such as protein degradation due to the release of lytic contents during cell death. Cells were harvested by centrifugation at 4500 g for 21 mins at 21 °C (JLA 8100p rotor Avant;

J20 XP centrifuge). The 2 L of pellet was transferred to a 15 mL conical tube, flash frozen in liquid nitrogen, and stored at -80 °C until use.

3.4.3 Expression of CtCESA1 in *Pichia pastoris* yeast cells

3.4.3.1 *CtCESA1*-PICZA construct design for expression in *Pichia pastoris* and transformation into *Pichia pastoris* cell lines

CtCESA1 was designed with an additional Strep (Schmidt and Skerra, 2007) and 12xHIS tags at the N terminus for protein purification and codon optimized for expression in yeast (Figure 3.12). A TEV sequence was added in between the protein purification tags and the *CtCESA1* sequence to prevent steric hindrance during protein folding. The alcohol oxidase 1 (AOX1) constitutive promoter upstream was used to drive the gene expression (Figure 3.12) (reviewed by Vogl and Glieder 2013). To improve translation efficiency, the entire *CtCESA1* sequence was codon optimized for expression in yeast and commercially synthesized into the PiczA *Pichia* expression vector (Thermo Fisher GeneArt). The PiczA vector is for non-secretory protein expression. To obtain enough plasmid, the synthesized codon optimized *CtCESA1* in PiczA was then transformed into *E. coli* Top10 cells, cultured in 120 mL LB and Zeocin, and their plasmids were subsequently harvested using the Qiagen maxiprep kit. A large amount of plasmid (~ 3 k ng/ μL) was utilized for yeast cell transformation.

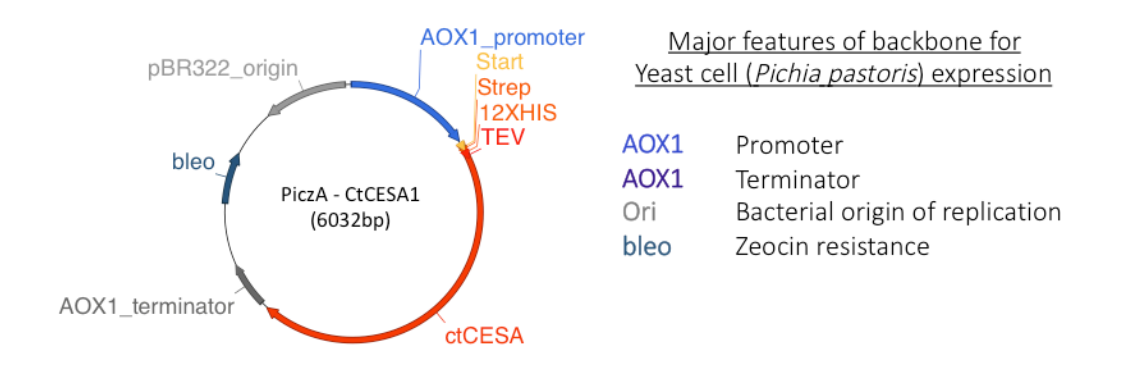

Figure 3.12 Vector map of *CtCESA1* in the PiczA backbone used for expression in *Pichia pastoris*. Major features of the vector backbone are indicated on the vector and written to the right. The PiczA is directly transformed into the Pichia cells where it drives the expression of the CtCESA1 protein with a highly active alcohol oxidase (AOX1) promoter.

To prepare the CtCESA1 containing PiczA vector for transformation into the yeast cells for expression, the 30 µg of plasmid was linearized with PMEI (New England Biolabs), incubated at 37 °C overnight and then cleaned the next day (QIAquick PCR purification kit).

To prepare the Pichia methanol metabolizing strain for transformation a sufficient amount of untransformed electrocompetent Pichia needed to be produced. To culture sufficient amounts of Pichia, a colony was picked and inoculation into a 50 mL of yeast extract peptone dextrose (YEPD) media (1 % yeast extract, 2 % peptone, 2 % dextrose/glucose) and grown (220 RPM shaking; 30 °C) until the culture reached early stationary phase of growth (~0.6E8 cells/ml). Cells were then harvested in a sterile 50 mL conical tube by centrifugation (4 k RPM; 10 mins). The supernatant was discarded. The pellet was subsequently washed a total of two times with 40 mL cold pre-autoclaved Milli-Q H2O by resuspension then pelleted by centrifugation (4 k RPM;

10 mins). The supernatant was removed and the pellet was washed with 5 mL of cold 1 M sorbitol by resuspension and pelleted (4 k RPM; 10 mins). The supernatant was discarded and the pellet was resuspended in 150 µL of cold 1 M sorbitol. The *CtCESA1* encoding PicZa was then transformed into the *Pichia* cells by electroporation. 5 µg of previously prepared PMEI digested PicZA containing the *CtCESA1* sequence was incubated with 40 µL of Pichia in a cuvette and electroshocked (V = 1.5k V, 25 µF, 200 Ohms, 7.8 milliseconds). 3 mL of cold 1 M sorbitol was added directly after, the solution was left at 30 °C for 1 h and then spread on selection plates (1 % yeast extract, 2 % peptone, 2 % dextrose (glucose), 1 M sorbitol, 2 % agar) containing 100 µg/ml Zeocin or 500 µg/ml Zeocin and grown for 3 – 10 days at 30 °C. Increasing Zeocin selection concentration was used to choose colonies with multiple insertion events of DNA from the plasmid into the yeast genome via recombination at the AOX1 site. Pichia transformed with *CtCESA1* gene was cultured and stored as glycerol stocks for inoculating later large-scale cultures. To produce the initial culture, multiple colonies were selected and grown overnight in 250 mL of YEPD selection media (100 µg/ml zeocin) (30 °C; shaking 220 RPM). To make the glycerol stocks 500 µL of sample was added to 500 µL of 50 % glycerol topped with autoclave water, flash frozen, and stored in -80 °C.

To produce large-scale cultures of CtCESA1 expressing Pichia, starter cultures were produced from glycerol stocks and subsequently scaled up into larger cultures. To produce the initial starter cultures, 2 or 3 glycerol stocks of CtCESA1 producing *Pichia* was inoculated into 400 mL of YEPD with 100 µm/µL Zeocin and grown overnight (30 °C; 220 RPM). Once at OD 2 - 6, indicating log phase of growth, the starter culture was then pelleted (4 k RPM; 16 °C; 10 mins), supernatant removed, and resuspended in filter sterilized Yeast Nitrogen Base (Thermofisher)

with ammonium sulfate and 5 % methanol (YNBM). Methanol added here and beyond this point is used to induce the AOX1 promoter to begin protein expression of CtCESA1. The resuspended starter culture was then added to equally to 7 flasks containing approximately 150 mL – 200 mL of YNBM. The concentration of cells at the start of the large-scale cultures was measured to ~ 0.5 OD_{600}. The starter culture was grown for 2.5 days (30 °C; shaking 220 RPM) with 0.5 % methanol/total volume added. The culture was then pelleted by centrifugation (4.5 k RPM; 14 °C; 20 mins), the supernatant removed, the pellet flash frozen in liquid nitrogen, and then stored at -80 °C until protein purification.

3.4.4 Detergent screen of membranes containing heterologous expressed CtCESA1

Detergents used and their working concentrations were (Fig 4B): 40 mM DDM = Dodecyl Maltoside; 40 mM LDAO = Lauryldimethylamine-N-Oxide; 40 mM NaC = Sodium cholate; 2 % Triton-X 100; 2 % L + 0.4 % C = Lauryl Maltose Neopentyl Glycol + Cholesteryl Hemisuccinate; 2 % SDS = Sodium dodecyl sulfate. All detergents except for SDS were sol-grade (from Anatrace), resuspended in water, and rotated at 4 °C to resuspend. These detergent tests were done on both insect cells and yeast cells expressing CtCESA1.

First, the membrane fraction of the cells expressing CtCESA1 was isolated. 50 mL of cold membrane resuspension buffer (MRB = 20 mM tris + 100 mM NaCl in distilled H20 at 7.5 pH) was added to a frozen cell pellet of CtCESA1. Half a crushed protease inhibitor pellet (cOmplete ULTRA tablet Roche) and the frozen CtCESA1 cell pellet in MRB was resuspend with a dounce. The resuspended solution was lysed by passing through a microfluidizer at 30 k PSI three times.

The lysate was spun down 42 k RPM for 1 h at 4 °C (Ti 500 rotor, Beckman Optima LE 80K ultracentrifuge). The supernatant containing the cellular contents was discarded and the pellet containing the lysed membrane fraction with the embedded CtCESA1 was kept for subsequent detergent testing. The membrane pellet was resuspended in a cold dounce with MRB. To resuspend the samples with the detergent, 300 µL of resuspended cells, the appropriate amount of detergent, and MRB to 700 µL was dounced with a clean cold dounce for each sample of insect cell membranes with CtCESA1. This process was also done on cell membranes with PttCESA8 using L+C detergent only. Samples were then transferred to 1.5 mL ultracentrifuge polypropylene tubes and incubated rotating for 1 h at 4 °C except the samples with SDS that were incubated at room temperature to avoid detergent precipitation. The samples were then spun at 65 k RPM at 4 °C for 30 mins to separate out the detergent solubilized protein in the supernatant from the insolubilized protein remaining in the membrane fraction pellet. The top half fraction of the supernatant and pellet fraction with 2 % SDS of all samples were western blotted. L+C was used for all further protein purification with CtCESA1 purifications.

3.4.5 Protein purification of CtCESA1 from insect or yeast cell membranes

An overview of the CtCESA1 purification from heterologous cells membranes requires lysing harvested cells, solubilization of CtCESA1 from the cell membranes, a blunt purification of CtCESA1 using the histidine tagged affinity to a Ni^{2+} column, and a refined purification of the protein with size exclusion chromatography (SEC). The detergent LMNG + CHS is used to solubilize the protein initially, but is later switched out of the solution for GDN, a synthetic derivative of digitonin. GDN has been shown to be superior in maintaining native protein

structure and integrity in comparative detergent studies (Chae et al. 2012) and we use it here to encase CtCESA1 during enzyme activity assays.

3.4.5.1 Lysis of host cells expressing CtCESA1

The initial lysis steps are similar to the extraction method for the detergent test. 1.5 L of frozen pelleted CtCESA1 expressing insect cells was warmed and dounced with MRB and 1.5 tablets of protease inhibitor to resuspend. Cells were then lysed in a microfluidizer at 30 k PSI three times and spun down at 10 k g for 10 mins at 4 °C to separate the membrane fraction cellular debris (JA20 rotor; Avanti J20XP centrifuge). The supernatant was decanted and then spun again at 42 k RPM for 1 h at 4 °C (ti45 rotor; Avanti J20XP centrifuge). The supernatant was discarded and the pellet with the membrane fraction was dounced with solubilization buffer (40 mL of 1 % LMNG + 0.2 % CHS, 5 mM β-mercaptoethanol, and MRB to volume). This solution was left to solubilize rocking for 1 h at 4 °C and then spun down at 42K RPM for 30 mins at 4 °C to separate insoluble protein in the pellet from solubilized CtCESA1 in solution.

3.4.5.2 Gravity flow column purification of CtCESA1 against the HIS tag

Prior to gravity flow column purification, HIS tagged CtCESA1, 2.5 mL of nickel beads (HisPur Ni-NTA Resin Thermofisher) were washed with ddH20 and 20 mM of imidazole. 20 mM of imidazole was also added to the supernatant containing solubilized CtCESA1 and the prepared nickel beads and left to batch bind rocking for 1 h at 4 °C. Imidazole competitively binds to the nickel beads and adding it into batch binding reduces contamination by preventing lower affinity

proteins from adhering to the nickel beads. After batch binding with the nickel beads, this first flow through solution was passed through a gravity flow column. The beads were subsequently washed with 4 solutions and eluted (wash 1 = 20 mM imidazole; wash 2 = 20 mM imidizol; wash 3 = 1 M NaCl; wash 4 = 40 mM imidizol; elution = 400 mM imidizol; to volume with MRB + 0.03 % GDN + 5 mM β-mercaptoethanol for each solution). The pellet, supernatant, and nickel column purification steps were western blotted.

3.4.5.3 Size exclusion chromatography for CtCESA1 protein purification

CtCESA1 isolation was further refined using size exclusion chromatography (SEC). To prepare the sample, the elution fraction from purified insect cells was concentrated to 0.5 mL using a 100 kDa concentrator (Amicon™ Ultra Centrifugal Filter Units) at 4 k RPM and 4 °C in a swinging bucket tabletop centrifuge. The concentrated protein was passed through a 0.2 μm cellulose filter column on a tabletop centrifuge to remove large impurities or protein aggregates that cause issues with internal SEC pressure control. This solution was injected into the SEC and size gradient column (HiLoad 16/60 Superdex 200 prep grade) pre-equilibrated with gel filtration buffer (GFB = 0.03 % GDN + 5 mM β-mercaptoethanol in MRB filtered through 0.2 μm cellulose acetate to remove large debris). Fractions were collected as the protein solution ran through the column. These fractions were western blotted. Appropriate fractions containing CtCESA1 were concentrated to approximately 0.13 mg/mL using a 100 kDa concentrator at 4 k RPM and 4 °C. This solution was used for subsequent enzyme activity assays

3.4.6 Reconstitution of CtCESA1 into proteoliposomes

The purified CtCESA1 protein was placed in two separate conditions that were tested for glucosyltransferase activity. In the first condition, CtCESA1 surrounded by GDN detergent was used to test for protein activity within the hour after SEC purification and concentration. In the second condition, CtCESA1 was reconstituted from GDN detergent to *E. coli* lipids. CtCESA1 surrounded in the organismal total membrane lipids would better mimic the native functional environment of the protein. To reconstitute CtCESA1 in these *E. coli* proteoliposomes, 2 volumes of CtCESA1 in GDN detergent from the concentrated SEC fractions were incubated with 4 mg/mL of *E. coli* total lipid extract (in 40 mM DDM; Avanti) and Bio-Beads (Bio-Rad) rotating overnight at 4 °C. The samples were visually inspected for increased turbidity, indicating successfully formed proteoliposomes. The supernatant containing CtCESA1 reconstituted in proteoliposomes were used for subsequent activity assays.

3.4.7 Glucosyltransferase activity assay of CtCESA1, PttCESA8, and BCSA/BCSB

To test for glucosyltransferase activity purified protein 5 µL of detergent enclosed CtCESA1 or 10 µL of proteoliposome reconstituted CtCESA1 was incubated for at least 10 h at 30 °C with 50 µM UDP-glucose, 0.25 µCi radioactive ^{3}H-labelled glucose (Perkin Elmer; [6-^{3}H]-Uridine Diphospho-D-Glucose), 500 µM of supplied cation (Ca^{+2}, Mg^{+2}, Mn^{+2}), or nothing (Ø), and topped to the final volume with the GFB. Purified PttCESA8 and *E. coli* BCSA/BCSB protein were tested in the same conditions as known functional glucosyltransferases. *E. coli* BCSA/BCSB protein had an 30 µM additional cyclic di-GMP added. Reactions were stopped by

addition of 100 μM of SDS. Thin layer chromatography (TLC) was then used to separate polymeric glucans from monomeric glucose. During TLC, one of each cation treatment sample (i.e. Ø , Ca^{+2}, Mg^{+2}, Mn^{+2}) were blotted separately and run together. The sample was blotted onto chromatography paper (Whatman-2MM) developed in a mobile phase of 60 % ethanol. Samples were then dried and submerged in scintillation fluid and incorporation of radioactive ^{3}H-glucose was measured with a scintillator.

3.4.8 SDS-PAGE and western blotting conditions

To resolve the full length CtCESA1, samples in reducing loading buffer with β-mercaptoethanol were run on 10 % SDS-PAGE gels (resolving gel: 10 % acrylamide, 0.375 M tris pH 8.8, 0.1 % SDS, 0.1 % NH_4SO_5, 4 % TEMED; stacking gel: 5 % acrylamide, 0.25 M tris pH 6.8, 0.1 % SDS, 0.1 % NH_4SO_5, 4 % TEMED) at 100 V for 1 h. Samples were loaded identically in two simultaneously run 10 % gels. One gel was stained with Coomassie Blue to visualize the protein load. The other gel was used for a western blot.

For Coomassie stained gels, gels were submerged in staining solution (0.1 % Coomassie Blue R-250, 20 % MeOH, 10 % HOAc), microwaved briefly, and incubated rocking for 5 mins. Samples were subsequently destained (50 % MeOH, 10 % HOAc) until bands were clear with no background dye present. Gels were imaged on a flatbed scanner.

For western blotting, the proteins were transferred from the SDS-PAGE gel onto a nitrocellulose film by wet transfer at 100 V for 1 h. After transferring, the blot was washed 1 x with TBS-T (10

mM Tris pH 7.4, 150 mM NaCl, 0.05 % tween-20), blocked in 5 % milk in TBS-T for 40 mins to reduce unspecific protein binding, washed 2 x with TBS-T for 10 mins each, then incubated with incubate with primary mouse anti-HIS or anti-Strep at 4 °C rocking overnight. The primary antibodies were diluted 1:1000 in 5 % bovine serum albumen (BSA) and 1 x TBS-T. The next day the blot was washed 2 x with TBS-T for 10 mins each, then incubated with secondary antibody for 1 h at room temperature (1:3000 dilution). The blot was then washed 4 x with TBS-T for 10 mins each before imaging. For detergent screen samples, an additional 2 % SDS was added the supernatant and pellet samples prior to loading on SDS-PAGE.

Chapter 4: The red algal CtCESA1 and land plant CESAs from *Arabidopsis thaliana* are not functionally redundant but share some functional similarities

4.1 Introduction

Cellulose production is a complex multi-layered process that has primarily been studied in the model plant *Arabidopsis thaliana.* Cellulose is produced at the plasma membrane (PM) by cellulose synthase enzymes (CESAs) organized into multimeric cellulose synthase complexes (CSC). In *A. thaliana*, CESAs that form the cellulose synthase complexes in primary cell walls are encoded by a different set of genes *(CESA1, 3*, and the partially redundant *CESA2, 5, 6, or 9*) than the *CESA* found in secondary cell walls (*CESA 4, 7, 8*) (Table 4.1) (reviewed in McFarlane et al., 2014). Mutant studies have demonstrated that these CESA types are functionally distinct and cannot fully compensate for one another (Kumar et al., 2017; Park and Ding, 2020) but have minor partial functional compatibility between some CESAs (Carroll et al., 2012; Kumar et al., 2017). When describing the CESA encoding genes, capitalized names represent the wild type copy while lowercase represents a mutant allele. A single nucleotide substitution causing a premature stop codon and truncated protein in the primary cell wall CESA encoding gene *CESA6*, in a mutant allele called *procuste1-1* (*cesa6*) causes decreased hypocotyl lengths in dark-grown, etiolated, seedlings (Table 4.1) (Fagard et al., 2000; MacKinnon et al., 2006). Insertional mutations (T-DNA) that disrupt genes *CESA3* and *CESA1* encoding primary cell wall CESAs (*cesa3-2* or *cesa1-1*) are gametophytic lethal, causing deformations in the haploid pollen and pollen tube formation that render their pollen non-viable (Table 4.1) (Persson et al., 2007b).

Mutations in any secondary cell wall CESA encoding gene *CESA4, 7* or *8* cause collapsed xylem, called an irregular xylem (*irx*) phenotype, and dwarfed plant size (Table 4.1) (Turner and Somerville, 1997; Taylor et al., 2003; Brown et al., 2005). Plants with disruption of *CESA* genes encoding either primary or secondary cell wall *CESA* proteins have defective cellulose production. These mutational studies demonstrate the essential roles CESAs have in proper cell and tissue development in *Arabidopsis*.

Table 4.1 Description of *A. thaliana cesa* mutants used in mutant phenotype rescue assays. Information from the Arabidopsis Information Resource (ABRC) as indicated.

Gene		Mutant allele	Insertion type (ABRC code)	Background	Phenotype information from ABRC
Primary Cell Wall CESAs	*CESA3* (At5g05170)	*cesa3-2*	T-DNA: SALK_137389	Columbia	Gametophytic lethal. 50 % pollen grains from the heterozygous plants are significantly deformed, without pollen tube formation. (Persson et al., 2007b)
	CESA6 (At5g64740)	*cesa6/ prc1-1*	Y752→ STOP: CS297	Columbia	Stunted hypocotyl and roots; incomplete cell wall; cellulose deficiency (Desnos et al., 1996; Fagard et al., 2000)
Secondary Cell Wall CESAs	*CESA7* (At5g17420)	*Cesa7/ irx3-4*	T-DNA: SALK_029940	Columbia	Irregular collapsed xylem; cellulose deficiency, dwarf plants (Brown et al., 2005)

Cellulose production starts with the trafficking of CESAs from the endoplasmic reticulum to the plasma membrane (for an overview, see Lampugnani, Khan, Somssich, & Persson, 2018; McFarlane, Doring, & Persson, 2014), and it requires a host of accessory proteins that support the cellulose synthase complex (CSC) machinery (for a review see Lampugnani et al., 2019).

However, it is unclear how closely our current paradigm of land plant cellulose production reflects the process of cellulose production in more distant lineages such as the red algae.

From a functional perspective, cellulose deposition occurs in similar subcellular domains between the land plant *Arabidopsis thaliana* and red algae in the genus *Calliarthron*. In land plants, cellulose is produced in both primary cell walls (PCW) as well as in secondary plant cell walls (SCW) that are deposited after the PCW. In recent years, distinct PCWs and SCWs have also been identified in some red algae such as *Calliarthron cheilosporioides* (Martone, 2007; Martone et al., 2009), the sister species to *Calliarthron tuberculosum* (Gabrielson et al., 2011). There are many developmental similarities between secondary cell walls in the joint (or genicular) tissues in *C. tuberculosum* and the interfascicular fibers, supportive tissue, and xylary, water-conducting, tissues in *A. thaliana*. The cell wall thickens after cell expansion to act in structural cellular support and lignin is found in both organisms (Martone, 2007; Martone et al., 2009; Meents et al., 2018; Martone et al., 2019). Given these similarities in cell wall specialization, it is possible that CESAs from the red algal and land plant lineages share comparable roles in both PCW and SCWs.

However, two major pieces of evidence suggest that cellulose production in Rhodophytes (red algae) has likely evolved divergently from the system of cellulose production in *A. thaliana*. First, the land plant CESA proteins likely evolved and became functionally specialized after divergence from the red algal CESAs lineage, as demonstrated in phylogenetic analysis (Ch 2). Second, many accessory proteins essential to cellulose production in the green plant lineage (encompassing the green algae to vascular plants) are not present in the red algae (Ch 2 and

Lampugnani et al., 2019). If these significant differences exist, to what extent are the components between the eukaryotic cellulose production systems interchangeable, if at all?

The identification and biochemical evidence for *CtCESA1* in *C. tuberculosum* presents an exciting opportunity to explore the functional compatibilities between CESAs from red algae and land plants. In this chapter I used *A. thaliana cesa* mutants, with mutations either in PCW or SCW *CESA* genes, to investigate a red algal CESA's functionality in a vascular plant. Previous phylogenetic analysis (Ch 2 Figure 2.2) has identified that CtCESA1 is equally similar/dissimilar to the PCW and SCW CESAs from *A. thaliana*. Here, I introduced *C. tuberculosum*'s CtCESA1 gene into the well-studied *A. thaliana* PCW mutant *procuste* (*cesa6*) to test if the algal gene can function in *A. thaliana*. If the introduction of *CtCESA1* can rescue the absence of CESA6 in the *cesa6/prc1-1* mutant, I predict that the *cesa6/prc1-1* plants transformed with *CtCESA1* would have increased hypocotyl lengths in dark-grown conditions. A second PCW CESA (*cesa3-2*), which is required for pollen formation (Persson et al., 2007b) was also tested. As the *cesa3-2* mutant is homozygous lethal, if the *CtCESA1* can rescue the absence of *CESA3*, I predict that homozygous *cesa3-2* mutants will be recovered in the presence of the *CtCESA1*. Finally, the secondary cell wall *cesa7* mutant, named irregular xylem (*irx3-4*), was used to test for functional rescue in the SCW context. If the *CtCESA1* gene could functionally rescue the absence of CESA7 in the *cesa7/irx3-4* mutant plants, I predict that the transformant plants would increase in size.

4.2 Results

4.2.1 *CtCESA1* is able to partially rescue the primary cell wall *cesa6* mutant

Mutations in the primary cell wall *cesa6/prc1-1* had reduced hypocotyl lengths relative to Col (Figure 4.1 and Figure 4.2) in dark-grown, etiolated, seedlings as previously reported (Fagard et al., 2000; MacKinnon et al., 2006). To test if the red algal CtCESA1 can functionally rescue the *Arabidopsis* primary cell wall CESA6, *CtCESA1* driven by the constitutively active ubiquitin 10 promoter (*UBQ10::CtCESA1*) was introduced into homozygous *cesa6/prc1-1* mutant plants by agrobacterium mediated transformation. The constitutively active *UBQ10* promoter approximates the overall expression patterns of PCW *CESA*s, which are expressed in many cell types and developmental stages (Norris et al., 1993; Grefen et al., 2010). 6 transformed lines (called the T1 generation), were recovered by gentamycin antibiotic selection of seeds from the originally transformed plant (T0 generation). Each transformed line represented an independent transgenic event. Genotyped T1 plants contained one copy of the *CtCESA1* gene, confirmed by PCR, and were grown to their T2 generation to test for the presence of homozygous mutants.

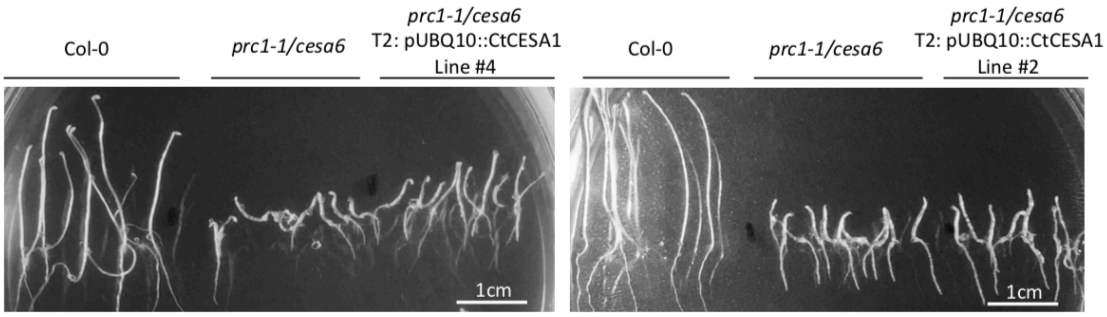

Figure 4.1 Partial rescue of the *cesa6/prc1-1* mutant phenotype with the *CtCESA1* gene. Shown are representative seedlings of Col and *cesa6/prc1-1* compared with two transformed lines in the *cesa6/prc1-1* background T2:*UBQ10::CtCESA1 line #4 and line #2.*

Welch's analysis of variance (ANOVA) indicated that there were significant differences in hypocotyl length between dark-grown seedlings of Col, *cesa6/prc1-1*, and T2:*UBQ10::CtCESA1* in the *cesa6/prc1-1* background, consistent within six independently transformed lines (Figure 4.2A) (line # 1 $F_{2,55} = 569$, $p = 2 \times 10^{-16}$; line # 2 $F_{2,37} = 377$, 2×10^{-16} ; line # 3 $F_{2,54} = 593$, $p = 2 \times 10^{-16}$; line # 4 $F_{2,15} = 184$, $p = 2 \times 10^{-11}$; line # 5 $F_{2,21} = 239$, $p = 2 \times 10^{-15}$; line # 8 $F_{2,5} = 31$, $p = 0.001$). Post hoc analysis with Games-Howell pairwise comparison tests indicated that hypocotyl lengths were different between all three genotypes Col, *cesa6/prc1-1*, and *cesa6/prc1-1* T2: *UBQ10::CtCESA1* for each transformant line tested. Welch's ANOVA was used as analysis assumptions of normality were met (Shapiro Wilks test $p > 0.1$), but unequal variances existed between the three genotypes (Levene's test $p < 0.01$). The complete experiment was replicated three times for line # 1, two times for line # 2, and four times for line # 3. The consistent approximately 0.1 cm difference in mean between the *cesa6/prc1-1* mutant and the

independently produced lines of T2: *UBQ10::CtCESA1* in the *cesa6/prc1-1* background demonstrated replicable partial phenotype rescue (Figure 4.2A).

Overall, the mean and standard deviation of the hypocotyl length was 1.66 ± 0.23 cm for wild type (Col, n = 99), 0.27 ± 0.05 cm for *cesa6/prc1-1* (n = 155), and 0.34 ± 0.06 cm for the combined lines of T2: *UBQ10::CtCESA1* in the *cesa6/prc1-1* background (n = 229) (Figure 4.2B). *cesa6/prc1-1* and *cesa6/prc1-1* T2: *UBQ10::CtCESA1* had overlapping distributions in their hypocotyl length measurements (Figure 4.2B). This was likely due to the segregating *CtCESA1* in the transformants, with some T2 offspring not carrying the *CtCESA1* insertion, resulting in differential levels of rescue.

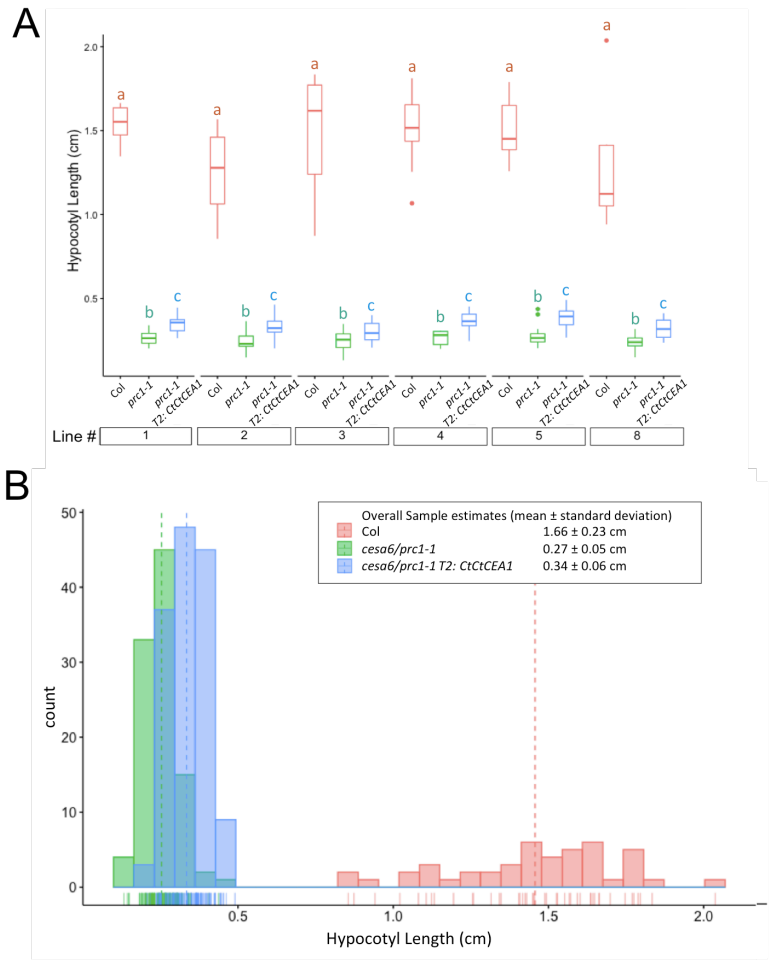

Figure 4.2 Etiolated hypocotyl lengths of Col, *cesa6/prc1-1*, and the T2: *UBQ10::CtCESA1* in the *cesa6/prc1-1* background are all statistically different. (A) Boxplot of hypocotyl measurements for each genotype separated by each independently transformed line. Statistical differences determined with Welch's ANOVA followed by post hoc analysis using Games-Howell multiple comparison tests ($p < 0.01$). (B) Histogram of all measured hypocotyl lengths from (A) colored by genotype with mean indicated by the dotted line. In total Col (n = 99), *cesa6/prc1-1* (n = 155), and T2: *UBQ10::CtCESA1* in the *cesa6/prc1-1* background (n = 229).

4.2.2 CtCESA1 did not rescue the primary cell wall *cesa3* mutant

The wild type (WT) copy of *CESA3* is required for proper pollen formation (Persson et al., 2007b). For this reason, only heterozygous plants with one mutant *cesa3* and one WT copy of *CESA3* or homozygous wild type plants with two copies of the *CESA3* can be produced. To test if the red algal CtCESA1 can functionally rescue the *Arabidopsis* primary cell wall CESA3, I introduced the *CtCESA1* gene, driven by the constitutively active *UBQ10* promoter (*UBQ10::CtCESA1*), into the heterozygous *cesa3-2* mutant background. The resulting *Arabidopsis* plants heterozygous for the *cesa3-2* gene with the introduced *UBQ10::CtCESA1* are called the T1 generation. If any homozygous *cesa3-2 mutants* were seen in subsequent generations from the transformed lines (i.e. T2, T3, etc. generations), this would suggest that its survival was due to the presence of the *CtCESA1* gene rescuing the lethality.

The first transformed generation (T1) of *UBQ10::CtCESA1* in the heterozygous *cesa3-2* background was visually indistinguishable from wild type (Col) (Figure 4.3). Presence of the *cesa3-2* T-DNA insertion (representative genotyping in Figure 4.4) and *CtCESA1* was verified with genotyping by PCR amplification. Total copy number of *CtCESA1* gene insertions was not measured. In the T1, I saw plants that were homozygous *CESA3* and plants that were heterozygous for the *cesa3-2* mutation, but no plants that were homozygous *cesa3-2* mutants, indicating an inability to rescue the mutant genotype with a single copy of *CtCESA1* (n=8 T1 lines). All survivors from the T1 that were genotyped carried the *UBQ10::CtCESA1* insertion.

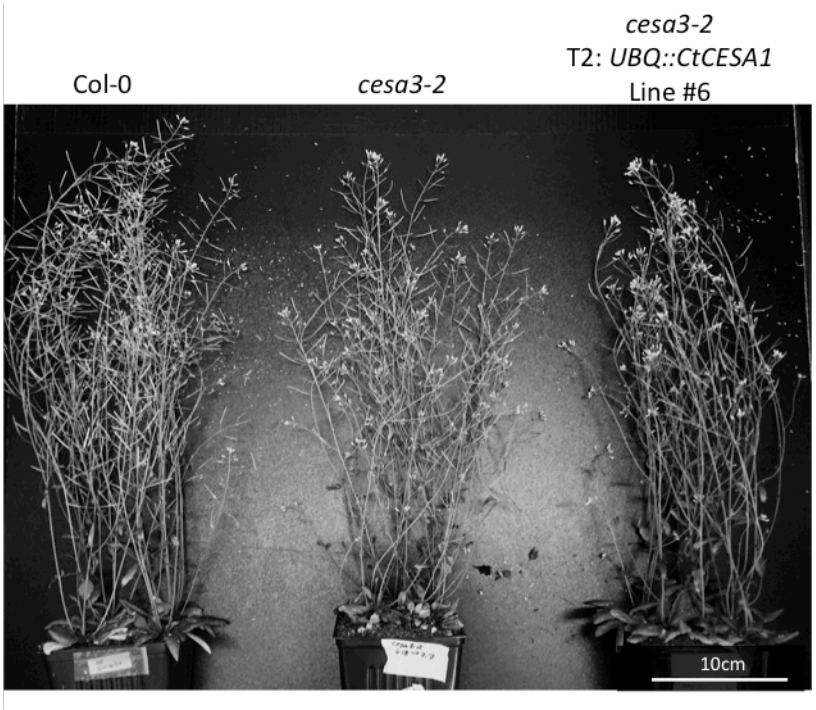

Figure 4.3 No visual phenotypic differences exist between wild type (Col), heterozygous *cesa3-2*, and the *CtCESA1* transformed lines in the heterozygous *cesa3-2* background (*cesa3-2* T2: *UBQ10::CtCESA1*). Plants shown were six weeks old.

The T1's were selfed to generate the T2: *UBQ10::CtCESA1* in the *cesa3-2* background and the segregating *cesa3-2* background was genotyped by PCR amplification (Figure 4.4). The no-template control showed no banding indicating that the amplification seen in other lanes was not due to contamination (lane blank Figure 4.4). As a control, the heterozygous *cesa3-2* was also selfed. It generated the predicted offspring population segregating with heterozygous *cesa3-2* and homozygous *CESA3* but no homozygous *cesa3-2* (Table 4.2) confirming the previously identified pollen lethal *cesa3* mutant phenotype (Persson et al., 2007b). Upon genotyping three independently transformed lines of T2: *UBQ10::CtCESA1* (line #1, #2, and #6) for the segregating *cesa3-2* background, no homozygous *cesa3-2* were recovered (Table 4.2). If any

homozygous *cesa3-2* mutants had been recovered with the introduction of *UBQ10::CtCESA1*, this would suggest that its survival was due to the presence of the *CtCESA1* gene rescuing the *cesa3-2* lethality. However, given that genotyping of the T1 or T2 generation of *UBQ10::CtCESA1* introduced into the *cesa3-2* background failed to find any *cesa3-2* homozygotes, this suggests that the *CtCESA1* gene cannot rescue the lethal phenotype of the *cesa3-2* homozygous mutants.

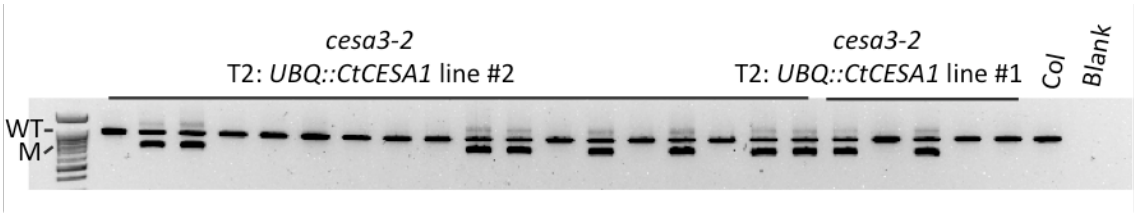

Figure 4.4 Representative genotyping for the presence of the T-DNA insertion in the segregating *cesa3-2* population in the second generation (T2) *UBQ10::CtCESA1* background. Two independently transformed lines, #2 and #1, are shown. The presence of the T-DNA insertion in the *CESA3* gene is indicated by the presence of a lower band and an upper band indicates the presence of a wild type copy of a *CESA3* gene. Template DNA is indicated above with each lane representing an individual plant (col – wild type control, blank – no template control). A 100 bp ladder is shown.

Table 4.2 Genotypes of *cesa3-2* and independently transformed lines T2: *UBQ10::CtCESA1* in a *cesa3-2* background. No homozygous *cesa3-2* seedlings were recovered.

Genotype	Homozygous Wild Type	Heterozygous *cesa3-2* insertion	Homozygous *cesa3-2* insertion
cesa3-2 (control)	12	15	0
cesa3-2 T2: *UBQ10::CtCESA1 line #1*	9	19	0
cesa3-2 T2: *UBQ10::CtCESA1 line #2*	14	13	0
cesa3-2 T2: *UBQ10::CtCESA1* line #6	15	19	0

4.2.3 CtCESA1 did not rescue the secondary cell wall *cesa7* mutant

Mutations in the secondary cell wall *cesa7/irx3-4* plants were dwarf compared to wild type (Col) (Figure 4.5), as previously observed (Brown et al., 2005). To test if the red algal CtCESA1 can functionally rescue the *Arabidopsis* secondary cell wall CESA7, *CtCESA1* driven by the secondary cell wall specific CESA7 promoter (*proCESA7::CtCESA1*) was introduced into the homozygous *cesa7/irx3-4* mutant. The CESA7 promoter was used so that the transgenic *CtCESA1* gene would have a temporal-spatial expression pattern similar to the native *CESA7* gene (Smith et al., 2013).

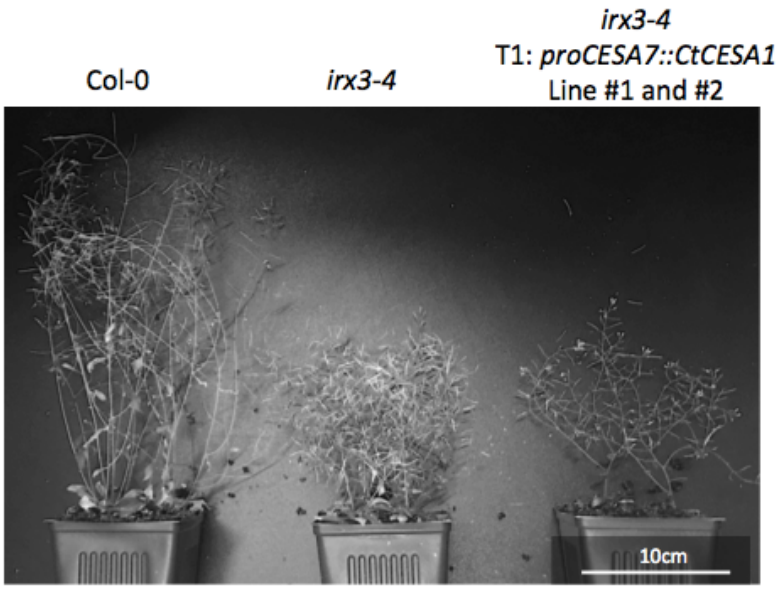

Figure 4.5 *cesa7/irx3-4* plants are dwarf compared to wild type (Col). *The cesa7/irx3-4* and the *CtCESA1* transformed lines in the *cesa7/irx3-4* background (*cesa7/irx3-4* T1: *proCESA7::CtCESA1*) are similar in plant size. Plants shown were nine weeks old.

To assess if *CtCESA1* could rescue the mutant dwarf phenotype, stem lengths of mature plants with the genotype Col, *cesa7/irx3-4*, and the first generation of *CtCESA1* transformed lines in the *cesa7/irx3-4* background were measured and compared (Figure 4.6). The mean and standard deviation of stem lengths was 36.6 ± 3.7 cm for wild type (Col, n = 8), 14.7 ± 0.95 cm for *cesa7/irx3-4* (n = 4) and 15.0 ± 3.64 cm for T1: *proCESA7::CtCESA1* in the *cesa7/irx3-4* background (n = 8) (Figure 4.5). The sample size for the T1: *proCESA7::CtCESA1* in the *cesa7/irx3-4* background consisted of multiple T1 plants combined. The data was normal (Shapiro-Wilk test $p > 0.1$) and had shared equal variance (Levene's Test $p > 0.1$).

Analysis of variance (ANOVA) indicated that there were significant differences between the three groups ($F_{2,17} = 100.7$, $p = 3.8 \times 10^{-10}$). Post hoc analysis with Tukey's pairwise comparison tests indicated that the stem lengths of the *cesa7/irx3-4* were not statistically significant from T1: *proCESA7::CtCESA1* in the *cesa7/irx3-4* background, but both were statistically different from those in Col (Figure 4.6). This suggests that the introduction of *CtCESA1* was unable to rescue the stem length phenotype of the *cesa7* mutant.

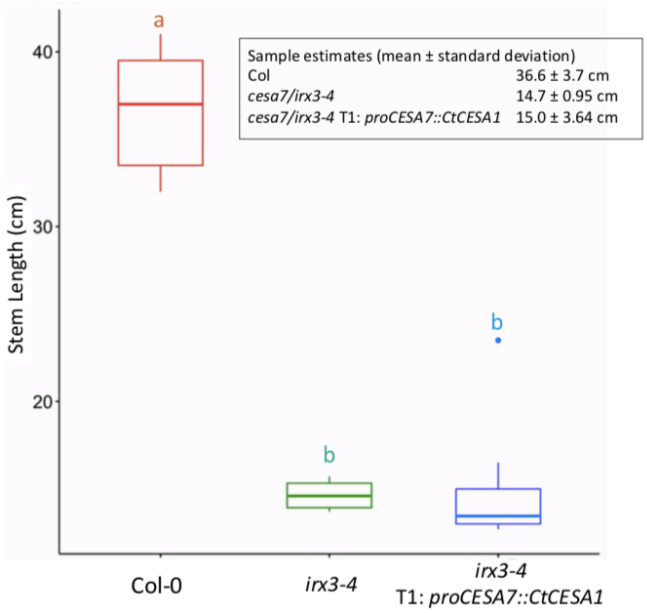

Figure 4.6 Wild type (Col) stem lengths are longer than *cesa7/irx3-4*, and the *CtCESA1* transformed lines in the *cesa7/irx3-4* background (*cesa7/irx3-4* T1: *proCESA7::CtCESA1*). Statistical differences determined with ANOVA followed by post hoc analysis using Tukey multiple comparison tests ($p < 0.001$) Col (n = 8), *cesa7/irx3-4* (n = 4), and the first generation of *CtCESA1* transformed lines in the *cesa7/irx3-4* background (n = 8).

4.3 Discussion

In this work, I used three well-characterized *Arabidopsis thaliana* mutants to test if the CESA from the red alga *Calliarthron tuberculosum* could function in a land plant and rescue the cellulose deficient mutant phenotypes. The results show that CtCESA1 could partially rescue the primary cell wall *cesa6* mutant, but not primary cell wall *cesa3* or secondary cell wall *cesa7*. It is not surprising that the algal gene could not rescue both primary and secondary cell wall mutants, as my phylogenetic analysis (chapter 2 of this thesis) strongly supports that the specialization of the land plant primary and secondary cell wall CESAs occurred independently from the evolution of red algal CESAs.

It is surprising that the *CtCESA1* was able to partially rescue the *cesa6/prc1-1* mutant, given that even *CESAs* within *A. thaliana*, which share high sequence conservation (Sethaphong et al., 2013), are, for the most part, functionally non-interchangeable in mutant studies. Mutant rescue studies of *A. thaliana* secondary cell wall *cesa* mutants show that only their wild type counterpart can functionally rescue their mutant phenotypes (Kumar et al., 2017). For example, *cesa7/irx3* mutant phenotypes can be rescued by a wild type *CESA7* but not a *CESA4* or *CESA8* sequence. Moreover, the *A. thaliana* primary cell wall CESA3 cannot rescue the *cesa1* mutant phenotype (Burn et al., 2002). In gene swap assays that explore the interchangeability of *A. thaliana* PCW and SCW CESAs, the SCW CESA7 was able to partially rescue a temperature sensitive *cesa3* but not a *cesa1* or *cesa6* mutant and the primary cell wall CESA1 was able to partially rescue a *cesa8,* but not a *cesa4* or *cesa7* mutant (Carroll et al., 2012). Clearly, there is little functional compatibility between class specific CESAs from *A. thaliana*. Because *A. thaliana* CESAs show

high functional specificity, it becomes clear that partial and not full rescue of *cesa* mutant phenotypes by *CtCESA1* would be expected. Given that red algal CtCESA1 and the *A. thaliana CESA3, CESA6,* and *CESA7* are highly dissimilar, it raises questions as to how this red algal CESA was able to partially rescue one, *cesa6/prc1-1*, mutant but not the other, *cesa7* or *cesa3*, which I discuss below.

4.3.1 Mechanisms surrounding cellulose production are likely functionally distinct between land plants and red algae

The introduction of *CtCESA1*, driven by the appropriate promoter for comparable expression with their *A. thaliana* host, was unable to rescue the primary cell wall *cesa3* and secondary cell wall *cesa7* mutant phenotypes measured, and only moderately rescued the primary cell wall *cesa6* mutants. These results show that although CtCESA1 is generally unable to compensate for the loss of the native *CESA* gene in *A. thaliana* and that their functional compatibility is therefore strongly limited. This limited compatibility could be due to several different factors such as recycling of CtCESA1 back into the cell, differences in secondary modifications, or low activity of CtCESAs unable to compensate for the loss of cellulose in the *A. thaliana cesa* mutants.

Perhaps a certain threshold of increased cellulose production is required to alleviate the *A. thaliana cesa* mutant phenotypes. From activity assays in chapter 3, the CtCESA1 enzymatic activity is substantially lower than the land plant secondary cell wall specific CESA, PttCESA8. If the measured levels of *in vitro* activity (Ch 3 Figure 3.8) reflect *in vivo* activity, then the low rate of cellulose production by CtCESA1 could be insufficient to compensate for the loss of

cellulose in the *A. thaliana* mutant *cesa3 or cesa7*. This may also explain why there was only a moderate increase in hypocotyl length (~ 7 % recovery; Figure 4.2) in the *cesa6* mutants after introducing CtCESA1 as opposed to full functional rescue seen in *cesa6* mutants with their wild type *CESA6* gene introduced (Park and Ding, 2020).

There are multiple alternative explanations for *CtCESA1*'s inability to rescue the *A. thaliana cesa* mutants, although these are more speculative. Post-translational modifications of *A. thaliana* CESAs are known to occur, such as protein phosphorylation (Chen et al., 2010; Jones et al., 2016) and acylation (Kumar et al., 2016), and affect CESA behaviour at, or delivery to the plasma membrane. However, the presence of CESA modifications is entirely unknown for red algal CESAs. For this reason, it is difficult to predict whether the inability to rescue the *A. thaliana cesa* mutants is due to the impaired enzymatic function of CtCESA1 caused by the absence or improper presence of post-translation modifications on the CtCESA1 protein. The role that CESA accessory proteins play in the observed results is also unclear. In chapter 2, sequence queries showed multiple CESA accessory proteins that facilitate proper cellulose synthesis in *A. thaliana* are largely absent in *C. tuberculosum*'s transcriptome (Figure 1.4). The accessory proteins present in *C. tuberculosum*'s transcriptome had large sequence differences, which likely reflect enzymatic differences, from their land plant counterpart. The absence of identified protein homologs and large sequence differences may suggest improper protein-protein interactions between the land plant accessory proteins and the CtCESA1 may have impaired CtCESA1 function at the *A. thaliana* plasma membrane. Lastly, there may be a low abundance of CtCESA1 at the plasma membrane. Given that CtCESA1 is trafficked to the plasma membrane, it is unclear if they remain there. The number of functional CtCESA1 could

have been reduced if they were consistently recycled back into the cell, though this will require further experimental *in planta* visualization to confirm.

As the CtCESA1 is mostly unable to rescue phenotypic deficiencies of *A. thaliana* PCW and SCW *cesa* mutants, the hypothesis that eukaryotic CESAs from evolutionarily distinct lineages have major functional incompatibilities is supported. Localization experiments of a fluorescently tagged CtCESA1 protein in the *A. thaliana cesa* background could demonstrate protein expression of CtCESA1 and help clarify some outstanding questions surrounding lack of functional rescue.

4.3.2 CESAs from distant lineages likely share deeply conserved trafficking routes to the plasma membrane

The CtCESA1 was able to partially rescue the primary cell wall *cesa6* phenotype. Because CESAs generally function at the plasma membrane, CtCESA1's functional rescue of the *cesa6* mutant, regardless of how moderate, suggests that there are likely deeply conserved CESA trafficking mechanisms to the plasma membrane. Given the partial rescue of the *cesa6* background, we must assume that the CtCESA1 can arrive at the plasma membrane through the trafficking pathway present in land plants. However, the extent of the trafficking beyond delivery to the plasma membrane is unknown. Further observational evidence would confirm their presence at the plasma membrane in *A. thaliana*; this could be achieved by fluorescently tagging the protein (Xue et al. ongoing).

Prior to insertion at the plasma membrane, land plant CESAs interact to form trimmers (Purushotham et al., 2020), and these trimmers associate into rosette shaped cellulose synthase complexes (CSCs) (Nixon et al., 2016). From freeze fracture electron microscopy images, red algae are also thought to form larger and linear CSC complexes (Tsekos and Reiss, 1992; Tsekos, 1999). It remains unknown if the CtCESA1 self assembles into a CSC or if the CtCESA1 is able to integrate into the *A. thaliana* CSC complex. The AtSTELLO proteins in *A. thaliana* may facilitate CtCESA1 CSC formation. In *A. thaliana*, CSC complex formation is thought to occur in the Golgi, facilitated by STELLO proteins (Zhang et al., 2016). Given that a candidate STELLO was identified in *C. tuberculosum*'s transcriptome dataset (Ch 2), perhaps the land plant STELLO could act analogously.

Structural evidence and biochemical evidence from PttCESA8 suggest that their CESAs likely form trimers via their plant conserved region (P-CR), 7th transmembrane domain (Purushotham et al., 2020), and zinc-finger (Znf) domain at their N terminus (Purushotham et al., 2016). CtCESA1 does not have a P-CR or N-terminal Znf domain but does have an analogous 7th transmembrane domain. Given this, it is unclear whether this 7th transmembrane domain alone would be sufficient to facilitate integration into the *A. thaliana* CESA trimer. With the fluorescently tagged CtCESA1, we can begin to elucidate its CSC formation and associations. Land plant CSCs form small puncta (~ 25 nm) when visualized at the plasma membrane (Paredez et al., 2006; Watanabe et al., 2018) whereas red algal CSCs are thought to be 4 – 5 x larger (~ 100 nm – 125 nm) (Tsekos, 1999). Upon visualization at the plasma membrane, if CtCESA1 is capable of self-associating into a CSC complex we would expect to see larger

fluorescent structures whereas if it were associating into CSCs composed of *Arabidopsis* CESAs we would likely see smaller fluorescent puncta at the plasma membrane.

4.4 Methods

4.4.1 Sample collection and cDNA production

Calliarthron samples were collected from Botanical Beach, Port Renfrew, BC (location 48°31'31.1"N 124°26'26.4"W and permit number 109636) situated on Pacheedhat First Nations lands. Samples were cleaned with 70 % ethanol and then kept at -80 °C in a dry shipper in order to ensure RNA stability. Samples were ground in a chilled mortar and pestle with liquid nitrogen.

RNA was extracted using the cetyltrimethylammonium bromide (CTAB) method (Barbier et al., 2019). Extracted RNA was treated with DNAse (DNase I, RNase-free) and cDNA was subsequently produced using reverse transcriptase (SuperScript™ III Reverse Transcriptase) with anchored oligo dT 22.

4.4.2 CtCESA1 nucleotide sequence

The CtCESA1 nucleic sequence was determined by sequencing PCR amplified samples from field collected *Calliarthron* samples. The nucleotide transcriptome data that the protein sequence has been derived can be found at European Nucleotide Archive (project number PRJEB39919).

4.4.3 *Arabidopsis thaliana* plant growth conditions

Seeds were sterilized with chlorine gas for 2 h then sown on plates containing germination selection media (GM) (1 × Murashige-Skoog (MS), 1 % Sucrose, 1 x Gamborg's Vitamin mix, 0.05 % MES, 0.8 % agar at pH 5.8) and antibiotic selection when appropriate (described below). Seeds were cold treated in the dark at 4 °C for 4 days, then grown at 21 °C vertically until seedlings with first true leaves grew. Seedlings were then transplanted onto soil and grown in long day light cycles with 18 h light and 6 h dark at 21 °C and 18 °C respectively and 70 % humidity.

4.4.4 Construct design

The putative *C. tuberculosum* CESA1 must have similar spatiotemporal expression relative to the *A. thaliana* mutant they are transformed into (PCW mutant vs. SCW mutant) to properly assess functional rescue. *A. thaliana* PCW *CESA*s are expressed during most stages of cellular development and in most cell types. As such primary cell wall mutants will be transformed with the *CtCESA* driven by a ubiquitin promoter (*proUBQ10::CtCESA1*) conferring constitutive expression allowing expression to always be turned on (Christensen & Quail, 1996; Grefen et al., 2010). The ubiquitin10 promoter was used over another popular constitutive promoter cauliflower mosaic virus 35S as 35S promoters can cause silencing of the sequence it drives (Grefen et al., 2010). However, the SCW *CESA*s are only expressed at specific developmental stages. Because CESA7 is a SCW specific CESA, its promoter will allow the *CtCESA*

(*proCESA7::CtCESA1*) to be expressed in a secondary cell wall specific spatio-temporal nature (Smith et al., 2013).

The gateway binary vector system was used to create the promoter-*CtCESA1* constructs. Gateway cloning is based on site-specific recombination sites used to combine sequences of interest in a direction-specific manner (Reece-Hoyes & Walhout, 2018). This involves the production of an entry vector with the sequence of interest. This entry vector is then used to shuttle its contents into later destination vectors that contain the final combined sequence (i.e. the promoter, sequence of interest, and additional modifications).

4.4.4.1 CtCESA1-pDONR/Zeo

To produce the entry vector CtCESA1-pDONR/Zeo, CtCESA1 was amplified from C. tuberculosum cDNA using primer P1F and P1R (addition of attb adaptors Table 4.3) and cloned into pDONR™/Zeo (Invitrogen cat. No 12535-035) in a BP gateway reaction. Plasmids were transformed into chemically competent Top10 E. coli cells and grown on LB plates under Zeocin selection (50 μg/mL). Colonies were cultured in LB under Zeocin selection (50 μg/mL) and subsequently had their plasmids harvested. Plasmids were then verified by sequencing using the primers found in Table 4.3.

4.4.4.2 proUBQ10::CtCESA1

Sequence verified CtCESA1-pDONR/Zeo was cloned into the proUBQ10::ccdb destination vector (Grefen et al., 2010) using a LR reaction to generate the proUBQ10::CtCESA1 construct. Plasmids were transformed into chemically competent Top10 E. coli cells and grown on LB plates under spectinomycin selection (100 μg/mL). Colonies were cultured in LB under spectinomycin selection (100 μg/mL) and subsequently had their plasmids harvested. Plasmids were then verified by sequencing using the primers found in Table 4.3.

4.4.4.3 proCESA7::CtCESA1

A multisite gateway protocol (Cheo et al., 2004) was used to generate the final proCESA7::CtCESA1 construct (for a diagrammatic scheme of this process see Figure 4.7). This first involved amplifying the CESA7 promoter sequence (proCESA7) with attb adapter sites from the proCESA7::mNG:CESA7 construct (Watanabe et al. unpublished) using primers P2F and P2R (Table 4.3). The CESA7 promoter sequence is originally from the genomic encoded sequence 1127 bp region directly upstream the CESA7-encoding gene (Smith et al., 2013). The adapter added proCESA7 sequence was inserted into pDONR P4-P1R in a BP reaction (Figure 4.7) to generate the proCESA7-pDONR/P4-P1R construct. The plasmid was transformed into Top10 E. coli cells and grown on LB plates under gentamycin selection (50 μg/mL). Colonies were cultured in LB under gentamycin selection (50 μg/mL) and subsequently had their plasmids harvested. Plasmids were then verified by sequencing using the primers found in Table 4.3.

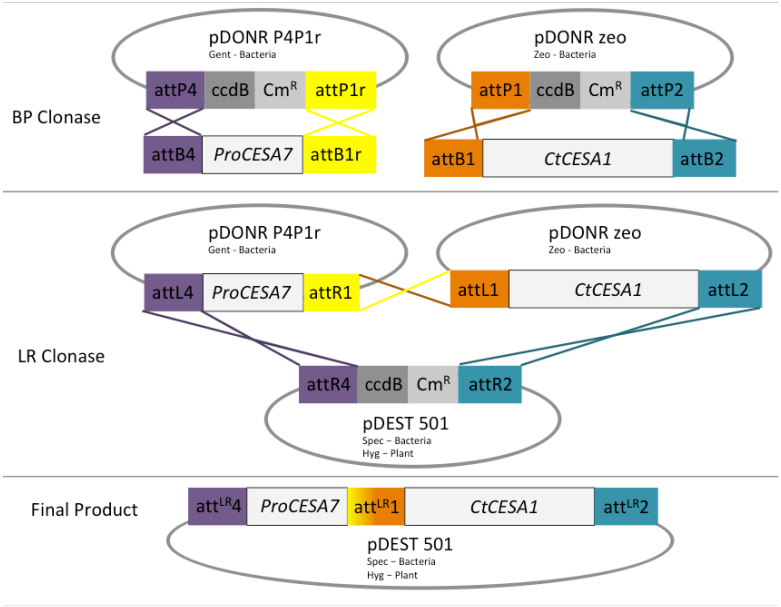

Figure 4.7 Schematic of *proCESA7::CtCESA1* construct generation using two fragment gateway.

The proCESA7-pDONR/P4-P1R, CtCESA1-pDONR/Zeo, and pDEST501 were incubated together in a LR reaction (Figure 4.7) to generate the final pCESA7::CtCESA1 construct. The plasmid was transformed into Top10 E. coli cells and grown on LB plates under spectinomycin selection (100 µg/mL). Colonies were cultured in LB under spectinomycin selection (100 µg/mL) and subsequently had their plasmids harvested. Plasmids were then verified by sequencing using the primers found in Table 4.3.

4.4.5 Generation of plant lines

cesa mutant seeds were obtained from the Arabidopsis Research Center (Table 4.1). The *pCESA7::CtCESA1* and *proUBQ10::CtCESA1* were transformed into electrocompetent *Agrobacterium tumefaciens* (strain GUV1301) separately. Correct colonies were selected on LB containing spectinomycin (100 µg/mL) and gentamycin (50 µg/mL) grown at 28 °C for 3 days. *pCESA7::CtCESA1* was transformed into SCW *cesa* and *proUBQ10::CtCESA1* was transformed into PCW *cesa* mutant backgrounds using the floral dip method (Clough and Bent, 1998). T1 seeds of *pCESA7::CtCESA1* in the SCW *cesa* backgrounds were selected on GM containing hygromycin (25 µg/mL). T1 seeds of *UBQ10::CtCESA1* in the PCW *cesa* backgrounds were selected on GM containing BASTA (25 µg/mL).

4.4.6 *cesa3-2* PCR genotyping assay

T1 plants of proUBQ10::CtCESA1 in a cesa3-2 background were grown up on soil. DNA was extracted from leaf tissue using the Edwards solution extraction method (Edwards et al., 1991). Samples were genotyped for the presence of a transfer DNA (T-DNA) insertion by PCR amplification with the primers Cesa3-2 F, Cesa3-2 R, and SALK lbb1.3 (Table 4.3). Samples were also genotyped for the presence of CtCESA1 by PCR with the primers SEQP1F and SEQP4R (Table 4.3). Plants heterozygous for cesa3-2 were kept as separate proUBQ10::CtCESA1 insertion lines and their seeds were harvested as T2 lines. T2 plants of proUBQ10::CtCESA1 in a cesa3-2 background were genotyped for the T-DNA insertion to look for the rescue of cesa mutant phenotypes.

4.4.7 *cesa6/prc1-1* etiolated hypocotyl growth assay

T2 plants of *proUBQ10::CtCESA1* in a *cesa6/prc1-1* background were assessed for hypocotyl length in a dark grown (etiolated) setting as previously described by Desnos et al. (1996). Arabidopsis wild type (Col), *cesa6/prc1-1*, and T2: *proUBQ10::CtCESA1* in a *cesa6/prc1-1* background were plated on the same GM agar plate in replicate. Seedlings were cold treated for 4 days then exposed to white light for 2 h to synchronize germination. Plates were then wrapped in two to three layers of aluminum foil and grown vertically at 21 °C for 5 days after which they were imaged. A ruler was included in each image to provide a scale. Hypocotyl lengths were measured using imageJ through the FIJI program (Schindelin et al., 2012). Hypocotyls were measured from the shoot apical meristem to the point between the hypocotyl and root hairs. Hypocotyls that were not lying flat against the growth media (indicated by tapering width) would not provide accurate total lengths and were not measured for this reason.

4.4.8 *cesa7/irx3-4* stem length measurement assay

Col, *cesa7/irx3-4*, and T1 plants of the *proCESA7::CtCESA1* in the *cesa7/irx3-4* background were grown to maturity. Stem lengths were measured from the base of the rosette to the bottom of the apical inflorescence for each plant.

Table 4.3 List of primers used in cloning *CtCESA1* into gateway vectors.

Primer	Sequence (5' -> 3')	Purpose
P1 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTC ATGGCGAACATTGGCCCC	Primer set 1 used to amplify CtCESA1 with attb1 attb2 sites for gateway cloning into pDONR™/Zeo
P1 R	GGGGACCACTTTGTACAAGAAAGCTGGGTT TTAAAACCTGCGACCCTCTACCG	
P2 F	GGGGACAACTTTgtatagaaaagttgacttacgaattgGGCTCCAACG	Primer set 2 used to amplify the promoter of *CESA7* with attb1 attb2 sites for gateway cloning into pDONR^{TM} P4-P1R
P2 R	GGGGACTGCTTTTTTGTACaaacttggGCGGTGATCAATGA GAGACG	
M13F	CCCAGTCACGACGTTGTAAAACG	List of primers used in sequencing CtCESA1 constructs
M13R	AGCGGATAACAATTTCACACAGG	
SEQP1F	GCGGTTGGACTACCCATCTC	
SEQP2R	GCGCTCAACACTTTTCGGAG	
SEQP3F	GCGCCGGATAACATCTACGA	
SEQP4R	GCCTCGCAACACATCAATGG	

Primer	Sequence (5' -> 3')	Purpose
Cesa3-2 F	TAGATGGGATTCAAGGACCTG	Primers used in genotyping *cesa3-2* plant lines
Cesa3-2 R	ACGGCTGGCAATGTACAATAC	
SALK lbb1.3	ATTTTGCCGATTTCGGAAC	
Irx3-4 F	AGAGAAGCTTAAGGAAACCGC	Primers used in genotyping *cesa7/irx3-4* plant lines
Irx3-4 R	GAACAACACAAGAGCAGAGGG	
SALK lbb1.3	ATTTTGCCGATTTCGGAAC	

Chapter 5: Summary and future directions

In this thesis, I sought to functionally characterize a CESA protein from the calcifying red algae *Calliarthron tuberculosum* and test differences in the overall cellulose synthesizing process between *C. tuberculosum*'s CESAs and the CESAs from land plants using *Arabidopsis thaliana* as a model.

The data in this thesis suggest that, like the well-described systems in land plants, CESA proteins mediate cellulose production in the calcifying red alga *C. tuberculosum*. Here, three candidate CESAs were recovered from *C. tuberculosum*'s transcriptome, *CtCESA1*, *CtCESA2* and *CtCESA3*. The *in vitro* expressed and purified CtCESA1, which is most closely related to other red algal CESAs and oomycete CESAs, had low levels of glucan synthase activity (Ch 3). This suggests that the other red algal CESAs, from *Pyropia yezoensis* and *Chondrus chrispus* for example, and the oomycete CESAs that CtCESA1 shared a close relationship with, likely have CESA activity. Moreover, the close relationship identified between the pathogenic oomycetes and red algal CESAs, distinct from CESAs of other lineages, supports the hypothesis that the oomycete CESAs identified may have been acquired from an ancestral red algal CESA. As the identified oomycetes require cellulose for their pathogenicity (Grenville-Briggs et al., 2008), insight into red algal CESAs may lead to a greater understanding of the mechanisms involved in oomycete infection as well as host responses.

This thesis also suggests that while there are likely some deeply conserved mechanisms for CESA trafficking, those conserved mechanisms are overshadowed by lineage specific

differences between the land plant and red algal cellulose machinery (CESAs) and their overall system of cellulose production. At the core enzymatic level, analysis of CESA evolution showed that CESAs sequences from land plants and red algae are largely distinct (Ch 2). Moreover, there were large functional differences in the rate of activity between the *Populus tremula x tremuloides* PttCESA8 and *C. tuberculosum* CtCESA1 (Ch 3). The overall systems surrounding CESA trafficking, microtubule interactions, and crystallization likely also have large dissimilarities as homologs were identified for some but not all of the cellulose synthase accessory proteins in *C. tuberculosum* (Ch 2). These differences were highlighted by studies *in planta* where the introduction of CtCESA1 was unable to functionally rescue the primary cell wall *cesa3-2* mutant or the secondary cell wall *cesa7/irx3-4* mutant. However, CtCESA1's ability to partially complement the primary cell wall *cesa6/procuste1-1* mutant did suggest that there are deeply conserved pathways in trafficking, allowing the CtCESA1 to localize correctly to function. Our understanding of cellulose synthesis in land plants should be applied with caution and limitation when looking at evolutionarily distant CESAs such as those in red algae, brown algae, oomycetes, tunicates, and otherwise. These proposed lineage specific differences further highlight the need for studies in these groups to properly assess their mechanisms of cellulose synthesis.

The findings from this thesis bring up many remaining questions. Here I identify some remaining hypothesis and propose experiments to address them.

5.1 Does *C. tuberculosum*'s putative CBM48 domain bind starch?

The red algal *CESA* sequences have unique features not found in other *CESA*s. Elucidating functional differences between CESAs from different species is relevant for engineering CESA proteins and their economically valuable cellulose products. Red algal *CESA*s encode a putative carbohydrate domain 48 (put-CBM48) at their N-terminal domain. This put-CBM48 is predicted to extend into the cell and bind starch (Ch 2). Florideophyceae red algae, to which *C. tuberculosum* belongs, possess naked granules of a unique starch called Floridean starch. Unlike land plants that store their starch in the chloroplast, red algal starch floats freely in the cell (Viola et al., 2001).

This Floridean starch is structurally different from land plant starch resulting in differences in digestibility (Meeuse et al., 1960). Land plant starch is composed of both amylose (linear α-1,4-linked glucose) and amylopectin (α-1,4-linked glucose with branched α-1,4-linked glucose) at varying ratios (Tester et al., 2004) while red algal Floridean starch lacks amylose and is composed of amylopectin-like molecules (also α-1,4-linked glucose with branched α-1,6-linked glucose) (Yu et al., 2002). One hypothesis is that this put-CBM48 facilitates binding to intracellular stores of Floridian starch specifically as opposed to plant like starches.

To test this hypothesis, one could isolate the *put-CBM48* domain found at the N-terminus of *Calliarthron*'s *CtCESA1* sequence and test if it can bind starch. The region containing the *put-CBM48* domain could be isolated, heterologously expressed in a bacterial system such as *E. coli*, and its protein purified. The binding affinity of this purified put-CBM48 protein could then be

tested against various glucans including a land plant starch, such as potato starch, pure amylose, pure amylopectin, and Floridean starch. Control glucans that can also be tested are callose and cellulose, which are glucan chains with different interlinking bonds than starch molecules. Protein pull-down assays can be performed to test the starch binding capacity of the put-CBM48. In this assay, the put-CBM48 protein could be incubated with one of the carbohydrates mentioned above, and then separated by centrifugation. If the put-CBM48 could bind the carbohydrate, it should be pulled down with and found in the carbohydrate fractionation. Alternatively, protein shift assays in blue native polyacrylamide gel electrophoresis (PAGE) can also be conducted to visualize if the put-CBM48 protein can bind starch. In this experiment one of the aforementioned carbohydrates would be solubilized into a non-denaturing PAGE gel. As the put-CBM48 is loaded into and runs through the gel, the protein's movement should be retarded if binding the carbohydrate in the gel matrix.

5.2 Elucidating the untested roles of the *CtCESA2* and *CtCESA3* candidates

In this study, I identified two additional *CESA* sequence candidates in *C. tuberculosum*'s transcriptome, *CtCESA2* and *CtCESA3*. As these sequences were not closely related to *CtCESA1*, it is unclear if they have glucan synthase activity. One hypothesis is that *CtCESA2* and *CtCESA3* also function in cellulose synthesis but at distinct developmental stages. From studies in *A. thaliana* we know that their *CESA*s are transcriptionally regulated to operate either during primary cell wall synthesis or secondary cell wall synthesis, but not both (Brown et al., 2005; Watanabe et al., 2018). As *C. tuberculosum* has both PCWs and SCWs (Martone, 2007; Martone et al., 2009), perhaps it also has multiple cellulose synthases that are transcriptionally regulated

to function during either PCW or SCW synthesis. To test this hypothesis, one could compare the transcription levels of *CtCESA1*, *2*, and *3* in both young (containing only primary cell walls) and older (containing both primary and secondary cell walls) *C. tuberculosum* tissue using quantitative reverse transcription PCR (RT-qPCR). If there were specificities in developmental synthesis, we would expect expression to occur differentially between the two tissue types. Alternatively, they could be *CESA-like* sequences in *C. tuberculosum* whose proteins produce non-cellulose glucans, though this would require more in-depth functional protein characterization.

5.3 Localization and behavior of *Calliarthron tuberculosum*'s CESAs protein in *Arabidopsis thaliana cesa* mutants

When expressed in *A. thaliana*, *CtCESA1* partially rescued the primary cell wall *Atcesa6* mutant but was unable to rescue the primary cell wall *Atcesa3* or secondary cell wall *Atcesa7* mutants. Given that the CESA could complement the *Atcesa6* mutant functionally, we assume that the CtCESA1 protein was trafficked to its functional location at the plasma membrane. If this were the case, we would assume that CtCESA1 protein was also trafficked to the plasma membrane in these other *Atcesa7* and *Atcesa3* mutants but, for some other reason, was not able to complement the lack of native cellulose. Alternatively, CtCESA1 may have been non-functional due to intracellular mis-localization in the *Atcesa7* and *Atcesa3* mutants. To establish the localization of CtCESA1 in these *A. thaliana cesa* mutants, CtCESA1 can be fused to a fluorescent tag and introduced to these *A. thaliana cesa* mutants in the same *Agrobacterium tumefaciens*-mediated manner described in chapter 4. Spinning disk fluorescent confocal microscopy could then be

utilized to view CtCESA1 protein localizations in the *A. thaliana cesa* mutants. The protocols for this imaging have already been optimized (Watanabe et al., 2015; Watanabe et al., 2018). One could then visualize the CtCESA1 localizations along the expected Golgi to vesicular transport to plasma membrane trafficking route in the *A. thaliana cesa* mutants and confirm either its correct localization at the plasma membrane or intracellular mislocalization.

Once this fluorescently tagged CtCESA1 is introduced into these *A. thaliana cesa* mutants, we could also visualize their behavior at the plasma membrane. CESAs in cellulose synthesizing complexes (CSC) migrate along the plasma membrane continuously and linearly with a velocity of 265 nm/min and 231 nm/min for secondary and primary cell wall CESAs respectively in *A. thaliana* (Watanabe et al., 2015). To produce these measurements, fluorescently tagged CESA proteins were visualized under fluorescent confocal microscopy and their movements tracked (using kymograph analysis). These same methods could be applied to the fluorescently tagged CtCESA1 in the *A. thaliana cesa* mutants to measure their velocity and monitor their movements at the plasma membrane.

Bibliography

Amikam D, Galperin MY (2006) PilZ domain is part of the bacterial c-di-GMP binding protein. Bioinformatics **22**: 3–6

Anderson CT, Kieber JJ (2020) Dynamic construction, perception, and remodeling of plant cell walls. Annu Rev Plant Biol **71**: 39–69

Arioli T, Peng L, Betzner AS, Burn J, Wittke W, Herth W, Camilleri C, Höfte H, Plazinski J, Birch R, et al (1998) Molecular analysis of cellulose biosynthesis in Arabidopsis. Science (80-) **279**: 717–720

Atalla RH, VanderHart DL (2011) Native cellulose : a composite of two distinct crystalline. Science (80-) **223**: 283–285

Balatinecz JJ, Kretschmann DE (2001) Properties and utilization of poplar wood.

Barbier FF, Chabikwa TG, Ahsan MU, Cook SE, Powell R, Tanurdzic M, Beveridge CA (2019) A phenol/chloroform-free method to extract nucleic acids from recalcitrant, woody tropical species for gene expression and sequencing. Plant Methods **15**: 9–12

Bashline L, Li S, Anderson CT, Lei L, Gu Y (2013) The endocytosis of cellulose synthase in arabidopsis is dependent on μ2, a clathrin-mediated endocytosis adaptin. Plant Physiol **163**: 150–160

Bashline L, Li S, Zhu X, Gu Y (2015) The TWD40-2 protein and the AP2 complex cooperate in the clathrin-mediated endocytosis of cellulose synthase to regulate cellulose biosynthesis. Proc Natl Acad Sci U S A **112**: 12870–12875

Benfey PN, Linstead PJ, Roberts K, Schiefelbein JW, Hauser MT, Aeschbacher RA (1993) Root development in Arabidopsis: Four mutants with dramatically altered root

morphogenesis. Development **119**: 57–70

Benziman M, Haigler CH, Brown RM, White AR, Cooper KM (1980) Cellulose biogenesis: Polymerization and crystallization are coupled processes in Acetobacter xylinum. Proc Natl Acad Sci U S A **77**: 6678–6682

Bi G, Liu G, Zhao E, Du Q (2016) Complete mitochondrial genome of a red calcified alga Calliarthron tuberculosum (Corallinales). Mitochondrial DNA **27**: 2554–2556

Bock CW, Katz AK, Markham GD, Glusker JP (1999) Manganese as a replacement for magnesium and zinc: Functional comparison of the divalent ions. J Am Chem Soc **121**: 7360–7372

Borowitzka MA (1977) Algal calcification. Oceanogr Mar Biol Annu Rev 189–223

Borowitzka MA, Vesk M (1978) Ultrastructure of the corallinaceae. I. The vegetative cells of Corallina officinalis and C. cuvierii. Mar Biol **46**: 295–304

Brawley SH, Blouin NA, Ficko-Blean E, Wheeler GL, Lohr M, Goodson H V., Jenkins JW, Blaby-Haas CE, Helliwell KE, Chan CX, et al (2017) Insights into the red algae and eukaryotic evolution from the genome of Porphyra umbilicalis (Bangiophyceae, Rhodophyta). Proc Natl Acad Sci U S A **114**: E6361–E6370

Bringmann M, Li E, Sampathkumar A, Kocabek T, Hauser M-T, Persson S (2012) POM-POM2/CELLULOSE SYNTHASE INTERACTING1 is essential for the functional association of cellulose synthase and microtubules in Arabidopsis. Plant Cell **24**: 163–177

Brown DM, Zeef LAH, Ellis J, Goodacre R, Turner SR (2005) Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. Plant Cell **17**: 2281–2295

Brown RM, Montezinos D (1976) Cellulose microfibrils: Visualization of biosynthetic and

orienting complexes in association with the plasma membrane. Proc Natl Acad Sci U S A **73**: 143–147

Burki F, Kaplan M, Tikhonenkov D V., Zlatogursky V, Minh BQ, Radaykina L V., Smirnov A, Mylnikov AP, Keeling PJ (2016) Untangling the early diversification of eukaryotes: A phylogenomic study of the evolutionary origins of centrohelida, haptophyta and cryptista. Proc R Soc B Biol Sci. doi: 10.1098/rspb.2015.2802

Burn JE, Hocart CH, Birch RJ, Cork AC, Williamson RE (2002) Functional analysis of the cellulose synthase genes CesA1, CesA2, and CesA3 in Arabidopsis. Plant Physiol **129**: 797–807

Cai G, Faleri C, Casino CDC, Emons AMC, Cresti M (2011) Distribution of callose synthase, cellulose synthase, and sucrose synthase in tobacco pollen tube is controlled in dissimilar ways by actin filaments and microtubules. Plant Physiol **155**: 1169–1190

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics **25**: 1972–1973

Carroll A, Mansoori N, Li S, Lei L, Vernhettes S, Visser RGF, Somerville C, Gu Y, Trindade LM (2012) Complexes with mixed primary and secondary cellulose synthases are functional in Arabidopsis plants. Plant Physiol **160**: 726–737

Carroll A, Specht CD (2011) Understanding plant cellulose synthases through a comprehensive investigation of the cellulose synthase family sequences. Front Plant Sci **2**: 1–11

de Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res **34**: W362–W365

Cereghino JL, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast Pichia pastoris. FEMS Microbiol Rev **24**: 45–66

Chan CX, Yang EC, Banerjee T, Yoon HS, Martone PT, Estevez JM, Bhattacharya D (2011) Red and green algal monophyly and extensive gene sharing found in a rich repertoire of red algal genes. Curr Biol **21**: 328–333

Chen S, Ehrhardt DW, Somerville CR (2010) Mutations of cellulose synthase (CESA1) phosphorylation sites modulate anisotropic cell expansion and bidirectional mobility of cellulose synthase. Proc Natl Acad Sci U S A **107**: 17188–17193

Cheo DL, Titus SA, Byrd DRN, Hartley JL, Temple GF, Brasch MA (2004) Concerted assembly and cloning of multiple DNA segments using in vitro site-specific recombination: Functional analysis of multi-segment expression clones. Genome Res **14**: 2111–2120

Cho SH, Purushotham P, Fang C, Maranas C, Díaz-Moreno SM, Bulone V, Zimmer J, Kumar M, Nixon BT (2017) Synthesis and self-assembly of cellulose microfibrils from reconstituted cellulose synthase. Plant Physiol **175**: 146–156

Cifuentes C, Bulone V, Emons AMC (2010) Biosynthesis of callose and cellulose by detergent extracts of tobacco cell membranes and quantification of the polymers synthesized in vitro. J Integr Plant Biol **52**: 221–233

Clough SJ, Bent AF (1998) Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J **16**: 735–743

Collén J, Porcel B, Carré W, Ball SG, Chaparro C, Tonon T, Barbeyron T, Michel G, Noel B, Valentin K, et al (2013) Genome structure and metabolic features in the red seaweed Chondrus crispus shed light on evolution of the Archaeplastida. Proc Natl Acad Sci U S A **110**: 5247–5252

Cosgrove DJ (2005) Growth of the plant cell wall. Nat Rev Mol Cell Biol **6**: 850–861

Cosgrove DJ (2018) Diffuse growth of plant cell walls. Plant Physiol **176**: 16–27

Cosgrove DJ, Jarvis MC (2012) Comparative structure and biomechanics of plant primary and secondary cell walls. Front Plant Sci **3**: 1–6

Crowell EF, Bischoff V, Desprez T, Rolland A, Stierhof YD, Schumacher K, Gonneau M, Höfte H, Vernhettes S (2009) Pausing of Golgi bodies on microtubules regulates secretion of cellulose synthase complexes in Arabidopsis. Plant Cell **21**: 1141–1154

Crowell EF, Timpano H, Desprez T, Franssen-Verheijen T, Emons AM, Höfte H, Vernhettes S (2011) Differential regulation of cellulose orientation at the inner and outer face of epidermal cells in the Arabidopsis hypocotyl. Plant Cell **23**: 2592–2605

Demeritt ME (1981) Growth of hybrid poplars in Pennsylvania and Maryland clonal tests. US Department of Agriculture, Forest Service

Denny M, Mach K, Tepler S, Martone P (2013) Indefatigable: An erect coralline alga is highly resistant to fatigue. J Exp Biol **216**: 3772–3780

Desnos T, Orbović V, Bellini C, Kronenberger J, Caboche M, Traas J, Höfte H (1996) Procuste1 mutants identify two distinct genetic pathways controlling hypocotyl cell elongation, respectively in dark- and light-grown Arabidopsis seedlings. Development **122**: 683–693

Desprez T, Juraniec M, Crowell EF, Jouy H, Pochylova Z, Parcy F, Höfte H, Gonneau M, Vernhettes S (2007) Organization of cellulose synthase complexes involved in primary cell wall synthesis in Arabidopsis thaliana. Proc Natl Acad Sci U S A **104**: 15572–15577

Desprez T, Vernhettes S, Fagard M, Refrégier G, Desnos T, Aletti E, Py N, Pelletier S, Höfte H (2002) Resistance against herbicide isoxaben and cellulose deficiency caused by

distinct mutations in same cellulose synthase isoform CESA6. Plant Physiol **128**: 482–490

Ding SY, Himmel ME (2006) The maize primary cell wall microfibril: A new model derived from direct visualization. J Agric Food Chem **54**: 597–606

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. doi: 10.1186/1471-2105-5-113

Edwards K, Johnstone C, Thompson C (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res **19**: 1349

Endler A, Kesten C, Schneider R, Zhang Y, Ivakov A, Froehlich A, Funke N, Persson S (2015) A mechanism for sustained cellulose synthesis during salt stress. Cell **162**: 1353–1364

Fagard M, Desnos T, Desprez T, Goubet F, Refregier G, Mouille G, McCann M, Rayon C, Vernhettes S, Höfte H (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

Finn RD, Clements J, Eddy SR (2011) HMMER Web Server: Interactive Sequence Similarity Searching. Nucleic Acids Res **39**: W29–W37

Fisher K, Martone PT (2014) Field study of growth and calcification rates of three species of articulated coralline algae in British Columbia, Canada. Biol Bull **226**: 121–130

Flores ML, Stortz CA, Cerezo AS (2000) Studies on the skeletal cell wall of the cystocarpic stage of the red seaweed Iridaea undulosa B.Part II. Fractionation of the cell wall and methylation analysis of the inner core-fibrillar polysaccharides. Int J Biol Macromol **27**: 21–27

Foston M, Hubbell CA, Davis M, Ragauskas AJ (2009) Variations in cellulosic ultrastructure

of poplar. Bioenergy Res **2**: 193–197

Fugelstad J, Bouzenzana J, Djerbi S, Guerriero G, Ezcurra I, Teeri TT, Arvestad L, Bulone V (2009) Identification of the cellulose synthase genes from the Oomycete Saprolegnia monoica and effect of cellulose synthesis inhibitors on gene expression and enzyme activity. Fungal Genet Biol **46**: 759–767

Fujita M, Himmelspach R, Ward J, Whittington A, Hasenbein N, Liu C, Truong TT, Galway ME, Mansfield SD, Hocart CH, et al (2013) The anisotropy1 D604N mutation in the Arabidopsis cellulose synthase1 catalytic domain reduces cell wall crystallinity and the velocity of cellulose synthase complexes. Plant Physiol **162**: 74–85

Gabrielson PW, Ann Miller K, Martone PT (2011) Morphometric and molecular analyses confirm two distinct species of Calliarthron (Corallinales, Rhodophyta), a genus endemic to the northeast pacific. Phycologia **50**: 298–316

Giuffre AJ, Hamm LM, Han N, De Yoreo JJ, Dove PM (2013) Polysaccharide chemistry regulates kinetics of calcite nucleation through competition of interfacial energies. Proc Natl Acad Sci **110**: 9261 LP – 9266

Gouy M, Guindon S, Gascuel O (2009) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol **27**: 221–224

Granja PL, Ribeiro CC, De Jéso B, Baquey C, Barbosa MA (2001) Mineralization of regenerated cellulose hydrogels. J Mater Sci Mater Med **12**: 785–791

Grant CE, Bailey TL, Noble WS (2011) FIMO: scanning for occurrences of a given motif. Bioinformatics **27**: 1017–1018

Grefen C, Donald N, Hashimoto K, Kudla J, Schumacher K, Blatt MR (2010) A ubiquitin-

10 promoter-based vector set for fluorescent protein tagging facilitates temporal stability and native protein distribution in transient and stable expression studies. Plant J **64**: 355–365

Grenville-Briggs LJ, Anderson VL, Fugelstad J, Avrova AO, Bouzenzana J, Williams A, Wawra S, Whisson SC, Birch PRJ, Bulone V, et al (2008) Cellulose synthesis in Phytophthora infestans is required for normal appressorium formation and successful infection of potato. Plant Cell **20**: 720–738

Gu Y, Kaplinsky N, Bringmann M, Cobb A, Carroll A, Sampathkumar A, Baskin TI, Persson S, Somerville CR (2010) Identification of a cellulose synthase-associated protein required for cellulose biosynthesis. Proc Natl Acad Sci U S A **107**: 12866–12871

Gutierrez R, Lindeboom JJ, Paredez AR, Emons AMC, Ehrhardt DW (2009) Arabidopsis cortical microtubules position cellulose synthase delivery to the plasma membrane and interact with cellulose synthase trafficking compartments. Nat Cell Biol **11**: 797–806

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol **35**: 518–522

Holm L (2020) using dali for protein structure comparison. Methods Mol. Biol. pp 29–42

Hsieh YSY, Harris PJ (2019) Xylans of red and green algae: What is known about their structures and how they are synthesised? Polymers **11**: 1–9

Hu H, Zhang R, Tao Z, Li X, Li Y, Huang J, Li X, Han X, Feng S, Zhang G, et al (2018) Cellulose synthase mutants distinctively affect cell growth and cell wall integrity for plant biomass production in arabidopsis. Plant Cell Physiol **59**: 1144–1157

Huang S, Kiemle SN, Makarem M, Kim SH (2020) Correlation between crystalline cellulose structure and cellulose synthase complex shape: a spectroscopic study with unicellular

freshwater alga Micrasterias. Cellulose **27**: 57–69

Jonas R, Farah LF (1998) Production and application of microbial cellulose. Polym Degrad Stab **59**: 101–106

Jones DM, Murray CM, Ketelaar KJ, Thomas JJ, Villalobos JA, Wallace IS (2016) The emerging role of protein phosphorylation as a critical regulatory mechanism controlling cellulose biosynthesis. Front Plant Sci **7**: 1–12

Jones L, Ennos AR, Turner SR (2001) Cloning and characterization of irregular xylem4 (irx4): A severely lignin-deficient mutant of Arabidopsis. Plant J **26**: 205–216

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res **28**: 27–30

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc **10**: 845

Kesten C, Wallmann A, Schneider R, McFarlane HE, Diehl A, Khan GA, van Rossum BJ, Lampugnani ER, Szymanski WG, Cremer N, et al (2019) The companion of cellulose synthase 1 confers salt tolerance through a Tau-like mechanism in plants. Nat Commun **10**: 1–14

Kimura S, Laosinchai W, Itoh T, Cui X, Linder CR, Brown JRM (1999) Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant Vigna angularis. Plant Cell **11**: 2075–2085

Kitamura M, Koyama T, Nakano Y, Uemura D (2007) Characterization of a natural inducer of coral larval metamorphosis. J Exp Mar Bio Ecol **340**: 96–102

Konar B, Foster MS (1992) Distribution and recruitment of subtidal geniculate coralline alage. J Phycol **28**: 273–280

Kondo T, Sawatari C (1996) A Fourier transform infra-red spectroscopic analysis of the character of hydrogen bonds in amorphous cellulose. Polymer (Guildf) **37**: 393–399

Kost TA, Condreay JP, Jarvis DL (2005) Baculovirus as versatile vectors for protein expression in insect and mammalian cells. Nat Biotechnol **23**: 567–575

Kotov V, Bartels K, Veith K, Josts I, Subhramanyam UKT, Günther C, Labahn J, Marlovits TC, Moraes I, Tidow H, et al (2019) High-throughput stability screening for detergent-solubilized membrane proteins. Sci Rep **9**: 1–19

Kumar M, Atanassov I, Turner S (2017) Functional analysis of cellulose synthase (CESA) protein class specificity. Plant Physiol **173**: 970–983

Kumar M, Turner S (2015) Plant cellulose synthesis: CESA proteins crossing kingdoms. Phytochemistry **112**: 91–99

Kumar M, Wightman R, Atanassov I, Gupta A, Hurst CH, Hemsley PA, Turner S (2016) S-acylation of the cellulose synthase complex is essential for its plasma membrane localization. Science (80-) **353**: 166–169

Labeeuw L, Martone PT, Boucher Y, Case RJ (2015) Ancient origin of the biosynthesis of lignin precursors. Biol Direct **10**: 1–21

Lahaye M, Rondeau-Mouro C, Deniaud E, Buléon A (2003) Solid-state 13C NMR spectroscopy studies of xylans in the cell wall of Palmaria palmata (L. Kuntze, Rhodophyta). Carbohydr Res **338**: 1559–1569

Lai-Kee-Him J, Chanzy H, Müller M, Putaux JL, Imai T, Bulone V (2002) In vitro versus in vivo cellulose microfibrils from plant primary wall synthases: Structural differences. J Biol Chem **277**: 36931–36939

Lampugnani ER, Flores-Sandoval E, Tan QW, Mutwil M, Bowman JL, Persson S (2019)

Cellulose synthesis – central components and their evolutionary relationships. Trends Plant Sci **24**: 402–412

Lampugnani ER, Khan GA, Somssich M, Persson S (2018) Building a plant cell wall at a glance. J Cell Sci. doi: 10.1242/jcs.207373

Langhans M, Förster S, Helmchen G, Robinson DG (2011) Differential effects of the brefeldin A analogue (6R)-hydroxy-BFA in tobacco and Arabidopsis. J Exp Bot **62**: 2949–2957

Lechat H, Amat M, Mazoyer J, Buléon A, Lahaye M (2000) Structure and distribution of glucomannan and sulfated glucan in the cell walls of the red alga Kappaphycus alvarezii. J Phycol **36**: 891–902

Lewis LA, McCourt RM (2004) Green algae and the origin of land plants. Am J Bot **91**: 1535–1556

Li L, Richard R. Drake J, Clement S, Brown RM (1993) β-glucan synthesis in the cotton fiber: iii. identification of udp-glucose-binding subunits of β-glucan synthases by photoaffinity labeling with [β-32P]5′-N_3-UDP-Glucose. Plant Physiol **101**: 1149–1156

Li L, Saga N, Mikami K (2008) Effects of cell wall synthesis on cell polarity in the red alga Porphyra yezoensis. Plant Signal Behav **3**: 1126–1128

Li X, Speicher TL, Dees DCT, Mansoori N, McManus JB, Tien M, Trindade LM, Wallace IS, Roberts AW (2019) Convergent evolution of hetero-oligomeric cellulose synthesis complexes in mosses and seed plants. Plant J **99**: 862–876

Little A, Schwerdt JG, Shirley NJ, Khor SF, Neumann K, O'donovan LA, Lahnstein J, Collins HM, Henderson M, Fincher GB, et al (2018) Revised phylogeny of the cellulose synthase gene superfamily: Insights into cell wall evolution. Plant Physiol **177**: 1124–1141

Liu L, Shang-Guan K, Zhang B, Liu X, Yan M, Zhang L, Shi Y, Zhang M, Qian Q, Li J, et al (2013) Brittle Culm1, a COBRA-Like protein, functions in cellulose assembly through binding cellulose microfibrils. PLOS Genet **9**: e1003704

Liu Z, Schneider R, Kesten C, Zhang Y, Somssich M, Zhang Y, Fernie AR, Persson S (2016) Cellulose-microtubule uncoupling proteins prevent lateral displacement of microtubules during cellulose synthesis in arabidopsis. Dev Cell **38**: 305–315

Luo A, Qiao H, Zhang Y, Shi W, Ho SY, Xu W, Zhang A, Zhu C (2010) Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. BMC Evol Biol **10**: 242

MacKinnon IM, Šturcová A, Sugimoto-Shirasu K, His I, McCann MC, Jarvis MC (2006) Cell-wall structure and anisotropy in procuste, a cellulose synthase mutant of Arabidopsis thaliana. Planta **224**: 438–448

Mahram A, Herbordt MC (2010) Fast and Accurate NCBI BLASTP: Acceleration with Multiphase FPGA-based Prefiltering. Proc. 24th ACM Int. Conf. Supercomput. ACM, New York, NY, USA, pp 73–82

Mansoori N, Timmers J, Desprez T, Kamei CLA, Dees DCT, Vincken JP, Visser RGF, Fte HH, Vernhettes S, Trindade LM (2014) KORRIGAN1 interacts specifically with integral components of the Cellulose synthase machinery. PLoS One **9**: 4–11

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, et al (2014) CDD: NCBI's conserved domain database. Nucleic Acids Res **43**: D222–D226

Martone PT (2010) Quantifying growth and calcium carbonate deposition of Calliarthron cheilosporioides (Corallinales, Rhodophyta) in the field using a persistent vital stain: Note. J

Phycol **46**: 13–17

Martone PT (2007) Kelp versus coralline: Cellular basis for mechanical strength in the wave-swept seaweed Calliarthron (Corallinaceae, Rhodophyta). J Phycol **43**: 882–891

Martone PT, Denny MW (2008) To bend a coralline: Effect of joint morphology on flexibility and stress amplification in an articulated calcified seaweed. J Exp Biol **211**: 3421–3432

Martone PT, Estevez JM, Lu F, Ruel K, Denny MW, Somerville C, Ralph J (2009) Discovery of lignin in seaweed reveals convergent evolution of cell-wall architecture. Curr Biol **19**: 169–175

Martone PT, Janot K, Fujita M, Wasteneys G, Ruel K, Joseleau J-P, Estevez JM (2019) Cellulose-rich secondary walls in wave-swept red macroalgae fortify flexible tissues. Planta. doi: 10.1007/s00425-019-03269-1

Martone PT, Navarro DA, Stortz CA, Estevez JM (2010) Differences in polysaccharide structure between calcified and uncalcified segments in the coralline calliarthron cheilosporioides (corallinales, rhodophyta). J Phycol **46**: 507–515

Matahwa H, Ramiah V, Sanderson RD (2008) Calcium carbonate crystallization in the presence of modified polysaccharides and linear polymeric additives. J Cryst Growth **310**: 4561–4569

Matthews PR, Schindler M, Howles P, Arioli T, Williamson RE (2010) A CESA from Griffithsia monilis (Rhodophyta, Florideophyceae) has a family 48 carbohydrate-binding module. J Exp Bot **61**: 4461–4468

McFarlane HE, Döring A, Persson S (2014) The cell biology of cellulose synthesis. Annu Rev Plant Biol **65**: 69–94

McNamara JT, Morgan JLW, Zimmer J (2015) A molecular description of cellulose

biosynthesis. Annu Rev Biochem **84**: 895–921

McNeil M, Darvill AG, Fry SC, Albersheim P (1984) Structure and function of the primary cell walls of plants. Annu Rev Biochem **53**: 625–663

Meents MJ, Watanabe Y, Samuels AL (2018) The cell biology of secondary cell wall biosynthesis. Ann Bot **121**: 1107–1125

Meeuse BJD, Andries M, Wood JA (1960) Floridean Starch. J Exp Bot **11**: 129–140

Miart F, Desprez T, Biot E, Morin H, Belcram K, Höfte H, Gonneau M, Vernhettes S (2014) Spatio-temporal analysis of cellulose synthesis during cell plate formation in Arabidopsis. Plant J **77**: 71–84

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

Morgan JLW, McNamara JT, Fischer M, Rich J, Chen HM, Withers SG, Zimmer J (2016) Observing cellulose biosynthesis and membrane translocation in crystallo. Nature **531**: 329–334

Morgan JLW, Strumillo J, Zimmer J (2013) Crystallographic snapshot of cellulose synthesis and membrane translocation. Nature **493**: 181–186

Mueller SC, Brown RMJ (1980) Evidence for an intramembrane component associated with a cellulose microfibril-synthesizing complex in higher plants. J Cell Biol **84**: 315–326

Nash MC, Diaz-Pulido G, Harvey AS, Adey W (2019) Coralline algal calcification: A morphological and process-based understanding. PLoS One **14**: 1–61

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol **32**:

268–274

Nicol F, His I, Jauneau A, Vernhettes S, Canut H, Höfte H (1998) A plasma membrane-bound putative endo-1,4-β-D-glucanase is required for normal wall assembly and cell elongation in Arabidopsis. EMBO J **17**: 5563–5576

Nishiyama Y, Sugiyama J, Chanzy H, Langan P (2003) Crystal structure and hydrogen bonding system in cellulose iα from synchrotron x-ray and neutron fiber diffraction. J Am Chem Soc **125**: 14300–14306

Nixon BT, Mansouri K, Singh A, Du J, Davis JK, Lee JG, Slabaugh E, Vandavasi VG, O'Neill H, Roberts EM, et al (2016) Comparative structural and computational analysis supports eighteen cellulose synthases in the plant cellulose synthesis complex. Sci Rep **6**: 1–14

Nobles DR, Romanovicz DK, Brown J (2001) Cellulose in cyanobacteria. Origin of vascular plant cellulose synthase? Plant Physiol **127**: 529–542

Norris SR, Meyer SE, Callis J (1993) The intron of Arabidopsis thaliana polyubiquitin genes is conserved in location and is a quantitative determinant of chimeric gene expression. Plant Mol Biol **21**: 895–906

Okuda K, Tsekos L, Brown RM (1994) Cellulose microfibril assembly in Erythrocladia subintegra Rosenv.: An ideal system for understanding the relationship between synthesizing complexes (TCs) and microfibril crystallization. Protoplasma **180**: 49–58

Omadjela O, Narahari A, Strumillo J, Mélida H, Mazur O, Bulone V, Zimmer J (2013) BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for in vitro cellulose synthesis. Proc Natl Acad Sci U S A **110**: 17856–17861

Pang Z, McKee LS, Srivastava V, Klinter S, Díaz-Moreno SM, Orlean P, Liu X, Bulone V

(2020) Analysis of a cellulose synthase catalytic subunit from the oomycete pathogen of crops Phytophthora capsici. Cellulose **9**: 8551–8565

Paredez AR, Persson S, Ehrhardt DW, Somerville CR (2008) Genetic evidence that cellulose synthase activity influences microtubule cortical array organization. Plant Physiol **147**: 1723–1734

Paredez AR, Somerville CR, Ehrhardt DW (2006) Visualization of cellulose synthase demonstrates functional association with microtubules. Science (80-) **312**: 1491–1495

Park S, Ding SY (2020) The N-terminal zinc finger of CELLULOSE SYNTHASE6 is critical in defining its functional properties by determining the level of homodimerization in Arabidopsis. Plant J **103**: 1826–1838

Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM (1996) Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. Proc Natl Acad Sci **93**: 12637 LP – 12642

Peña V, Vieira C, Braga JC, Aguirre J, Rösler A, Baele G, De Clerck O, Le Gall L (2020) Radiation of the coralline red algae (Corallinophycidae, Rhodophyta) crown group as inferred from a multilocus time-calibrated phylogeny. Mol Phylogenet Evol **150**: 106845

Persson S, Caffall KH, Freshour G, Hilley MT, Bauer S, Poindexter P, Hahn MG, Mohnen D, Somerville C (2007a) The Arabidopsis irregular xylem8 mutant is deficient in glucuronoxylan and homogalacturonan, which are essential for secondary cell wall integrity. Plant Cell **19**: 237–255

Persson S, Paredez A, Carroll A, Palsdottir H, Doblin M, Poindexter P, Khitrov N, Auer M, Somerville CR (2007b) Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in Arabidopsis. Proc Natl Acad Sci U S A **104**: 15566–

15571

Polekhina G, Gupta A, Van Denderen BJW, Feil SC, Kemp BE, Stapleton D, Parker MW (2005) Structural basis for glycogen recognition by AMP-activated protein kinase. Structure **13**: 1453–1462

Polko JK, Kieber JJ (2019) The regulation of cellulose biosynthesis in plants. Plant Cell **31**: 282–296

Popper ZA (2008) Evolution and diversity of green plant cell walls. Curr Opin Plant Biol **11**: 286–292

Privé GG (2007) Detergents for the stabilization and crystallization of membrane proteins. Methods **41**: 388–397

Purushotham P, Cho SH, Díaz-Moreno SM, Kumar M, Nixon BT, Bulone V, Zimmer J (2016) A single heterologously expressed plant cellulose synthase isoform is sufficient for cellulose microfibril formation in vitro. Proc Natl Acad Sci U S A **113**: 11360–11365

Purushotham P, Ho R, Zimmer J (2020) Architecture of a catalytically active homotrimeric plant cellulose synthase complex. Science **2978**: 1–11

Qiu H, Price DC, Yang EC, Yoon HS, Bhattacharya D (2015) Evidence of ancient genome reduction in red algae (Rhodophyta). J Phycol **51**: 624–636

Qiu H, Yoon HS, Bhattacharya D (2016) Red algal phylogenomics provides a robust framework for inferring evolution of key metabolic pathways. PLoS Curr. doi: 10.1371/currents.tol.7b037376e6d84a1be34af756a4d90846

Rambaut A, Drummond A (2010) FigTree v1. 3.1 Institute of Evolutionary Biology. Univ. Edinburgh

Ramírez-Rodríguez EA, McFarlane HE (2020) Insights from the structure of a plant cellulose

synthase trimer. Trends Plant Sci **xx**: 3–5

Rice P, Longden I, Bleasby A (2000) EMBOSS: The european molecular biology open software suite. Trends Genet **16**: 276–277

Richmond TA, Somerville CR (2000) The cellulose synthase superfamily. Plant Physiol **124**: 495 LP – 498

Rigaud J-L, Lévy D (2003) Reconstitution of membrane proteins in liposomes. Methods Enzymol. pp 65–86

Roberts AW, Bushoven JT (2007) The cellulose synthase (CESA) gene superfamily of the moss Physcomitrella patens. Plant Mol Biol **63**: 207–219

Roberts E, Roberts AW (2009) A cellulose synthase (CESA) gene from the red alga porphyra yezoensis (rhodophyta). J Phycol **45**: 203–212

Roest S, Kapps-Fouthier S, Klopp J, Rieffel S, Gerhartz B, Shrestha B (2016) Transfection of insect cell in suspension for efficient baculovirus generation. MethodsX **3**: 371–377

Rombel IT, Sykes KF, Rayner S, Johnston SA (2002) ORF-FINDER: a vector for high-throughput gene identification. Gene **282**: 33–41

Roudier F, Fernandez AG, Fujita M, Himmelspach R, Borner GHH, Schindelman G, Song S, Baskin TI, Dupree P, Wasteneys GO, et al (2005) COBRA, an Arabidopsis extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation. Plant Cell **17**: 1749–1763

Roudier F, Schindelman G, DeSalle R, Benfey PN (2002) The COBRA family of putative GPI-anchored proteins in Arabidopsis. A new fellowship in expansion. Plant Physiol **130**: 538–548

Rudolph U, Schnepf E (1988) Investigations of the turnover of the putative Cellulose-synthesizing particle “rosettes” within the plasma membrane of Funaria hygrometrica protonema cells - I. Effects of monensin and cytochalasin B. Protoplasma **143**: 63–73

Sánchez-Rodríguez C, Bauer S, Hématy K, Saxe F, Ibáñez AB, Vodermaier V, Konlechner C, Sampathkumar A, Rüggeberg M, Aichinger E, et al (2012) CHITINASE-LIKE1/POM-POM1 and its homolog CTL2 are glucan-interacting proteins important for cellulose biosynthesis in Arabidopsis. Plant Cell **24**: 589–607

Sánchez-Rodríguez C, Shi Y, Kesten C, Zhang D, Sancho-Andrés G, Ivakov A, Lampugnani ER, Sklodowski K, Fujimoto M, Nakano A, et al (2018) The cellulose synthases are cargo of the TPLATE adaptor complex. Mol Plant **11**: 346–349

Sato K, Ito S, Fujii T, Suzuki R, Takenouchi S, Nakaba S, Funada R, Sano Y, Kajita S, Kitano H, et al (2010a) The carbohydrate-binding module (CBM)-like sequence is crucial for rice CWA1/BC1 function in proper assembly of secondary cell wall materials. Plant Signal Behav. doi: 10.4161/psb.5.11.13342

Sato K, Suzuki R, Nishikubo N, Takenouchi S, Ito S, Nakano Y, Nakaba S, Sano Y, Funada R, Kajita S, et al (2010b) Isolation of a novel cell wall architecture mutant of rice with defective Arabidopsis COBL4 ortholog BC1 required for regulated deposition of secondary cell wall components. Planta **232**: 257–270

Scheffzek K, Welti S (2012) Pleckstrin homology (PH) like domains - Versatile modules in protein-protein interaction platforms. FEBS Lett **586**: 2662–2673

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al (2012) Fiji: An open-source platform for biological-image analysis. Nat Methods **9**: 676–682

Schindelman G, Morikami A, Jung J, Baskin TI, Carpita NC, Derbyshire P, McCann MC, Benfey PN (2001) COBRA encodes a putative GPI-anchored protein, which is polarly localized and necessary for oriented cell expansion in Arabidopsis . Genes Dev **15**: 1115–1127

Schmidt TGM, Skerra A (2007) The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. Nat Protoc **2**: 1528–1535

Seddon AM, Curnow P, Booth PJ (2004) Membrane proteins, lipids and detergents: Not just a soap opera. Biochim Biophys Acta - Biomembr **1666**: 105–117

Sethaphong L, Haigler CH, Kubicki JD, Zimmer J, Bonetta D, DeBolt S, Yingling YG (2013) Tertiary model of a plant cellulose synthase. Proc Natl Acad Sci U S A **110**: 7512–7517

Smith RA, Schuetz M, Roach M, Mansfield SD, Ellis B, Samuels L (2013) Neighboring parenchyma cells contribute to Arabidopsis xylem lignification, while lignification of interfascicular fibers is cell autonomous. Plant Cell **25**: 3988–3999

Smith S V (1972) Production of calcium carbonate on the mainland shelf of Southern California. Limnol Oceanogr **17**: 28–41

Steneck s. R, Martone TP (2007) Algae, Calcified. Encycl. Tidepools Rocky Shores. University of California Press, pp 21–24

Suzuki S, Li L, Sun YH, Chiang VL (2006) The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in Populus trichocarpa. Plant Physiol **142**: 1233–1245

Szyjanowicz PMJ, McKinnon I, Taylor NG, Gardiner J, Jarvis MC, Turner SR (2004) The irregular xylem 2 mutant is an allele of korrigan that affects the secondary cell wall of

Arabidopsis thaliana. Plant J **37**: 730–740

Tatusov RL, Koonin E V, Lipman DJ (1997) A genomic perspective on protein families. Science **278**: 631–637

Taylor NG, Gardiner JC, Whiteman R, Turner SR (2004) Cellulose synthesis in the Arabidopsis secondary cell wall. Cellulose **11**: 329–338

Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR (2003) Interactions among three distinct CesA proteins essential for cellulose synthesis. Proc Natl Acad Sci U S A **100**: 1450–1455

Taylor NG, Laurie S, Turner SR (2000) Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in Arabidopsis. Plant Cell **12**: 2529–2539

Taylor NG, Scheible WR, Cutler S, Somerville CR, Turner SR (1999) The irregular xylem3 locus of Arabidopsis encodes a cellulose synthase required for secondary cell wall synthesis. Plant Cell **11**: 769–779

Tester RF, Karkalas J, Qi X (2004) Starch - Composition, fine structure and architecture. J Cereal Sci **39**: 151–165

Timell TE (1967) Recent progress in the chemistry of wood hemicelluloses. Wood Sci Technol **1**: 45–70

Tsekos I (1999) The sites of cellulose synthesis in algae: Diversity and evolution of cellulose-synthesizing enzyme complexes. J Phycol **35**: 635–655

Tsekos I, Reiss H-D (1994) Tip cell growth and the frequency and distribution of cellulose microfibril-synthesizing complexes in the plasma membrane of apical shoot cells of the red alga porphyra yezoensis. J Phycol **30**: 300–310

Tsekos I, Reiss H-D, Schnepf E (1993) Cell-wall structure and supramolecular organization of

the plasma membrane of marine red algae visualized by freeze-fracture. Acta Bot Neerl **42**: 119–132

Tsekos I, Reiss H (1992) Occurrence of the putative microfibril-synthesizing complexes (linear terminal complexes) in the plasma membrane of the epiphytic marine red alga Erythrocladia subiptegra Rosenv . 57–67

Tsirigos KD, Peters C, Shu N, Kall L, Elofsson A (2015) The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Res **43**: W401-7

Turner SR, Somerville CR (1997) Collapsed xylem phenotype of arabidopsis identifies mutants deficient in cellulose deposition in the secondary cell wall. Plant Cell **9**: 689–701

Turner SR, Taylor N, Jones L (2001) Mutations of the secondary cell wall. Plant Mol Biol **47**: 209–219

Usov AI, Bilan MI, Shashkov AS (1997) Structure of a sulfated xylogalactan from the calcareous red alga Corallina pilulifera P. et R. (Rhodophyta, Corallinaceae). Carbohydr Res **303**: 93–102

Vain T, Crowell EF, Timpano H, Biot E, Desprez T, Mansoori N, Trindade LM, Pagant S, Robert S, Höfte H, et al (2014) The cellulase KORRIGAN is part of the cellulose synthase complex. Plant Physiol **165**: 1521–1532

Vandavasi VG, Putnam DK, Zhang Q, Petridis L, Heller WT, Tracy Nixon B, Haigler CH, Kalluri U, Coates L, Langan P, et al (2016) A structural study of CESA1 catalytic domain of arabidopsis cellulose synthesis complex: Evidence for CESA trimers. Plant Physiol **170**: 123–135

Viola R, Nyvall P, Pedersén M (2001) The unique features of starch metabolism in red algae.

Proc R Soc B Biol Sci **268**: 1417–1422

Vogl T, Glieder A (2013) Regulation of Pichia pastoris promoters and its consequences for protein production. N Biotechnol **30**: 385–404

Wang Q, Sun H, Huang J (2017) Re-analyses of "algal" genes suggest a complex evolutionary history of oomycetes. Front Plant Sci **8**: 1–14

Watanabe Y, Meents MJ, McDonnell LM, Barkwill S, Sampathkumar A, Cartwright HN, Demura T, Ehrhardt DW, Samuels AL, Mansfield SD (2015) Visualization of cellulose synthases in Arabidopsis secondary cell walls. Science (80-) **350**: 198–204

Watanabe Y, Schneider R, Barkwill S, Gonzales-Vigil E, Hill JL, Lacey Samuels A, Persson S, Mansfield SD (2018) Cellulose synthase complexes display distinct dynamic behaviors during xylem transdifferentiation. Proc Natl Acad Sci U S A **115**: E6366–E6374

Wightman R, Turner S (2010) Trafficking of the plant cellulose synthase complex. Plant Physiol **153**: 427–432

Williams EA, Craigie A, Yeates A, Degnan SM (2008) Articulated coralline algae of the genus Amphiroa are highly effective natural inducers of settlement in the tropical abalone Haliotis asinina. Biol Bull **215**: 98–107

Yoon HS, Müller KM, Sheath RG, Ott FD, Bhattacharya D (2006) Defining the major lineages of red algae (Rhodophyta). J Phycol **42**: 482–492

Yu S, Blennow A, Bojko M, Madsen F, Olsen CE, Engelsen SB (2002) Physico-chemical characterization of floridean starch of red algae. Starch/Staerke **54**: 66–74

Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM, et al (2015) Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nat Biotechnol **33**: 531–537

Zhang Y, Nikolovski N, Sorieul M, Vellosillo T, McFarlane HE, Dupree R, Kesten C, Schneider R, Driemeier C, Lathe R, et al (2016) Golgi-localized STELLO proteins regulate the assembly and trafficking of cellulose synthase complexes in Arabidopsis. Nat Commun. doi: 10.1038/ncomms11656

Appendix

This appendix contains supplemental information for chapter 2 of this thesis. This includes appendix figure 1 which is an expansion on figure 2.2 that includes a larger taxonomic sampling of major clades. Some support values are not shown and clades have been collapsed in order to improve readability of the figure. The full, unmodified, tree can be accessed for more detailed view at https://github.com/janxue/2020_CESA_thesis_work. Support values shown are Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT) and ultrafast bootstrap respectively (Appendix Figure 1).

Appendix table 1, 2, and 3 contain the sequence identifiers for the HMM profile used to identify CESAs (Table 2.1), the CESA gene tree (Figure 2.2), and the HMM profile used to identify CESA accessory proteins (Table 2.4) respectively.

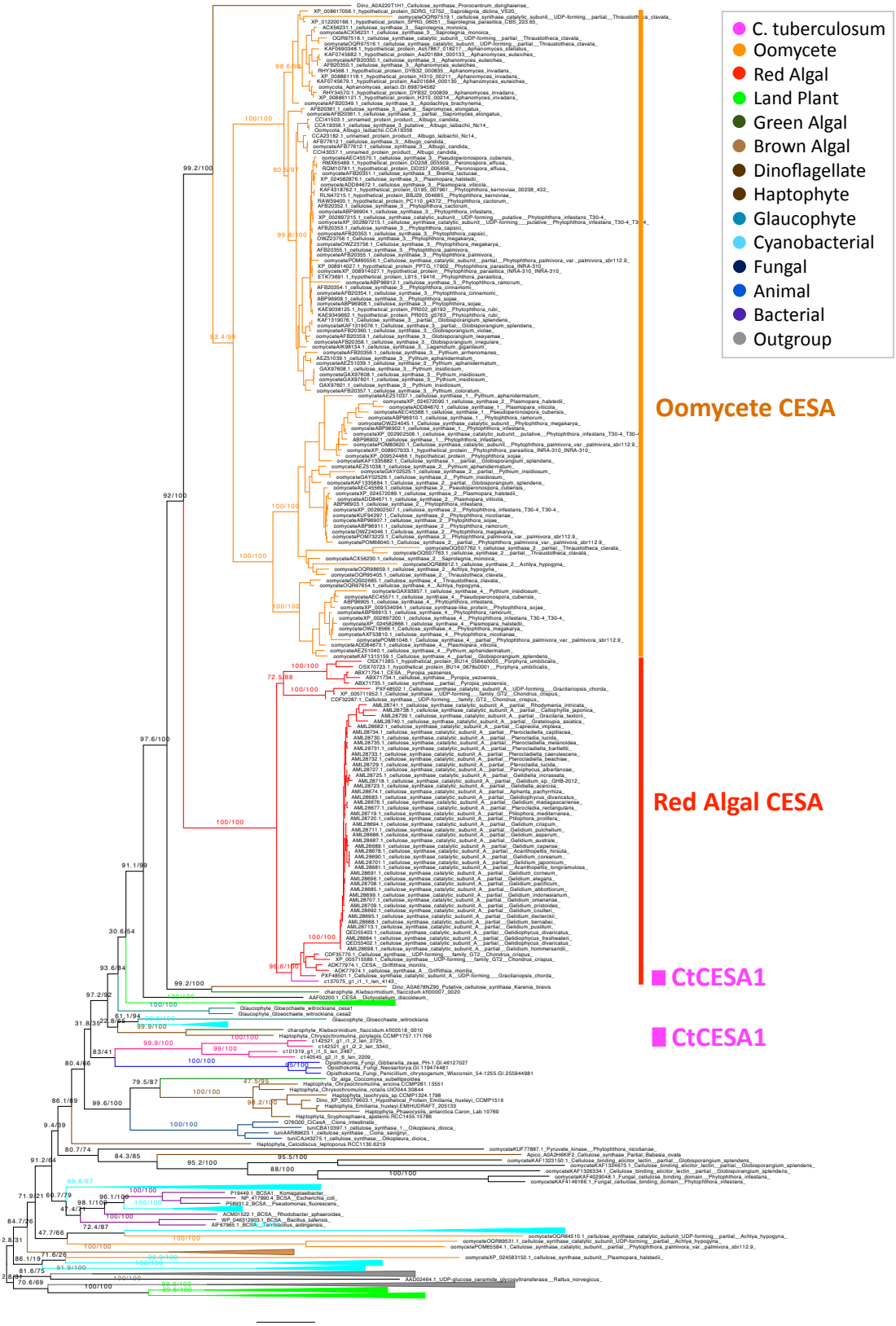
C. tuberculosum
Oomycete
Red Algal
Land Plant
Green Algal
Brown Algal
Dinoflagellate
Haptophyte
Glaucophyte
Cyanobacterial
Fungal
Animal
Bacterial
Outgroup
Oomycete CESA
Red Algal CESA
CtCESA1
CtCESA1
0.6

Figure S1 Rooted maximum likelihood CESA gene tree highlighting the position of red algal, oomycete, and *C. tuberculosum* (magenta) sequences. This tree builds on figure 2.2, expanded with additional taxonomic sampling. Support (SH-aLRT support / ultrafast bootstrap support) is indicated for major branching points. Some support values are omitted and some major clades containing CESA sequences have been collapsed to improve readability. Alignment sites with <80% coverage were removed. IQtree was used to search for the model and estimate the maximum likelihood tree. The model under BIC criterion was VT+F+G4. Log-likelihood of the ML tree = -254632.5801.

Table S1 List of organisms and their identifiers used as the HMM profile to search for the CESA in the *C. tuberculosum* transcriptomic dataset.

Protein Profile	Major Lineage	*Organism*	Identifier	GenBank Accession Number	Uniprot Accession Number
CESA	**Oomycete**	*Phytophthora infestans*	piCESA1		ABP96902.1
		Phytophthora infestans	piCESA2		ABP96903.1
		Phytophthora infestans	piCESA3		ABP96904.1
		Phytophthora infestans	piCESA4		ABP96905.1
	Red Algae	*Chondrus crispus*	CESA 1	CDF32287.1	
		Chondrus crispus	CESA 2	CDF35770.1	
		Griffithsia monilis	CESA	ADK77974.1	
		Pyropia yezoensis	CESA	ABX71734.1	
	Blue-Green Algae	*Anabaena variabilis*	CESA	ABA21191	
		Nostoc sp.	CESA	ACC80169	
		Synechococcus sp.	CESA	ACB00100	
	Brown algae	*Ectocarpus siliculosus*	CESA1	CBJ28249.1	
		Ectocarpus siliculosus	CESA2	CBJ28245.1	
		Ectocarpus siliculosus	CESA3	CBJ30572.1	
		Ectocarpus siliculosus	CESA4	CBJ30317.1	
	Angiosperm	*Gossypium hirsutum*	CESA1	AAB37766.1	
		Gossypium hirsutum	CESA3	AAD39534.2	
		Gossypium hirsutum	CESA4	AAL37718.1	
		Gossypium hirsutum	CESA5	ADD71499.1	
		Gossypium hirsutum	CESA6	AFB18635.1	
		Gossypium hirsutum	CESA7	AFB18636.1	
		Gossypium hirsutum	CESA8	AFB18637.1	

Protein Profile	Major Lineage	*Organism*	Identifier	GenBank Accession Number	Uniprot Accession Number
		Gossypium hirsutum	CESA10	AFB18638.1	
		Arabidopsis thaliana	CESA1 (At4g32410; Rsw1)	AAC39334.1	
		Arabidopsis thaliana	CESA2 (At4g39350)	AAC39335.1	
		Arabidopsis thaliana	CESA3 (At5g05170; Cev1; Rsw5)	AAC39336.1	
		Arabidopsis thaliana	CESA4 (At5g44030; Irx5)	AAO15532.1	
		Arabidopsis thaliana	CESA5 (At5g09870)	BAB09408.1	
		Arabidopsis thaliana	CESA 6 (At5g64740; Prc1; Irx2)	BAB10307.1	
		Arabidopsis thaliana	CESA7 (At5g17420; Irx3)	AAD32031.1	
		Arabidopsis thaliana	CESA8 (IAt4g18780; Irx1)	AAM20487.1	
		Arabidopsis thaliana	CESA9 (At2g21770)	AAD20396.1	
		Arabidopsis thaliana	CESA10 (At2g25540)	AAD20713.1	
	Bryophyte	*Physcomitrella patens*	CESA3		A9RGN5
		Physcomitrella patens	CESA4	ABI78957.1	
		Physcomitrella patens	CESA5	ABI78958.1	
		Physcomitrella patens	CESA6	ABI78959.1	
		Physcomitrella patens	CESA7	ABI78960.1	
		Physcomitrella patens	CESA8	ABI78961.1	
	Bacteria	*Komagataeibacter*	BCSA1	P19449.1	
		Escherichia coli	BCSA	NP_417990.4	

Protein Profile	Major Lineage	*Organism*	Identifier	GenBank Accession Number	Uniprot Accession Number
		Rhodobacter	BCSA	ACM01522.1	
		Pseudomonas	BCSA	P58931.2	
		Terribacillus	BCSA	AIF67965.1	
		Bacillus	BCSA	WP_046312903.1	

Table S2 List of organisms and their identifiers used in *CESA* gene tree analysis, see Figure 2.2.

Major Lineage	*Organism*	Identifier	GenBank	Uniprot
Oomycete	*Phytophthora infestans*	CESA1		ABP96902.1
	Phytophthora infestans	CESA2		ABP96903.1
	Phytophthora infestans	CESA3		ABP96904.1
	Phytophthora infestans	CESA4		ABP96905.1
Rhodophyta (Red Algae)	*Chondrus crispus*	CESA 1	CDF32287.1	
	Chondrus crispus	CESA 2	CDF35770.1	
	Griffithsia monilis	CESA	ADK77974.1	
	Pyropia yezoensis	CESA	ABX71734.1	
Blue-Green Algae (Cyanobacteria)	*Anabaena variabilis*	CESA	ABA21191	
	Nostoc sp.	CESA	ACC80169	
	Synechococcus sp.	CESA	ACB00100	
	Coleofasciculus chthonoplastes	CESA	EDX74293.1	
	Trichormus variabilis	CESA	ABA21191.1	
Phaeophyceae (Brown algae Multicellular)	*Ectocarpus siliculosus*	CESA1	CBJ28249.1	
	Ectocarpus siliculosus	CESA2	CBJ28245.1	
	Ectocarpus siliculosus	CESA3	CBJ30572.1	
	Ectocarpus siliculosus	CESA4	CBJ30317.1	
Angiosperm	*Gossypium hirsutum*	CESA1	AAB37766.1	
	Gossypium hirsutum	CESA3	AAD39534.2	
	Gossypium hirsutum	CESA4	AAL37718.1	
	Gossypium hirsutum	CESA5	ADD71499.1	
	Gossypium hirsutum	CESA6	AFB18635.1	
	Gossypium hirsutum	CESA7	AFB18636.1	

Major Lineage	*Organism*	Identifier	GenBank	Uniprot
	Gossypium hirsutum	CESA8	AFB18637.1	
	Gossypium hirsutum	CESA9	AFB18639.1	
	Gossypium hirsutum	CESA10	AFB18638.1	
	Gossypium hirsutum	CSL D5	ADQ28096.1	
	Arabidopsis thaliana	CESA1 (At4g32410; Rsw1)	AAC39334.1	
	Arabidopsis thaliana	CESA2 (At4g39350)	AAC39335.1	
	Arabidopsis thaliana	CESA3 (At5g05170; Cev1; Rsw5)	AAC39336.1	
	Arabidopsis thaliana	CESA4 (At5g44030; Irx5)	AAO15532.1	
	Arabidopsis thaliana	CESA5 (At5g09870)	BAB09408.1	
	Arabidopsis thaliana	CESA 6 (At5g64740; Prc1; Irx2)	BAB10307.1	
	Arabidopsis thaliana	CESA7 (At5g17420; Irx3)	AAD32031.1	
	Arabidopsis thaliana	CESA8 (IAt4g18780; Irx1)	AAM20487.1	
	Arabidopsis thaliana	CESA9 (At2g21770)	AAD20396.1	
	Arabidopsis thaliana	CESA10 (At2g25540)	AAD20713.1	
	Arabidopsis thaliana	CslA1 (At4g16590; β-1,4-mannan synthase / glucomannan synthase)	AAO42230.1	
	Arabidopsis thaliana	CslA2 (At5g22740; glucomannan synthase)	BAB11680.1	
	Arabidopsis thaliana	CslA7 (At2g35650; glucomannan synthase)	AAL24081.1	
	Arabidopsis thaliana	CslA9 (At5g03760; glucomannan synthase)	CAB82941.1	
	Arabidopsis	CslA14 (At3g56000)	NP_191159.2	

Major Lineage	*Organism*	Identifier	GenBank	Uniprot
	Arabidopsis thaliana			
	Arabidopsis thaliana	CslB2 (At2g32620)	AAC25944.1	
	Arabidopsis thaliana	CslB3 (At2g32530)	AAL90907.1	
	Arabidopsis thaliana	CslB4 (At2g32540)	AAC25936.1	
	Arabidopsis thaliana	CslB5 (At4g15290)	NP_193264.3	
	Arabidopsis thaliana	CslB6 (At4g15320)	CAB78574.1	
	Arabidopsis thaliana	CslC4 (At3g28180; xyloglucan β-1,4-glucan synthase)	BAB01433.1	
		CslC8 (At2g24630)	AAD23884.1	
	Arabidopsis thaliana	CslC12 (At4g07960)	AAD15482.1	
	Arabidopsis thaliana	CslD1 (At2g33100)	AAC04910.1	
	Arabidopsis thaliana	CslD2 (At5g16910; glucomannan synthase)	CAC01704.1	
	Arabidopsis thaliana	CslD3 (At3g03050; glucomannan synthase)	AAF26119.1	
	Arabidopsis thaliana	CslD4 (At4g38190)	CAB37559.1	
	Arabidopsis thaliana	CslD5 (At1g02730; glucomannan synthase)	AAF02892.1	
	Arabidopsis thaliana	CslE1 (At1g55850)	AAF79313.1	
	Arabidopsis thaliana	CslG1 (At4g24010)	AAB63622.1	
	Arabidopsis thaliana	CslG2 (At4g24000)	AAM20086.1	
	Arabidopsis thaliana	CslG3 (At4g23990)	NP_194130	
Bryophyte	*Physcomitrella patens*	CESA3		A9RGN5
	Physcomitrella patens	CESA4	ABI78957.1	
	Physcomitrella patens	CESA5	ABI78958.1	

Major Lineage	*Organism*	Identifier	GenBank	Uniprot
	Physcomitrella patens	CESA6	ABI78959.1	
	Physcomitrella patens	CESA7	ABI78960.1	
	Physcomitrella patens	CESA8	ABI78961.1	
	Physcomitrella patens	CslA1	ABD79099.1	
	Physcomitrella patens	CslA2	ABD79100.1	
	Physcomitrella patens	CslC1	ABI55233.1	
	Physcomitrella patens	CslC2	ABI55234.1	
	Physcomitrella patens	CslC3	ABI55235.1	
	Physcomitrella patens	CslD1	ABI75151.1	
	Physcomitrella patens	CslD2	ABI75152.1	
	Physcomitrella patens	CslD3	ABI75153.1	
	Physcomitrella patens	CslD4	ABI75154.1	
	Physcomitrella patens	CslD5	ABI75155.1	
	Physcomitrella patens	CslD6	ABI75156.1	
	Physcomitrella patens	CslD7	ABI75157.1	
	Physcomitrella patens	CslD8	ABI75158.1	
Bacteria	*Komagataeibacter*	BCSA1	P19449.1	
	Escherichia coli	BCSA	NP_417990.4	
	Rhodobacter	BCSA	ACM01522.1	
	Pseudomonas	BCSA	P58931.2	
	Terribacillus	BCSA	AIF67965.1	
	Bacillus	BCSA	WP_046312903.1	
Chordate	*Ciona intestinalis*	CESA		Q76G00
Mycetozoa	*Dictyostelium discoideum*	CESA	AAF00200.1	

OUTGROUP

Major Lineage	*Organism*	Identifier	GenBank	Uniprot
	Saccharomyces cerevisiae	CHS1 (Chitin Synthase)	AAA34491.1	
	Saprolegnia monoica	CHS2 (Chitin Synthase)	AAC49743.1	
	Aspergillus fumigatus	CHCG (Chitin Synthase)	CAA63928.1	
	Xenopus laevis	DG42 (Chitin Synthase)	AAA49699.1	
	Azorhizobium caulinodans	NodC (Chitin Synthase)	AAB51164.1	
	Streptococcus pyogenes	HAS	AAA17981.1	
	Mus musculus	HAS1	BAA11654.1	
	Homo sapiens	HAS1	AAC50706.1	
	Streptococcus dysgalactiae	HAS	AAB87874.1	
	Rattus norvegicus	GT-21	AAD02464.1	

Table S3 List of organisms and their identifiers used as the HMM profile to search for the CESA accessory proteins in the *C. tuberculosum* transcriptome dataset.

Protein Profile	Major Lineage	*Organism*	Identifier	GenBank Accession Number	Uniprot Accession Number
Companion of cellulose synthase (CC)	**Angiosperm**	*Arabidopsis thaliana*	AtCC1		AEE32122.1
		Arabidopsis thaliana	AtCC2		AAO42869.1
		Arabidopsis thaliana	AtCC3		ABD59052.1
		Arabidopsis thaliana	AtCC4		AAX23904.1
		Oryza sativa	OsCC1	BAS91763.1	
		Oryza sativa	OsCC2	BAS89256.1	
		Oryza sativa	OsCC3	BAS80202.1	
Cellulose-microtubule uncoupling (CMU)[1]	**Angiosperm**	*Arabidopsis thaliana*	AtCMU1	O81629.1	
		Arabidopsis thaliana	AtCMU2		AAQ22597.1
		Arabidopsis thaliana	AtCMU3		AAO64874.1
COBRA (COB) and COBRA like (COBL)	**Angiosperm**	*Arabidopsis thaliana*	AtCOB		Q94KT8-1
		Arabidopsis thaliana	AtCOBL1		Q9SRT7-1
		Arabidopsis thaliana	AtCOBL2		Q8L8Q7-1
		Arabidopsis thaliana	AtCOBL3		Q9LFW3-1
		Arabidopsis thaliana	AtCOBL4		O04500-1
		Arabidopsis thaliana	AtCOBL5		Q8GZ17-1
		Arabidopsis thaliana	AtCOBL6		Q9LIB6-1
		Arabidopsis thaliana	AtCOBL7		Q9FJ13-1
		Arabidopsis thaliana	AtCOBL8		Q9LJU0-1
		Arabidopsis thaliana	AtCOBL9		Q9T045-1
		Arabidopsis thaliana	AtCOBL10		
		Arabidopsis	AtCOBL11		

Protein Profile	Major Lineage	*Organism*	Identifier	GenBank Accession Number	Uniprot Accession Number
		thaliana			
		Zea mays	ZmCOB		ABJ99754.1
		Zea mays			
		Zea mays	ZmCOBL3		ABL59983.1
		Zea mays	ZmCOBL4		ABL59984.1
		Zea mays	ZmCOBL5		ABL59985.1
		Zea mays	ZmCOBL6		ABL59986.1
		Zea mays	ZmCOBL7		AAQ81633.1
		Zea mays			
		Oryza Sativa	OsCOB		AAQ56121.1
Cellulose synthase interactive (CSI)	**Angiosperm**	*Arabidopsis thaliana*	AtCSI		F4IIM1.1
		Oryza sativa	OsCSI		EEC80251.1
Chitinase-like (CTL)	**Angiosperm**	*Arabidopsis thaliana*	AtCTL1		Q9MA41.1
		Arabidopsis thaliana	AtCTL2		Q9LSP9.1
		Brachypodium distachyon	BdCTL		KQJ90816.1
		Hordeum vulgare	HvCTL1		BAK00996.1
		Medicago truncatula	MtCTL		ACJ84914.1
Korrigan (KOR)	**Angiosperm**	*Arabidopsis thaliana*	AtKOR		Q38890
		Brassica napus	BnKOR		CAB51903.1
		Gossypium hirsutum	GhKOR		AAS87601.1
		Oryza sativa	OsKOR		BAF37260.1
		Populus alba	PaKOR		ADB82903.1
		Populus tremuloides	PtKOR		AAS45400.1
		Solanum Lycopersicum	SlKOR		AAC49704.1
STELLO	**Angiosperm**	*Arabidopsis thaliana*	AtSTELLO1		O22943.1
		Arabidopsis thaliana	AtSTELLO2		Q9SCN0.1