The most abundant cyst wall proteins of *Acanthamoeba castellanii* are cellulose-binding lectins from three gene families that localize to distinct structures in cyst walls

Short title: Identification and characterization of *Acanthamoeba* cyst wall proteins

Pamela Magistrado-Coxen[1,#a], Yousuf Aqeel[1], Angelo Lopez[1,2], John R. Haserick[1,2,#b], Breeanna R. Urbanowicz[3], Catherine E. Costello[2], and John Samuelson[1*]

[1] Department of Molecular and Cell Biology, Boston University Goldman School of Dental Medicine, Boston, Massachusetts, United States of America

[2] Department of Biochemistry, Boston University School of Medicine, Boston, Massachusetts, United States of America

[3] Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia, United States of America

[#a] Current Address: Sarepta Therapeutics, Andover, Massachusetts, United States of America

[#b] Current Address: Glyde Bio, Inc., Cambridge, Massachusetts, United States of America

* Corresponding author

E-mail: jsamuels@bu.edu

Abstract

Acanthamoeba castellanii, cause of keratitis and blindness, is an emerging pathogen because of its association with contact lens use. The cyst wall contributes to pathogenesis as cysts are resistant to sterilizing reagents in lens solutions and to antibiotics applied to the eye. Here we used structured illumination microscopy (SIM) and probes for glycopolymers to show that purified cyst walls of *A. castellanii* retain endocyst and ectocyst layers and conical structures (ostioles) that connect them. Mass spectrometry showed candidate cyst wall proteins (CWPs) are dominated by three families of lectins (named here Luke, Leo, and Jonah), because each binds to microcrystalline cellulose +/- chitin. Luke lectins contain two or three carbohydrate-binding modules (CBM49), which were first identified in a tomato cellulase. Leo lectins have two unique domains with eight Cys residues each (8-Cys) +/- a Thr-, Lys-, and His-rich spacer. Jonah lectins contain one or three choice-of-anchor A (CAA) domains previously of unknown function. Representative members of each family were tagged with green fluorescent protein (GFP) and expressed under their own promoters in transfected parasites. A representative Jonah lectin with one CAA domain is made early during encystation and localizes to the ectocyst layer. In contrast, Leo and Luke lectins are made later and localize to the endocyst layer and ostioles. Probes for CWPs (anti-GFP antibodies) and for glycopolymers (maltose-binding protein-fusions with CWPs) suggest Jonah lectin and the glycopolymers to which it binds are accessible in the ectocyst layer, while Luke and Leo lectins and their epitopes are mostly inaccessible in the ectocyst layer and ostioles. In summary, the most abundant *A. castellanii* CWPs are three sets of lectins, which have conserved (CBM49s of Luke), newly characterized (CAA of Jonah), or unique carbohydrate-binding modules (8-Cys of Leo).

Author summary

Fifty years ago, the cyst wall of *Acanthamoeba castellanii* was shown to contain cellulose and have an ectocyst layer, an endocyst layer, and conical ostioles that attach them. The goals here were to identify abundant cyst wall proteins (CWPs) and begin to determine how the wall is assembled. We used wheat germ agglutinin to show cyst walls also contain chitin fibrils. When trophozoites are starved of nutrients, they become immotile and make CWPs and glycopolymers in dozens of small vesicles. The primordial cyst wall is composed of a single, thin layer containing cellulose, chitin, and an abundant CWP we called Jonah. The primordial wall also has small, flat ostioles that contain another abundant CWP we called Luke. Jonah (the best candidate for diagnostic antibodies) is accessible in the ectocyst layer of mature cyst walls, while Luke and a third abundant CWP we termed Leo are present but mostly inaccessible in the endocyst layer and ostioles. While *A. castellanii* cyst walls contain cellulose (like plants) and chitin (like fungi), the glycopolymers are made in vesicles rather than at the plasma membrane, and the CWPs (Luke, Leo, and Jonah lectins) are unique to the protist.

Introduction

Acanthamoebae, which include the genome project *A. castellanii* Neff strain studied here, are soil protists named for acanthopods (spikes) on the surface of trophozoites [1]. In immunocompetent persons, *Acanthamoebae* is a rare but important cause of corneal inflammation (keratitis), which is difficult to treat and so may lead to scarring and blindness [2-5]. In immunosuppressed patients, *Acanthamoebae* may cause granulomatous encephalitis [6]. *Acanthamoeba* is considered to be an emerging pathogen, because 80 to 90% of infections are associated with growing contact lens use and is often transmitted via dirty hands or contaminated lens solutions [7-9]. Water for hand washing is often scarce in places where *Acanthamoeba* is endemic. However, we recently showed that alcohols in concentrations present in hand sanitizers kill *A. castellanii* trophozoites and cysts, providing a possible route to reducing infection [10-12].

Acanthamoebae host numerous pathogenic bacteria and so may contribute to pneumonia (*Legionella pneumophilia*), diarrhea (*Vibrio cholera* and *Campylobacter jejuni*), or disseminated disease (*Listeria monocytogenes*) [13-15]. *Acanthamoebae* also contain enormous double-stranded DNA viruses, which can cause respiratory infections [16, 17]. For 30+ years, *Acanthamoeba* was used as a model system to identify and characterize the role of actin and actin-associated proteins in the cytoskeleton, during phagocytosis, and in cell motility [18, 19]. Finally, the *A. castellanii* genome contains >500 genes derived from bacteria by horizontal gene transfer (HGT), which is the greatest number of any human pathogen [1].

When *A. castellanii* trophozoites are deprived of nutrients in solution or on agar plates, they form cysts [20-22]. Transmission electron microscopy (TEM) shows cyst walls have two

fibril-dense layers (outer ectocyst and an inner endocyst), which are separated by a relatively fibril-free middle layer [23-25]. The endocyst and ectocyst layers are connected to each other by ostioles, which are conical and fibrillar. There is no cell division during encystation and excystation, so a single trophozoite passes through an individual ostiole when it exits the cyst [26]. The cyst wall of *A. castellanii* protects free-living parasites from osmotic shock when exposed to fresh water, from drying when exposed to air, or from starvation when deprived of bacteria or other food sources. Furthermore, the cyst wall also acts as a barrier, sheltering parasites from killing by disinfectants used to clean surfaces, sterilizing agents in contact lens solutions, or antibiotics applied directly to the eye [4, 5, 27-30].

We are interested in the cyst wall proteins (CWPs) of *A. castellanii* for four reasons. 1) Although monoclonal antibodies to *A. castellanii* have been made, the vast majority react to trophozoites, and no CWPs have been molecularly identified [31-35]. Indeed the only cyst-specific protein identified, which was named for its 21-kDa predicted size (CSP21), is unlikely to be a CWP, as it lacks a signal peptide [36, 37]. 2) *A. castellanii* and related species are the only human pathogens that contain cellulose in their wall [38-40]. Whole genome sequences predict that there is a set of proteins with tandemly repeated carbohydrate-binding modules (CBM49) that was previously shown to be a cellulose-binding domain at the C-terminus of the tomato cellulase SlCel9C1 [41-43]. It is possible then that the *A. castellanii* CBM49 domain proteins are lectins that bind cellulose fibrils in the cyst wall [44]. 3) The whole genome of *A. castellanii* predicts a chitin synthase, a chitinase, and two chitin deacetylases, suggesting the possibility that chitin and chitin-binding proteins are present in the cyst wall [1, 45-49]. 4) We wanted to test the idea that structured illumination microscopy (SIM), which samples the entire wall of dozens of cysts without need for sectioning, may be able to replace, at least partially,

immunoelectron microscopy, which samples 100-nm sections of just a few cysts, for localization of proteins in the ectocyst and endocyst layers and ostioles [50].

The experimental design here was simple. We used SIM and probes for glycopolymers to localize structures in developing and mature cyst walls of *A. castellanii*. Specifically, we focused on probes for chitin and cellulose, which are a β-1,4-linked homopolymers of N-acetylglucosamine and D-glucose, respectively, that exist as hydrogen-bonded chains called microbrils. SIM and TEM were used to check the intactness of purified cyst walls, which were treated with trypsin to identify candidate CWPs by mass spectrometry. We chose representative proteins from each of three families of abundant candidate CWPs and checked the timing of their expression, determined their localization and accessibility in the cyst wall (or not), and determined whether each protein binds cellulose and/or chitin. In this way, we began to answer five basic questions concerning *A. castellanii* CWPs: What are their identities? When are they made? Where are CWPs located in the developing and mature cyst wall? Why are CWPs located there? Which CWP is the best target for production of diagnostic antibodies?

Results

SIM of encysting *Acanthamoebae* shows glycopolymers are made in secretory vesicles and the primordial cyst wall is a single, thin layer with small, flat ostioles. To visualize the structures of developing and mature cyst walls of *A. castellanii* by SIM, we used probes that bind chitin (wheat germ agglutinin, WGA) and β-1,3 and β-1,4 polysaccharides (calcofluor white, CFW) in the walls of *Saccharomyces* and *Entamoeba* cysts [51-54]. CFW, a fluorescent brightener, has also been used to diagnose *A. castellanii* cysts in eye infections

[55]. We also made a glutathione-S-transferase (GST) fusion-protein, which contains the N-terminal CBM49 of a candidate CWP of *A. castellanii* (S1 Fig and S1 Excel file) [56]. The GST-AcCBM49 expression construct was designed to replicate that used to determine the carbohydrate binding properties of SlCBM49, which is a C-terminal CBM of the *Solanum lycopersicum* (tomato) cellulase SlGH9C [41].

GST-AcCBM49, WGA, and CFW bind to a small number of organisms in trophozoite cultures that are spontaneously encysting. In contrast, all three probes label dozens of vesicles of organisms induced to encyst by placing them on non-nutrient agar plates for three to six hr (Fig 1A). After 12 hr encystation, WGA-labeled vesicles are still prominent, but those labeled with GST-AcCBM49 are fewer and smaller (Fig 1B). GST-AcCBM49, WGA, and CFW each labels primordial cyst walls, which have a single, thin layer that is closely apposed to the surface of the encysting cell. The primordial wall contains small, flat ostioles, which label with CFW but not with GST-AcCBM49 or WGA. After 36 hr encystation, GST-AcCBM49 predominantly labels the ectocyst layer, while WGA predominantly labels the endocyst layer that is closely apposed to the cell inside (Fig 1C). WGA labels dome-shaped ostioles in a punctate manner, while CFW diffusely binds the endocyst layer and ostioles. After 72 hr encystation, mature cyst walls are labeled in a similar manner, except that WGA labels an epitope that forms a sharp ring around ostioles (Fig 1D).

Conclusions. These results show glycopolymers labeled with WGA and GST-AcCBM49 are made in secretory vesicles of encysting *A. castellanii*. The primordial cyst wall is single-layered and has the beginnings of ostioles (stained by CFW), which are small and flat but are present in the same number and are distributed in a similar fashion to that of mature

ostioles. The ectocyst layer of mature cyst walls is primarily labeled by GST-AcCBM49, while the endocyst layer and conical ostioles primarily label with WGA and stain with CFW.

SIM and TEM show purified *A. castellanii* cyst walls preserve distinct endocyst and ectocyst layers, as well as ostioles. We used SIM and TEM to check the integrity and cleanliness of cyst wall preparations, which were made by sonicating cysts and separating walls from cellular contents by density centrifugation and retention on a membrane with 8 micron pores. SIM shows that in purified cyst walls, like in intact cyst walls, GST-AcCBM49 predominantly labels the ectocyst layer, WGA highlights the ostioles, and CFW stains the endocyst layer (Fig 1E).

For TEM, intact cysts and purified cyst walls were frozen under high pressure, and fixatives were infiltrated at low temperature [57]. Walls of intact cysts and purified cyst walls each have an ectocyst layer and an endocyst layer, as well as conical ostioles that link the two layers (Fig 2A) [23-25]. At higher magnification, bundles of fibrils are apparent in the endocyst layer, which is thicker (~300 nm) and less electron-dense than the ectocyst layer (~100 nm) (Fig 2C). The space between the endocyst and ectocyst layers, which is often wider in purified walls, is electron-lucent with a few thin fibrils. The purified walls are missing amorphous material that fills the space between the inner aspect of the cyst wall and the plasma membrane of the single trophozoite inside (Figs 2B and 2D). At the edge of the ostiole, the endocyst layer splits into an external face that meets the ectocyst layer (Fig 2E). In the center of the ostiole, the ectocyst layer forms a narrow cap over the internal face of the endocyst layer, which splits again.

Conclusions. SIM and TEM both show that ectocyst and endocyst layers, as well as ostioles, are preserved in purified cyst walls, which are relatively free of cellular material. The purified walls were then used to identify candidate CWPs by mass spectrometry.

Mass spectrometry shows candidate CWPs of *A. castellanii* are encoded by multigene families and contain tandem repeats of short domains (CBM49, 8-Cys, and CAA). Trypsin treatment of purified *A. castellanii* cyst walls followed by liquid chromatography mass spectrometry (LC-MS/MS) analysis of the released peptides gave similar results across five biological experiments (Table 1 and S1 Excel file). While some proteins remain in cyst walls after trypsin digestion, their identities are similar to those detected in the soluble fractions by in gel-digests with trypsin or chymotrypsin. Candidate CWPs with the most unique peptides identified by LC-MS/MS belong to three families, which we named Luke, Leo, and Jonah lectins, because each binds cellulose +/- chitin (see below). Although it is impossible to draw a line that separates actual CWPs from contaminating proteins, secreted proteins with 18+ unique peptides include six Leo lectins, four Luke lectins, and three Jonah lectins. The vast majority of proteins with <18 unique peptides are predicted to be cytosolic (including the 21-kDa cyst-specific protein (CSP21) localized below) and so are likely intracellular contaminants of cyst wall preparations. The exception to this hypothesis, we think, are additional Luke, Leo, and Jonah lectins, which are most likely less abundant CWPs. For readers interested in cytosolic proteins of *A. castellanii*, we have added S2 Excel file, which contains all the mass spectrometry data including a “dirty” cyst wall preparation that was generated without using a Percoll gradient or porous filter.

188

Table 1. Candidate CWPs identified by mass spectrometry

	ID	# Unique peptides	Coverage (%)	Mass (kDa)
Jonah				
three CAAs				
	ACA1_157320	147	38	146
one CAA				
	ACA1_164810	83	56	58
	ACA1_261530	18	23	55
	ACA1_133400	9	24	44
	ACA1_377440	6	11	47
Luke				
three CBM49s				
	ACA1_245650	72	74	44
	ACA1_160160	8	25	43
	ACA1_187760	7	25	42
	ACA1_252830	6	19	44
	ACA1_031530	6	21	43
	ACA1_253650	5	20	42
	ACA1_253500	5	19	42
	ACA1_061050	3	12	43
	ACA1_287530	2	14	43
two CBM49s				
	ACA1_377670	78	68	29
	ACA1_096300	47	77	28
	ACA1_246110	22	70	27
Leo				
two 8-Cys domains				
	ACA1_074730	34	82	20
	ACA1_351320	24	44	20
	ACA1_394030	24	44	20
	ACA1_394280	24	36	24
	ACA1_083920	19	68	20
	ACA1_394560	1	10	19
two 8-Cys + TKH-rich spacer				
	ACA1_188350	21	20	59
	ACA1_374130	7	20	52
	ACA1_188550	7	15	46
	ACA1_188370	6	9	68
	ACA1_116240	5	18	56
	ACA1_365840	3	18	44
	ACA1_117050	3	33	36
	ACA1_096640	2	27	37

Luke lectins are comprised of an N-terminal signal peptide, followed by two (Luke(2)) or three CBM49s (Luke(3)) separated by Ser- and Pro-rich spacers (Fig 3 and S2 Fig) [37, 42, 43, 58]. The N-terminal CBM49 of Luke lectins contains three conserved Trp resides conserved in Sl CBM49 from tomato [41]. A fourth conserved Trp is present in the CBM49 of *D. discoideum* cellulose-binding proteins [43]. The other CBM49s (middle and/or C-terminal) of Luke lectins have two conserved Trp residues. Luke lectins are acidic (pI 5 to 6) and have formula weights (FWs) from 27 to 29-kDa (Luke(2)) or 42 to 44-kDa (Luke(3)). There are no predicted transmembrane helices or GPI-anchors in the Luke or Leo lectins [59, 60]. LC-MS/MS of the released cell wall peptides identified at least one unique peptide corresponding to all 12 genes encoding Luke lectins, although the number of unique peptides varied from 78 to two (Table 1 and S1 Excel file). In general, Luke lectins with two CBM49s have more unique peptides than Luke lectins with three CBM49s. One to four unique peptides were derived from three CBM49-metalloprotease fusion-proteins, which consist of an N-terminal signal peptide followed by a single CBM49 with four conserved Trp residues and a metalloprotease (ADAM/reprolysin subtype) with a conserved catalytic domain (HEIGHNLGGNH) [58]. We used the same Luke(2) lectin (ACA1_377670) to perform RT-PCR, make rabbit anti-peptide antibodies, and make maltose-binding protein (MBP)- and green fluorescent protein (GFP)-fusions (S1 Fig) [18, 56, 61-63].]. In addition, we used Luke(3) lectin (ACA1_245650) to make a GFP-fusion (S2 Fig).

Leo lectins have an N-terminal signal peptide followed by two repeats of a unique 8-Cys domain, some of which are separated by a long Thr-, Lys-, and His-rich spacer (Fig 3 and S2

Fig). Leo lectins without a spacer are acidic (pI ~4.8) and have FWs from 19 to 24-kDa, while Leo lectins with the TKH-rich spacer (Leo(TKH)) are basic (pI ~8.3) and have FWs from 36- to 59-kDa. Leo lectins are encoded by 16 genes, of which 14 proteins were identified by our LC-MS/MS analysis. While the number of unique peptides varies from 34 to one, Leo lectins without the spacer generally have more unique peptides than Leo(TKH)s. We used a Leo lectin with no spacer (ACA1_074730) to perform RT-PCR, make rabbit anti-peptide antibodies, and make MBP- and GFP-fusions (S1 Fig).

Jonah lectins have an N-terminal signal peptide followed by one or three choice-of-anchor A (CAA) domains (Fig 3 and S2 Fig) [58]. The binding activity of the CAA domain, which is adjacent to a collagen-binding domain in a microbial surface component recognizing the adhesive matrix molecule (MSRAMM) of *Bacillus anthracis*, is not known [64]. Jonah(1) lectins with a single CAA domain are acidic (pI ~6), have a FW from 44 to 58-kDa and have an N-terminal Thr-, Lys-, and Cys-rich domain. Jonah(3) lectins with three CAA domains are basic (pI ~8.8), have a FW of ~146-kDa, and contain Ser- and Pro-rich spacers between CAA domains, as well as and hydrophobic regions that may be transmembrane helices [60]. Jonah lectins are encoded by eight genes, of which five can be identified by our LC-MS/MS analysis based on one to 147 unique peptides. We used a Jonah(1) lectin (ACA1_164810) with a single CAA domain to perform RT-PCR, make rabbit anti-peptide antibodies, and make MBP- and GFP-fusions (S1 Fig).

Other secreted proteins with 18+ unique peptides detected by LC-MS/MS, which are candidate CWPs, include a laccase with three copper oxidase domains (ACA1_068450), a protein with a C-terminal ferritin-like domain (ACA1_292810), a Kazal-type serine protease

inhibitor (ACA1_291590), a conserved uncharacterized protein (ACA1_068630), and a protein unique to *A. castellanii* (ACA1_145900) [58, 65-67].

Conclusions. These results suggest that the most abundant candidate CWPs of *A. castellanii* contain tandem repeats of conserved domains (CBM49 in Luke lectins and CAA in Jonah lectins) or a unique domain (8-Cys in Leo lectins). Peptides corresponding to nearly all members of each gene family were detected by mass spectrometry, although the relative abundances of unique peptides for each CWP vary by more than an order of magnitude, suggesting marked differences in gene expression. Because we did not (and cannot) separate cyst walls into component parts (endocyst and ectocyst layers and ostioles) prior to LC-MS/MS analysis of tryptic peptides, we used SIM and GFP-tags to localize one or two representative members of each family of CWPs in cyst walls of transfected *A. castellanii* (see below).

Origins and diversity of genes that encode Luke, Leo, and Jonah lectins. Leo lectins, which have two domains with 8-Cys each, appear to be unique to *A. castellanii*, as no homologs were identified when BLAST analysis was performed using the nonredundant (NR) database at NCBI (https://www.ncbi.nlm.nih.gov/). The origin of genes encoding Luke lectins is difficult to infer, because its CBM49s show only a 31% identity over a short (77-amino acid) overlap with a predicted cellulose-binding protein of *D. discoideum* (expect value of BLASTP is just 7e-05) [45]. In contrast, the CAA domain of Jonah lectins appears to derive from bacteria by horizontal gene transfer (HGT), as no other eukaryote contains CAA domains, and there is a 28% identity over a bigger (263-aa) overlap with a choice-of anchor A family protein of *Saccharibacillus sp. O16* (5e-12). The *A. castellanii* laccase (also known as copper oxidase), whose signals were abundant in the mass spectra, is likely the product of HGT, as there is a 44% identity over a large (526-aa) overlap with a copper oxidase of *Caldicobacteri oshimai*

(6e-135). The uncertainty is based upon the presence of similar enzymes in plants, one of which (*Ziziphus jujube*) shows a 39% identity over a 484-aa overlap (4e-101) with the *A. castellanii* laccase.

No pairs of genes within each lectin family are syntenic, suggesting that individual genes rather than clusters of genes were duplicated. With the exception of two Luke lectins (ACA1_253500 and ACA1_253650) that are 98% identical and two Leo lectins (ACA1_074770 and ACA1_083920) that are 85% identical, members of each family of CWPs differ in amino acid sequence by >40%. Genes that encode CWPs also vary in the number of introns (zero to two in Luke, two to four in Leo, and zero to 24 in Jonah). Searches of genomic sequences of 11 strains of *Acanthamoebae*, deposited in AmoebaDB without protein predictions by Andrew Jackson of the University of Liverpool, using TBLASTN and sequences of representative Luke, Leo, and Jonah lectins localized in the next section showed four results [47]. First, although stop codons are difficult to identify using this method, all 11 strains appear to encode each CWP. Second, most strains show 100 to 200-amino acid stretches of each CWP that are 80 to 90% identical to the *A. castellanii* Neff strain studied here. These stretches exclude low complexity spacers that are difficult to align. Third, some of the strains show greater differences from the Neff strain in each CWP, consistent with previous descriptions of *Acanthamoeba* strain diversity based upon 18S rDNA sequences [68]. Fourth, while coding sequences and 5' UTRs are well conserved, introns are very poorly conserved, with the exception of branch-point sequences.

Conclusions. Genes encoding Jonah lectin and laccase likely derive by HGT, while genes encoding Leo lectins appear to originate within *Acanthamoeba*. Although CBM49s of Luke lectins share common ancestry with plants and other Amoebazoa, their precise origin is

not clear. For the most part, gene duplications that expanded each family within the *Acanthamoeba* genome occurred a long time ago, as shown by big differences in amino acid sequences of paralogous proteins and variations in the number of introns. However, the set of Luke, Leo, and Jonah lectins identified by mass spectrometry and the sequence of representative CWPs localized in the next section both appear to be conserved among 11 sequenced isolates. This result suggests antibodies to a particular CWP in the *A. castellanii* Neff strain would likely recognize the same protein in other strains of *Acanthamoebae*.

SIM shows a representative Jonah lectin is made early and is present in the ectocyst layer of mature cyst walls, while representative Luke and Leo lectins are made later and are present in the endocyst layer and ostioles. To localize candidate CWPs, we expressed representative Luke(2), Leo, and Jonah(1) lectins each with a GFP-tag under its own promoter (446-, 486- and 571-bp of the 5'UTR, respectively) in transfected trophozoites of *A. castellanii*, using an episomal vector that was selected with G418 (S1 Fig and S3 Excel file) [62, 63]. GFP-tagged candidate CWPs expressed under their own promoter are absent in the vast majority of log-phase trophozoites, which divide but do not become cysts (data not shown). GFP-tagged CWPs are present in small numbers in trophozoites in stationary cultures, where some organisms have begun to encyst spontaneously. In contrast, the vast majority of mature cysts express GFP-tagged CWPs in their walls, although the strength of the signal varies among the lectins assayed (S3 Fig).

After three to six hr encystation, Jonah(1)-GFP (single CAA domain) is present in hundreds of small vesicles, which are distinct from the WGA-labeled vesicles (Fig 4A). At 15 hr, Jonah(1)-GFP is present in a fibrillar pattern in the single layer of the primordial cyst wall (Fig 4B). CFW-labeled ostioles are not visible with Jonah(1)-GFP or WGA. At 24 hr, Jonah(1)-

GFP is present in the ectocyst layer, which is its final destination in the mature cyst wall (Figs 4C and 4D). At this time, the endocyst layer labeled by WGA and CFW has separated from the ectocyst layer.

Luke(2)-GFP (two CBM49s) and Leo-GFP (two 8-Cys domains) are both made later than Jonah-GFP. Indeed vesicles containing these two lectins are too small and too few in number to be visualized easily with SIM even after 12 hr encystation. At 15 hr, Luke(2)-GFP forms a sharp outline around the edges of a subset of ostioles, which are small and flat (Fig 5A). At 18 and 24 hr, Luke(2)-GFP expands across the endocyst layer and ostioles, which has begun to separate from the ectocyst layer (Figs 5B and 5C). In mature cysts, Luke(2)-GFP forms dense rings around the ostioles (Fig 5D). The distribution of Leo-GFP is different from that of Luke(2)-GFP in a number of ways. At 15 hr, Leo-GFP is present in abundant small vesicles, some of which have reached the single-layered wall (Fig 6A). At 18 hr, Leo-GFP forms patches on the cyst wall, which are for the most part independent of ostioles (Fig 6B). Only at 36 hr does Leo-GFP accumulate at the edges of the ostioles, which is its location in mature cysts (Figs 6C and 6D). We counted an average of 8.8 +/- 2.5 ostioles per cyst wall by rotating SIM images of cysts expressing Luke-GFP and Leo-GFP or labeled with WGA, MBP-Luke, or MBP-Leo (24 cysts total).

Conclusions. These results show that ~500 bp of the 5’ UTR is sufficient to cause encystation-specific expression of GFP-tagged representatives of Luke, Leo, and Jonah lectins. Jonah(1)-GFP is made early and is present in the single-layered, primordial cyst wall and the ectocyst layer of mature cyst walls. Luke(2)-GFP, which is made later, is present in some of the small, flat ostioles in the primordial cyst wall and in the endocyst layer and conical ostioles of mature cyst walls. Leo-GFP, which is also made later, is absent from small, flat

ostioles, but is present in endocyst layer and ostioles of mature cyst walls. This is the first estimate of the number of ostioles in *Acanthamoeba* cyst walls, because ostioles have not previously been visualized by light microscopy and are impossible to count by TEM, as the 100-nm thick sections poorly sample the ostioles (Fig 2).

Control experiments suggest that the timing and locations of GFP-tagged constructs in mature cyst walls are accurate. RT-PCR showed that mRNAs of representative Luke, Leo, and Jonah lectins, as well as cellulose synthase (ACA1_349650), are absent or nearly absent from trophozoites but are present during the first three days of encystation (S4 Fig). These results are consistent with encystation-specific expression of GFP-tagged CWPs under their own promoter (Figs 4-6). In contrast, glyceraldehyde 3-phosphate dehydrogenase (GAPDH), which catalyzes the sixth step in glycolysis, is expressed in both trophozoites and encysting *A. castellanii* [62]. Monospecific, polyclonal rabbit antibodies to a 50-amino acid peptide of a representative Jonah lectin and a 16-amino acid peptide of a representative Leo lectin bind to Western blots of proteins from cysts but not trophozoites (S5 Fig). (Rabbit antibodies to a 50-amino acid peptide of Luke(2) lectin did not react to trophozoites or cysts). Although these rabbit anti-peptide antibodies do not work for localization of CWPs by SIM, they confirm the encystation-specific expression of the GFP-tagged proteins under their own promoter.

An abundant Luke lectin with three CBM49s tagged with GFP (Luke(3)-GFP) (ACA1_245650) and expressed under its own promoter is present in the endocyst layer and ostioles of mature cyst walls, which is the same place where Luke(2)-GFP with two CBM49s is located (Figs 3 and 5 and S2 Fig). In contrast, GFP-tagged cyst-specific protein CSP21 expressed under its own promoter is present in cytosolic accumulations of mature cysts (Fig 6)

[36, 69]. Under a constitutive GAPDH promoter, Luke(2)-GFP and Jonah(1)-GFP are both present in secretory vesicles of trophozoites, which do not change their “acanthamoeboid” appearance when observed by differential-inference contrast microscopy (S6 Fig) (Leo-GFP did not express well under the GAPDH promoter). In addition, when Jonah(1)-GFP was expressed using the GAPDH promoter, it localizes to the ectocyst layer of mature cyst walls, while Luke(2)-GFP localizes to both the endocyst layer and ostioles (Figs 4 and 5). These are the same locations where Jonah(1)-GFP and Luke(2)-GFP go when expressed under their own promoter. In contrast, when GFP is expressed alone under control of the GAPDH promoter, it remains localized in the cytosol of trophozoites and cysts (S6 Fig). A GFP-fusion protein appended with an N-terminal signal peptide from Luke(2) lectin, also expressed under a GAPDH promoter, localizes to secretory vesicles of trophozoites and cysts but not to cyst walls (Fig 6). Finally, while there are issues of accessibility, MBP-Jonah(1), MBP-Luke(2), and MBP-Leo bind to the same structures in the mature cyst wall, where GFP-tagged CWPs localize when expressed under their own promoter (see next section).

Conclusions. RT-PCR and Western blots with anti-CWP antibodies confirm the encystation-specific expression of GFP-tagged CWPs under their own promoters. The number of CBM49s does not appear to affect localization of Luke lectins, as Luke(2)-GFP and Luke(3)-GFP each localize to endocyst layer and ostioles. Because localizations of Luke(2)-GFP and Jonah(1)-GFP are the same under their own or GAPDH promoters, it appears that intrinsic properties of each CWP (e.g. carbohydrate-binding specificity tested below) rather than timing of expression determine its location in the cyst wall. As CSP21 is homologous to universal stress proteins and lacks an N-terminal signal peptide, its presence in the cytosol after nutrient deprivation was expected [69, 70].

Jonah lectin and glycopolymers are readily accessible in ectocyst layer of mature cyst walls, while Luke and Leo lectins and glycopolymers are poorly accessible in the endocyst layer and ostioles. To determine the accessibility of glycopolymers in the two layers of the cyst wall and ostioles, we fused each CWP to MBP and recombinantly expressed them in the periplasm of *E. coli*. Previously, we have used MBP-fusions to express a chitin-binding domain of an *Entamoeba* Jessie lectin, a β-1,3-linked GalNAc polymer-binding domain of *Giardia* cyst wall protein 1, and a β-1,3-glucan-binding domain of a *Toxoplasma* glucanase [71-73]. MBP-Jonah(1) labels small vesicles and the primordial cyst wall of encysting organisms (Fig 7). MBP-Jonah(1) also binds to the ectocyst layer of mature cell walls, which is the same location as Jonah(1)-GFP expressed under its own promoter (Fig 4). Because nearly 100% of the organisms are labeled with MBP-Jonah(1), glycopolymer(s) bound by Jonah lectins must not be completely covered by CWPs (S7 Fig). MBP-Luke(2) and MBP-Leo also label small vesicles and the primordial cyst wall of encysting organisms, but these probes label the endocyst layer and ostioles of <10% mature cyst walls (Fig 7 and S7 Fig). Although these are the same places where Luke(2)-GFP and Leo-GFP localize under their own promoter (Figs 5 and 6), these results suggest that glycopolymers bound by Luke and Leo lectins in the endocyst layer and ostioles are, for the most part, inaccessible to external probes. In a parallel fashion, anti-GFP antibodies show that Jonah(2)-GFP is accessible in the endocyst layer of nearly 100% of mature cysts with a detectable Jonah-GFP signal (Fig 4 and S3 Fig). In contrast, anti-GFP antibodies show Luke(2)-GFP and Leo-GFP are accessible in the endocyst layer and ostioles of 3 and 2%, respectively, of cysts with detectable GFP signals.

Conclusions. Jonah(1)-GFP and glycopolymer(s) bound by MBP-Jonah(1) are accessible in the ectocyst layer of mature cyst walls, while Luke(2)-GFP and Leo-GFP and

glycopolymers bound by MBP-Luke(2) and MBP-Leo are, for the most part, inaccessible in the endocyst layer and ostioles. It appears the CWPs and glycopolymers in the ectocyst layer block access of external probes to the endocyst layer and ostioles. Alternatively, CWPs and glycopolymers in the endocyst layer and ostioles are so tightly packed that they are inaccessible to external probes.

Luke, Leo, and Jonah lectins all bind microcrystalline cellulose, while binding of CWPs to chitin beads is variable. To test the binding of representative CWPs to commercially available glycopolymers, we isolated GFP alone and GFP-labeled Luke and Jonah lectins from lysed trophozoites that expressed each fusion-protein under the GAPDH promoter (S1 and S6 Figs). We also performed parallel experiments using MBP-CWP fusion-proteins recombinantly expressed in *E. coli*. The targets were microcrystalline cellulose (used to characterize binding activities of GST-SlCBM49 from tomato cellulase) and chitin beads (used to characterize myc-tagged Jacob and Jessie lectins of *E. histolytica*) [41, 74].

Western blots anti-GFP and anti-MBP antibodies show both Luke(2)-GFP and MBP-Luke(2), each of which is partially degraded, bind to microcrystalline cellulose and somewhat less well to chitin beads (Fig 8). Jonah(1)-GFP and MBP-Jonah(1) also bind to microcrystalline cellulose, while binding to chitin beads is mixed (Jonah(1)-GFP not at all and MBP-Jonah(1) to some degree). MBP-Leo binds less well to microcrystalline cellulose and weakly at best to chitin beads. GFP alone and MBP alone (negative controls) do not bind to microcrystalline cellulose or chitin beads.

Conclusions. Luke(2) and Jonah(1) lectins bind cellulose well, while Leo lectin binds cellulose less well. Luke(2) lectin also binds well to chitin beads, while binding of Leo and Jonah(1) lectins is much less to chitin beads.

Discussion

Numerous advantages of studying *A. castellanii* cyst walls. *A. castellanii* has a number of properties that make it possible to quickly identify and begin to characterize its CWPs. The protist grows axenically in medium without serum or vitamins and synchronously encysts when placed on non-nutrient agar plates [20, 21]. Cyst walls separate cleanly from cellular contents without losing essential structures (ectocyst and endocyst layers and ostioles), which we show can be readily visualized with SIM using probes for wall glycopolymers (GST-AcCBM49, WGA, and CFW). The most abundant CWPs belong to three protein families that contain tandem repeats of CBM49s (Luke lectins), 8-Cys domains (Leo lectins), or CAA domains (Jonah lectins). Although CRISPR/Cas9 methods are not yet available, stable expression of GFP-tagged proteins under their own promoters allows localization of candidate CWPs during cyst wall development using SIM [62, 63]. Expression of GFP-tagged proteins under a constitutive promoter in trophozoites or expression of recombinant proteins in the periplasm (MBP-fusions) of bacteria makes it possible to assay the ability of CWPs to bind microcrystalline cellulose or chitin beads by techniques that have been previously used to characterize other protist cyst wall lectins and carbohydrate-binding modules [41, 61, 74].

The ease of studying cyst walls of *A. castellanii* is in contrast to studies of cyst walls of *E. histolytica*, which do not encyst *in vitro* and so are modeled by cysts of the reptilian pathogen *E. invadens* [53, 71]. While *Giardia* cysts made in mice excyst well, cysts made *in vitro* do not [72]. *Toxoplasma* forms walled oocysts in cats, while *Cryptosporidium* forms oocysts in cows, mice, or gnotobiotic pigs [73, 75, 76]. Oomycetes (water molds that cause potato blight) also have cellulose and chitin in their walls, but the amount of chitin is very small,

and β-1,3-glucan is also present [77]. The spore coat of *Dictyostelium discoideum*, a cousin of *A. castellanii*, contains cellulose and a heteropolysaccharide of galactose and GalNAc, but it lacks chitin [78, 79].

Familiarity and novelty in *A. castellanii* CWPs. While *Giardia* has three copies of a unique cyst wall protein with Leu-rich repeats (CWP1 to CWP3) and *Entamoeba* has two copies each of unique Cys-rich, cyst wall proteins (Jacob and Jessie lectins), *A. castellanii* has many copies each of its three CWPs [71, 72, 80]. although we expected Luke lectins with two or three CBM49s would be present in the cellulose-rich cyst wall, we could not have predicted the other abundant CWPs, because the 8-Cys domains of Leo lectins are unique to *Acanthamoebae* and the CAA domain(s) of Jonah lectins were previously uncharacterized [41, 64].

While Luke lectins have two or three CBM49s, *D. discoideum* has dozens of proteins with a single CBM49 (S8 Fig) [42, 43, 81]. In addition, there is a rare uncharacterized protein with three CBM49s. The Luke(2) lectin binds cellulose and chitin, while the *D. discoideum* proteins with a single CBM49 bind cellulose. Chitin-binding was not tested DdCBM49 or SlCBM49 because this polysaccharide is not present in *D. discoideum* and tomato walls. Demonstration that CBM49s of the Luke(2) lectin likely also bind chitin fibrils is new, but is consistent with recent studies showing CBMs may bind more than one glycopolymer [82]. The metalloprotease fused to an N-terminal CBM49 of *A. castellanii* is absent in *D. discoideum*, while *D. discoideum* adds two CBM49s to a cysteine proteinase, which lacks these domains in *A. castellanii* [58]. The CBM49 may act to localize the metalloproteases to the *A. castellanii* cyst wall, as is the case for the chitin-binding domain in *Entamoeba* chitinases or glucan-binding domain in *Toxoplasma* glucanases [71, 73, 74]. Alternatively, the CBM49 may suggest

the metalloprotease is specific for glycopeptides rather than peptides. While the GH5 glycoside hydrolases of *A. castellanii* lack CBM49 domains, CBM49 is present at the C-terminus of GH9 glycoside hydrolases of *D. discoideum* and *S. lycopersicum* (tomato) [41].

Even though *A. castellanii* Leo lectins and *E. histolytica* Jacob lectins share no common ancestry, they have 8-Cys and 6-Cys lectin domains, respectively, often separated by low complexity sequences (S9 Fig) [74, 80]. *E. histolytica* low complexity sequences vary from strain to strain, contain cryptic sites for cysteine proteases, and are extensively decorated with *O*-phosphate-linked glycans [83]. We have not yet identified any Asn-linked or *O*-linked glycans on Leo lectins or any of the other CWPs. *A. castellanii* and oomycetes (*Pyromyces* and *Neocallmistic*) each contain proteins with arrays of CAA domains, but the sequences of the CAAs are so different that it is likely that concatenation of domains occurred independently (S10 Fig) [43, 58]. Although *A. castellanii* is exposed to collagen in the extracellular matrix of the cornea, the protist lacks a homolog of the collagen-binding domain that is adjacent to the CAA domain in the *Bacillus anthracis* collagen-binding protein [64].

Concatenation of carbohydrate-binding domains in Luke, Leo, and Jonah lectins, which has previously been shown in WGA, Jacob lectins of *E. histolytica*, and peritrophins of insects, most likely increases the avidity of the lectins for glycopolymers [53, 74, 80, 83, 84]. The large number of genes encoding Luke, Leo, and Jonah lectins may be necessary to increase the numbers of CWPs coating glycopolymers in the cyst wall. We showed that Luke(2)-GFP and Luke(3)-GFP localize to the same place in mature cyst walls. We did not test Leo(TKC) lectins with a long spacer or Jonah(3) lectins with three tandem repeats of CAA domains. Finally, other candidate CWPs, which are abundant but present at lower copy numbers (e.g. laccase or ferritin-domain protein), may have important functions in the cyst wall [65-67].

Insights into the structure and assembly of the *A. castellanii* cyst wall. Assembly of the cyst wall, which is comprised of two layers that are connected by ostioles, is a complicated process. Chitin (detected with WGA) and cellulose (detected with MBP-Luke, MBP-Leo, and/or MBP-Jonah) are both made in vesicles. Similarly, chitin is made in vesicles of encysting *Entamoeba*, and a β-1,3-linked GalNAc polymer is made in vesicles of encysting *Giardia* [71, 72]. In contrast, fungi make chitin and β-1,3-glucan at the plasma membrane, and cellulose synthesis in plants takes place at the plasma membrane [85, 86]. Because Luke, Leo, and Jonah lectins bind cellulose but also bind chitin to varying degrees, definitive localization of cellulose and chitin in vesicles of encysting organisms will depend upon localization of cellulose and chitin synthases, each of which is encoded by a single gene *A. castellanii* [45, 47, 58].

Small, flat ostioles in the single-layered primordial cyst wall label beautifully with CFW, but not with probes for chitin and cellulose, so the identity of the glycopolymer in early ostioles remains unknown. CFW has been shown to bind several β-1,3 and β-1,4 polysaccharides, including cellulose, chitin, mixed linkage glucans and galactoglucomannan [85]. Therefore, our data suggests that *A. castellanii* may be producing a previously uncharacterized CFW-binding polysaccharide in the ostioles. Alternatively, the fibrillar form of the polysaccharide in the ostioles is not recognized by the cellulose- or chitin-binding probes. Because GFP-tagged CWPs expressed under the GAPDH promoter are absent from the surface of trophozoites and ostioles label with CFW prior to being coated with Luke(2)-GFP under its own promoter, it appears that CWPs are binding to glycopolymers rather than vice versa. What is directing glycopolymers to form ostioles and later endocyst and ectocyst layers (e.g. Rho GTPases that

direct chitin synthases in bud necks and septa of *Saccharomyces*) is of great interest but was not determined [86, 87].

Because timing of expression does not affect localization of the CWPs tested here, intrinsic properties of each lectin appear to determine its localization. Binding studies with microcrystalline cellulose and chitin beads show CWPs bind best to cellulose but also bind chitin to varying degrees. These binding studies with commercial substrates do not explain why Jonah binds to glycopolymers in the ectocyst layer, while Luke and Jonah lectins bind to glycopolymers in the endocyst layer and ostioles. It is likely that each lectin recognizes specific fibrillar forms of cellulose and/or chitin, which remain uncharacterized. There may also be protein-protein interactions and/or lectin-glycoprotein interactions. As an example of protein-protein interactions, an *E. histolytica* Jessie lectin has a chitin-binding domain and a self-agglutinating “daub” domain, which makes cyst walls impermeable to probes as small as phalloidin (789-Da) [71]. As an example of lectin-glycan interactions, the Gal/GalNAc lectin on the plasma membrane of *E. histolytica* binds to glycans on Jacob lectins, and these in turn bind to chitin fibrils in the cyst wall [53].

Cellulose and chitin are likely present in both layers of the cyst wall and in ostioles, based upon the following evidence. While the chitin-binding lectin WGA strongly labels the endocyst layer and ostioles of mature cyst walls, WGA often labels the ectocyst layer, as well. Cellulose-binding CWPs are abundant in the ectocyst (Jonah(1)-GFP) and endocyst layer and the ostioles (Luke(2)-GFP and Leo-GFP). Jonah(1)-GFP and glycopolymers bound by MBP-Jonah(1) are both accessible in the ectocyst layer. In contrast, Luke(2)-GFP and Leo-GFP and glycopolymers bound by MBP-Luke(2) and MBP-Leo are, for the most part, inaccessible in the endocyst layer and the ostioles. These results suggest the ectocyst layer may be accessible to

bacterial glycoside hydrolases and proteases, while the endocyst layer and ostioles are not. Glycoproteins in the walls of *Giardia, Eimeria,* and *Saccharomyces* protect glycopolymers from exogenous or endogenous glycoside hydrolases [71-73, 86]. How *A. castellanii* glycoside hydrolases, which belong to CAZy families (www.cazy.org) GH1, GH5, GH8, GH10, and GH18, are involved in wall formation during encystation and enzyme-catalyzed destruction and/or reorganization of the wall during excystation is of great interest but beyond the scope of these studies [42, 43, 45, 47].

Implications for diagnostics and therapeutics. While the focus here was on the biochemistry and cell biology of the cyst wall, an abundant and accessible Jonah(1) lectin in the ectocyst layer appears to be an excellent target for diagnostic antibodies for *A. castellanii* cysts [31, 32, 34]. We expect each antibody would react with a single Jonah lectin, as paralogous CWPs, with a couple of exceptions, share less than 40% amino acid identities. *Giardia* CWP1 and *Entamoeba* Jacob2 lectin, which are abundant and accessible in cyst walls, are targets for diagnostic antibodies [88, 89]. In contrast, Luke(2) and Leo lectins are not accessible in the endocyst layer and ostioles, and so these abundant CWPs do not appear to be good targets for diagnostic antibodies. Evidence for chitin in the cyst wall, admittedly not extensively developed here, suggests the possibility that chitin synthase inhibitors might be used as therapeutics [90]. Preliminary studies have explored the possibility of cellulose synthase inhibitors as therapeutics versus *A. castellanii* cysts [91-94]. The SIM methods, which worked so well here to study *A. castellanii* cyst walls, might be used to study the fine structure and assembly of other protist walls. Finally, insights from studies of the development of the *A. castellanii* cyst wall, which is relatively simple, may inform studies of fungal and plant walls, which are highly complex.

Materials and methods.

Ethics Statement. Culture and manipulation of *A. castellanii* have been approved by the Boston University Institutional Biosafety Committee.

Culture of trophozoites and preparation of encysting organisms and cysts. *A. castellanii* Neff strain trophozoites were purchased from the American Type Culture collection. Trophozoites of *A. castellanii* MEEI 0184 strain, which was derived from a human corneal infection, were obtained from Dr. Noorjahan Panjwani of Tufts University School of Medicine [12]. Trophozoites were grown in T-75 tissue culture flasks at 30°C in 10 ml ATCC medium 712 (PYG plus additives), which contains 2% proteose peptone, 0.1% yeast extract, 0.1 M glucose, 4 mM $MgSO_4$, 0.4 mM $CaCl_2$, 3.4 mM sodium citrate, 0.05 mM $Fe(NH_4)_2(SO_4)_2$, 2.5 mM Na_2HPO_4, and 2.5 mM KH_2PO_4, pH 6.5 (Sigma-Aldrich Corporation, St. Louis, MO) [21]. Flasks containing log-phase trophozoites (free of cysts that form spontaneously in stationary cultures) were either chilled or scraped with a cell scraper to release adherent amoebae, which were concentrated by centrifugation at 500 x g for 5 min and washed twice with phosphate buffered saline (PBS). Approximately 10^7 amoebae obtained from a confluent flask were induced to encyst by incubation at 30ºC on non-nutrient agar plates [20]. After 6, 12, 15, 18, 24, 36, or 72 hr incubation, 15 ml of PBS were added to agar plates, which were incubated on a shaker for 30 min at room temperature (RT). Encysting organisms were removed using a cell scraper and concentrated by centrifugation at 1,500 x g for 10 min.

Preparation of mature cyst walls for SIM, TEM, and mass spectrometry. After 3 to 10 days incubation on non-agar plates, mature cysts were washed in PBS and suspended in lysis buffer (10 mM HEPES, 25 mM KCl, 1 mM $CaCl_2$, 10 mM $MgCl_2$, 2% CHAPS, and 1X

Roche protease inhibitor) (Sigma-Aldrich). For SIM, cysts in 500-µl lysis buffer were broken four times for 2 min each with 200 µl of 0.5 mm glass beads in a Mini-Beadbeater-16 (BioSpec Products, Bartlesville, OK). For TEM, where glass beads cannot be used, cysts in 200-µl lysis buffer were broken by sonication four times for 20 seconds each in continuous mode in a Sonicator Cell Disruptor (formerly Heat Systems Ultrasonic, now Qsonica, Newtown, CT). Broken cysts were added to the top a 15-ml falcon tube containing 60% sucrose and centrifuged at 4,000 x g for 10 min. The broken cyst wall pellet was suspended in PBS buffer and washed three times at 10,000 x g in a microcentrifuge. The cyst wall pellet was used directly for labeling for SIM or fixation for TEM. For mass spectrometry, the cyst wall pellet broken in the bead beater was overlaid on gradient containing 2 ml each of 20%, 40%, 60% and 80% Percoll (top to bottom) and centrifuged for 20 min at 3,000 x g. The layer between 60% and 80% Percoll, where the broken cyst walls were located, was collected and washed in PBS. The cyst wall preparation was suspended in 10 ml of PBS, placed in a syringe, and forced through a 25-mm diameter Whatman Nuclepore Track-Etched Membrane with 8-µm holes (Sigma-Aldrich). The cellular debris, which passed through the membranes, was discarded. The membrane was removed from the cassette, suspended in 5 ml of PBS, and vortexed to release cyst walls. The membrane was removed, and cyst walls were distributed in microfuge tubes and pelleted at 15,000 x g for 10 min. The pellet was suspended in 50 µl PBS and stored at -20°C prior to trypsin digestion and mass spectrometry analysis.

SIM of glycopolymers of encysting organisms, mature cysts, and purified cyst walls. A GST-AcCBM49 fusion-construct, which contains the N-terminal CBM49 of a representative Luke lectin minus the signal peptide, was prepared by codon optimization (76 to 330-bp coding region of ACA1_377670) (GenScript, Piscataway, NJ). It was cloned into pGEX-

6p-1 (GE Healthcare Life Sciences, Marlborough, MA) for cytoplasmic expression in BL21(DE3) chemically competent *E. coli* (Thermo Fisher Scientific, Waltham, MA) [41, 56]. Expression of GST-AcCBM49 and GST were induced with 1 mM IPTG for 4 hr at RT, and GST-fusions were purified on glutathione-agarose and conjugated to Alexa Fluor 594 succinimidyl esters (red) (Molecular Probes, Thermo Fisher Scientific). Approximately 10^6 trophozoites, encysting organisms, mature cysts, or cyst walls were washed in PBS and fixed in 1% paraformaldehyde buffered with phosphate for 15 min at RT. Pellets were washed two times with Hank's Buffered Saline Solution (HBSS) and incubated with HBSS containing 1% bovine serum albumin (BSA) for 1 hour at RT. Preparations were then incubated for 2 hr at 4^0C with 2.5 µg GST or GST-CBM49 conjugated to Alexa Fluor 594 and 12.5 µg of WGA (Vector Laboratories, Burlingame, CA) conjugated to Alexa Fluor 488 in 100 µl HBSS [52, 53]. Finally, pellets were labeled with 100 µg of calcofluor white M2R (Sigma-Aldrich) in 100 µl HBSS for 15 min at RT and washed five times with HBSS [51, 55]. Preparations were mounted in Mowiol mounting medium (Sigma-Aldrich) and observed with widefield and differential interference contrast microscopy, using a 100x objective of a Zeiss AXIO inverted microscope with a Colibri LED (Carl Zeiss Microcopy LLC, Thornwood, NY). Images were collected at 0.2-µm optical sections with a Hamamatsu Orca-R2 camera and deconvolved using ZEN software (Zeiss). Alternatively, SIM was performed with a 63-x objective of a Zeiss ELYRA S.1 microscope at Boston College (Chestnut Hill, MA), and 0.09-µm optical sections deconvolved using Zen software [50]. To count ostioles, we rotated 3D reconstructions of each cyst wall labeled with WGA, MBP-Luke, MBP-Leo, Luke-GFP, or Leo-GFP (24 cysts total), which were prepared as described below.

TEM of intact and purified walls. High-pressure freezing and freeze substitution were used to prepare cyst and cyst walls for TEM at the Harvard Medical School Electron Microscope facility [57]. To make them noninfectious, we fixed mature cysts in 1% paraformaldehyde for 10 min at RT and washed them 2 times in PBS. Cyst walls in PBS were pelleted, placed in 6-mm Cu/Au carriers, and frozen in an EM ICE high-pressure freezer (Leica Microsystems, Buffalo Grove, Il). Freeze substitution was performed in a Leica EM AFS2 instrument in dry acetone containing 1% $_{dd}H_2$0, 1% OsO_4, and 1% glutaraldehyde at -90°C for 48 hr. The temperature was increased 5°C/hour to 20°C, and samples were washed 3 times in pure acetone and once in propylene oxide for 10 min each. Samples were infiltrated with 1:1 Epon:propylene oxide overnight at 4°C and embedded in TAAB Epon (Marivac Canada Inc. St. Laurent, Canada). Ultrathin sections (80 to 100 nm thick) were cut on a Leica Reichert Ultracut S microtome, picked up onto copper grids, stained with lead citrate, and examined in a JEOL 1200EX transmission electron microscope (JEOL USA, Peabody, MA). Images were recorded with an AMT 2k CCD camera.

Mass spectrometry of tryptic and chymotryptic peptides from cyst walls. Broken cyst walls, prepared as above, were dissolved into 50 mM NH_4HCO_3, pH 8.0, reduced with 50 mM dithiothreithol (DTT) for 20 min at 60°C, alkylated with iodoacetamide (IAA) for 20 min at RT, and then digested with proteomics grade trypsin (Sigma-Aldrich) overnight at 37°C. Alternatively broken cyst walls either before or after digestion with trypsin were reconstituted in 1× reducing SDS/PAGE loading buffer and run on a 4–20% precast polyacrylamide TGX gel (Bio-Rad). Bands stained by colloidal Coomassie blue were excised and washed with 50 mM NH_4HCO_3/acetonitrile (ACN). Reduction, alkylation, and trypsin/chymotrypsin digestion were performed in-gel. Peptides were dried and desalted using C18 ZipTip concentrators (EMD

Millipore). Peptides from five biological replicates for both *in solution* and *in-situ* hydrolysis were dissolved in 2% ACN, 0.1% formic acid (FA) and separated using a nanoAcquity-UPLC system (Waters) equipped with a 5-µm Symmetry C18 trap column (180 µm x 20 mm) and a 1.7-µm BEH130 C18 analytical column (150 µm × 100 mm). Samples were loaded onto the precolumn and washed for 4 min at a flow rate of 4 µl/min with 100% mobile phase A (99% Water/1% ACN/0.1% FA). Samples were eluted to the analytical column with a gradient of 2–40% mobile phase B (99% ACN/1% Water/0.1% FA) delivered over 40 or 90 min at a flow rate of 0.5 µl/min. The analytical column was connected online to a QE or a QE-HF Mass Spectrometer (Thermo Fisher Scientific) equipped with a Triversa NanoMate (Advion Inc., Ithaca, NY) electrospray ionization (ESI) source, which was operated at 1.7 kV. Data were acquired in automatic Data Dependent top 10 (QE) or top 20 (QE-HF) mode. Automated database searches were performed using the PEAKS software suite version 8.5 (Bioinformatics Solutions Inc., Waterloo, ON, Canada). The search criteria were set as follows: trypsin/chymotrypsin as the enzyme with ≤ 3 missed cleavages and ≤ 1 non-specific cleavage, the error tolerances for the precursor of 5 ppm and 0.05 Da for fragment ions, carbamidomethyl cysteine as a fixed modification, oxidation of methionine, Pyro-glu from glutamine, and deamidation of asparagine or glutamine as variable modifications. The peptide match threshold (-10 logP) was set to 15, and only proteins with a minimum of two unique peptides were considered. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifiers PXD011826 and 10.6019/PXD011826 [95].

Bioinformatic characterization of candidate CWPs. Signal peptides and transmembrane helices were predicted using SignalP 4.1 and TMHMM, respectively [37, 60]. GPI anchors were searched for using big-PI [59]. AmoebaDB, which contains sequence information from the Neff strain, was used to identify genome sequences, predict introns, and identify paralogous proteins by BLASTP [45, 47]. TBLASTN were used to identify homologous proteins in 11 sequenced genomes, which have been assembled without protein predictions. BLASTP against the NR database at the NCBI was used to identify homologs of candidate CWPs in other species and conserved domains [58]. Carbohydrate-binding modules were searched using CAZy and InterPro databases [42, 43].

Expression and visualization of GFP-fusions in transfected *A. castellanii*. We used RT-PCR from RNA of encysting protists to obtain the coding sequences of a representative Luke lectin (840-bp CDS of ACA1_377670) (Luke(2), Leo lectin (562-bp CDS of ACA1_074730), and Jonah lectin (1596-bp CDS of ACA1_164810) (Jonah(1)). Using NEBuilder HiFi DNA assembly (New England Biolabs, Ipswich, MA), we cloned each CDS into the pGAPDH plasmid, which was a kind gift from Yeonchul Hong [62, 63]. pGAPDH contains a neomycin resistance gene under a TATA-box promoter (for selection with G418) and a glyceraldehyde 3-phosphate dehydrogenase promoter for constitutive expression of GFP-fusions (S1 Fig and S3 Excel file). The GFP tag was placed at the C-terminus of each CWP, and a polyadenylation sequence was added downstream of the GFP-fusion's stop codon. As controls, GFP alone and GFP with a 60-bp sequence encoding the signal peptide of Luke(2) lectin were cloned into the same pGAPDH vector. For expression of CWP genes under their own promoters, we replaced the GAPDH promoter with 446-bp from the 5 'UTR of the Luke(2) gene, 486-bp from the 5' UTR of the Leo gene, and 571-bp of the 5'UTR of the Jonah(1) gene,

each cloned from the genomic DNA. A second Luke lectin representative (500-bp 5' UTR and 1293-bp CDS of ACA1_245650; Luke(3)) was tagged with GFP. The 470-bp 5' UTR and 525-bp CDS of cyst-specific protein 21 (CSP21) (ACA1_075240) were also tagged with GFP [36, 69].

Transfections in *A. castellanii* were performed as described previously [62] with some modifications. Briefly, 5 x 10^5 log-phase trophozoites were allowed to adhere to 6-well plates in ATCC medium 712 for 30 min at 30°C. The adherent trophozoites were washed and replaced with 500 µl of non-nutrient medium (20 mM Tris-HCl [pH 8.8], 100 mM KCl, 8 mM $MgSO_4$, 0.4 mM $CaCl_2$ and 1 mM $NaHCO_3$). In an Eppendorf tube, 4 µg of Midiprep (PureLink HiPure Midiprep Kit, Thermo Fisher Scientific) plasmid DNA was diluted to 100 µl with non-nutrient medium. Twenty microliters of SuperFect Transfection Reagent (Qiagen Inc, Germantown, MD) was added to the DNA suspension, mixed gently by pipetting five times, and incubated for 10 min at RT. Six hundred microliters of non-nutrient medium was added to the DNA-SuperFect mix, and the entire suspension added to the trophozoites adhering to the 6-well culture plate. The culture plate was incubated for 3 hr at 30°C, after which the non-nutrient medium was replaced with ATCC medium 712 and incubated for another 24 hr at 30°C. To select for transfectants, we added 12.5 µg/ml of Gibco G418 antibiotic (Thermo Fisher Scientific) to the culture after 24 hr, and we changed the medium plus antibiotic every 4 days. After 2 to 4 weeks, the transfectants were growing robustly in the presence of the antibiotic, and trophozoites and/or cysts expressing GFP were detected by widefield microscopy. These organisms were induced to encyst, fixed, labeled with WGA and CFW, and examined by widefield microscopy and SIM, as described above.

Labeling of encysting parasites and mature cysts with CWPs fused to MBP. MBP-fusion constructs were prepared (S3 Excel file) by cloning the cDNA of a representative Luke(2) lectin (60 to 843-bp CDS of ACA1_377670) and a representative Jonah(1) lectin (70 to 1599-bp CDS of ACA1_164810) without their signal sequences into pMAL-p2x vector (New England Biolabs) for periplasmic expression in BL21-CodonPlus(DE3)-RIPL (Agilent Technologies, Lexington, MA) [61]. For the Leo-MBP fusion, the Leo CDS without the signal sequence (67 to 564-bp of ACA1_074730) was codon optimized (GenScript) and cloned into pMAL-p2x vector. MBP-Luke(2) was induced with 250 µM IPTG for 5 hr at RT; MBP-Jonah(1) was induced with 1 mM IPTG for 5 hr at RT; and MBP-Leo was induced with 250 µM IPTG for 3.5 hr at 37^0C. MBP-fusion proteins were purified with amylose resin following the manufacturer's instructions (GE Healthcare, Pierce, Agilent Technologies and New England Biolabs). Encysting organisms and mature cysts were fixed, blocked, and incubated with 15 µg of each MBP-CWP fusion conjugated to Alex Fluor 594 in for 2 hr at 4^oC. Preparations were labeled with WGA conjugated to Alexa Fluor 488 and CFW, as described above.

Labeling of mature cysts expressing GFP-fusions with anti-GFP antibodies. Mature cysts expressing GFP-fusions under their own promoter were blocked with BSA, incubated with 1:400 mouse anti-GFP IgG (Roche) for 1 hr at RT, washed and then incubated with 1:800 goat anti-mouse IgG-Alexa Fluor 594 (Molecular Probes, Invitrogen). Preparations were labeled with CFW and fixed in paraformaldehyde as the last step prior to mounting.

Binding of GFP- and MBP-fusions to microcrystalline cellulose and chitin beads. Approximately 2 x 10^6 trophozoites expressing GFP-tagged CWPs or GFP alone under a GAPDH promoter were lysed in 100 µl of 1% NP40. Mass spectrometry of GFP-fusions purified on GFP-Trap_A (ChromoTek GmbH, Planegg-Martinsried, Germany) showed that

each preparation was strongly enriched with GFP-fusions but also contained actin and actin-binding proteins. GFP-fusions from 2 x 10^6 lysed trophozoites or MBP-fusions (1 µg each in 100 µl of 1% NP40) were incubated with 0.5 µg Avicel microcrystalline cellulose (Sigma-Aldrich) or a 50-µl slurry of magnetic chitin beads (New England Biolabs) for 3 hr at 4^0C with rocking. Cellulose or chitin beads were centrifuged to collect the supernatant (unbound fraction) and pellet (bound fraction). The pellet was washed three times with 1% NP40. To solubilize proteins, the input material (total), unbound, and bound fractions were boiled in SDS sample buffer. Proteins were visualized in SDS-PAGE gels and Western blots using anti-GFP (Roche) or anti-MBP (New England Biolabs) antibodies.

Western blots of *A. castellanii* trophozoite and cyst lysates probed with anti-lectin rabbit antibodies. Log-phase trophozoites and 36-hr-old cysts were harvested, and the total protein solubilized in SDS sample buffer, run in SDS-PAGE gels, blotted on PVDF membranes, and blocked in 5% BSA. MBP-CWP fusions and MBP alone were run in adjacent lanes as positive and negative controls, respectively. The blots were probed with 1:100 dilutions of rabbit polyclonal antibodies (Li International) raised to 16- or 50-amino acid peptides of representative Luke (residues 230-279 of ACA1_377670), Leo (residues 124-139 of ACA1_074730) and Jonah (residues 362-411 of ACA1_164810). A 1:1000 dilution of anti-rabbit IgG-HRP (BioRad) was used as secondary antibody and Super Signal West Pico PLUS (ThermoFisher Scientific) for chemiluminescent detection. Coomassie stained gels were run in parallel for loading control.

References.

1. Clarke M, Lohan AJ, Liu B, Lagkouvardos I, Roy S, Zafar N, et al. Genome of *Acanthamoeba castellanii* highlights extensive lateral gene transfer and early evolution

of tyrosine kinase signaling. Genome Biol 2013;14(2):R11. doi: 10.1186/gb-2013-14-2-r11.

2. Clarke B, Sinha A, Parmar DN, Sykakis E. Advances in the diagnosis and treatment of *Acanthamoeba* keratitis. J Ophthalmol 2012:484892 doi: 10.1155/2012/484892.
3. Lorenzo-Morales J, Khan NA, Walochnik J. An update on *Acanthamoeba* keratitis: diagnosis, pathogenesis and treatment. Parasite 2015;22:10 doi: 10.1051/parasite/2015010.
4. Carrijo-Carvalho LC, Sant'ana VP, Foronda AS, de Freitas D, de Souza Carvalho FR. Therapeutic agents and biocides for ocular infections by free-living amoebae of *Acanthamoeba* genus. Surv Ophthalmol 2017;62(2):203-218 doi: 10.1016/j.survophthal.2016.10.00.
5. Carnt N, Robaei D, Minassian DC, Dart JKG. *Acanthamoeba* keratitis in 194 patients: risk factors for bad outcomes and severe inflammatory complications. Br J Ophthalmol. 2018;102(10):1431-35 doi: 10.1136/bjophthalmol-2017-310806.
6. Satlin MJ, Graham JK, Visvesvara GS, Mena H, Marks KM, Saal SD, et al. Fulminant and fatal encephalitis caused by *Acanthamoeba* in a kidney transplant recipient: case report and literature review. Transpl Infect Dis 2013;15(6):619-26 doi: 10.1111/tid.12131.
7. Lorenzo-Morales J, Martín-Navarro CM, López-Arencibia A, Arnalich-Montiel F, Piñero JE, Valladares B. *Acanthamoeba* keratitis: an emerging disease gathering importance worldwide? Trends Parasitol 2013:29(4):181-7 doi: 10.1016/j.pt.2013.01.006.
8. Cope JR, Collier SA, Rao MM, Chalmers R, Mitchell GL, Richdale K, et al. Contact lens wearer demographics and risk behaviors for contact lens-related eye infections--United

States, 2014. MMWR Morb Mortal Wkly Rep 2015;64(32):865-70 DOI: 10.15585/mmwr.mm6432a2.

9. Carnt N, Hoffman JJ, Verma S, Hau S, Radford CF, Minassian DC, et al. *Acanthamoeba* keratitis: confirmation of the UK outbreak and a prospective case-control study identifying contributing risk factors. Br J Ophthalmol 2018:312544. doi: 10.1136/bjophthalmol-2018-312544.
10. Pittet D, Allegranzi B, Boyce J. The World Health Organization Guidelines on Hand Hygiene in Health Care and their consensus recommendations. Infect Control Hosp Epidemiol 2009;30(7):611-22 doi: 10.1086/600379.
11. Mekonnen MM, Hoekstra AY. Four billion people facing water scarcity. Sci Adv 2016;2(2):e1500323 doi: 10.1126/sciadv.1500323.
12. Aqeel Y, Rodriguez R, Chatterjee A, Ingalls RR, Samuelson J. Killing of diverse eye pathogens (*Acanthamoeba castellanii*, *Fusarium solani*, and *Chlamydia trachomatis*) with alcohols. PLoS Negl Trop Dis 2017;11(2):e0005382 doi: 10.1371/journal.pntd.0005382.
13. Tosetti N, Croxatto A, Greub G. Amoebae as a tool to isolate new bacterial species, to discover new virulence factors and to study the host-pathogen interactions. Microb Pathog 2014;77:125-30 doi: 10.1016/j.micpath.2014.07.009.
14. Van der Henst C, Scrignari T, Maclachlan C, Blokesch M. An intracellular replication niche for *Vibrio cholerae* in the amoeba *Acanthamoeba castellanii*. ISME J 2016;10(4):897-910 doi: 10.1038/ismej.2015.165.
15. Vieira A, Seddon, Karlyshev AV. *Campylobacter-Acanthamoeba* interactions. Microbiology 2015;161(Pt 5):933-47 doi: 10.1099/mic.0.000075.

16. La Scola B. Looking at protists as a source of pathogenic viruses. Microb Pathog 2014;77:131-5 doi: 10.1016/j.micpath.2014.09.005.
17. Abergel C, Legendre M, Claverie JM. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. FEMS Microbiol Rev 2015;39(6):779-96 doi: 10.1093/femsre/fuv037.
18. Kong HH, Pollard TD. Intracellular localization and dynamics of myosin-II and myosin-IC in live *Acanthamoeba* by transient transfection of EGFP fusion proteins. J Cell Sci 2002;115(Pt 24):4993-5002 doi: 10.1242/jcs.00159.
19. Ostap EM, Maupin P, Doberstein SK, Baines IC, Korn ED, Pollard TD. Dynamic localization of myosin-I to endocytic structures in *Acanthamoeba*. Cell Motil Cytoskeleton 2003;54(1):29-40 doi: 10.1002/cm.10081.
20. Neff RJ, Ray SA, Benton WF, Wilborn M. Induction of synchronous encystment (differentiation) in *Acanthamoeba* sp. Methods Cell Biol 1964;1:55-83 doi: 10.1016/S0091-679X(08)62086-5
21. Jensen T, Barnes WG, Meyers D. Axenic cultivation of large populations of *Acanthamoeba* castellanii (JBM). J Parasitol 1970;56(5):904–6 doi: 10.2307/3277503.
22. Lloyd D. Encystment in *Acanthamoeba castellanii*: a review. Exp Parasitol 2014;145:Suppl:20-7 doi: 10.1016/j.exppara.2014.03.026.
23. Bowers B, Korn ED. The fine structure of *Acanthamoeba castellanii* (Neff strain). II. Encystment. J Cell Biol 1969;41(3): 786–805 doi: 10.1083/jcb.41.3.786.
24. Chávez-Munguía B, Omaña-Molina M, González-Lázaro M, González-Robles A, Bonilla P, Martínez-Palomo A. Ultrastructural study of encystation and excystation in

Acanthamoeba castellanii. J Eukaryot Microbiol 2005;52(2):153-8 DOI | 10.1111/j.1550-7408.2005.04-3273.

25. Lemgruber L, Lupetti P, De Souza W, Vommaro RC, da Rocha-Azevedo B. The fine structure of the *Acanthamoeba polyphaga* cyst wall. FEMS Microbiol Lett 2010;305(2):170-6 doi: 10.1111/j.1574-6968.2010.01925.
26. Chambers JA, Thompson JE. A scanning electron microscopic study of the excystment process of *Acanthamoeba castellanii*. Exp Cell Res 1972;73(2):415-21 doi: 10.1016/0014-4827(72)90066-3.
27. Coulon C, Collignon A, McDonnell G, Thomas V. Resistance of *Acanthamoeba* cysts to disinfection treatments used in health care settings. J Clin Microbiol 2010;48(8): 2689-2697 doi: 10.1128/JCM.00309-10.
28. Dupuy M, Berne F, Herbelin P, Binet M, Berthelot N, Rodier MH, et al. Sensitivity of free-living amoeba trophozoites and cysts to water disinfectants. Int J Hyg Environ Health 2014(2-3);217:335-9 doi: 10.1016/j.ijheh.2013.07.007.
29. Ezz Eldin HM, Sarhan RM. Cytotoxic effect of organic solvents and surfactant agents on *Acanthamoeba castellanii* cysts. Parasitol Res. 2014;113(5):1949-53 doi: 10.1007/s00436-014-3845-5..
30. Lonnen J, Putt KS, Kernick ER, Lakkis C, May L, Pugh RB. The efficacy of *Acanthamoeba* cyst kill and effects upon contact lenses of a novel ultraviolet lens disinfection system. Am J Ophthalmol 2014;158(3):460-8 doi: 10.1016/j.ajo.2014.05.032.

31. Flores BM, Garcia CA, Stamm WE, Torian BE. Differentiation of *Naegleria fowleri* from *Acanthamoeba* species by using monoclonal antibodies and flow cytometry. J Clin Microbiol 1990;28(9):1999-2005 doi: jcm.asm.org/content/jcm/28/9/1999.full.pdf.
32. Hiwatashi E, Tachibanabi H, Kanedab Y, Obazawaa H. Production and characterization of monoclonal antibodies to *Acanthamoeba castellanii* and their application for detection of pathogenic *Acanthamoeba spp*. Parasitol Internat 1997;46(3):197-205 doi: 10.1016/S1383-5769(97)00029-9.
33. Leher H, Zaragoza F, Taherzadeh S, Alizadeh H, Niederkorn JY. Monoclonal IgA antibodies protect against *Acanthamoeba* keratitis. Exp Eye Res 1999;69(1):75-84 doi: 10.1006/exer.1999.0678.
34. Turner ML, Cockerell EJ, Brereton HM, Badenoch PR, Tea M, Coster DJ, et al. Antigens of selected *Acanthamoeba* species detected with monoclonal antibodies. Int J Parasitol. 2005;35(9):981-90 doi: 10.1016/j.ijpara.2005.03.015.
35. Kang AY, Park AY, Shin HJ, Khan NA, Maciver SK, Jung SY. Production of a monoclonal antibody against a mannose-binding protein of *Acanthamoeba culbertsoni* and its localization. Exp Parasitol 2018;192:19-24 doi: 10.1016/j.exppara.2018.07.009. .
36. Hirukawa Y, Nakato H, Izumi S, Tsuruhara T, Tomino S. Structure and expression of a cyst specific protein of *Acanthamoeba castellanii*. Biochim Biophys Acta 1998;1398(1):47-56 doi: 10.1016/S0167-4781(98)00026-8.
37. Nielsen H. Predicting Secretory Proteins with SignalP. Methods Mol Biol 2017;1611:59-73 doi: 10.1007/978-1-4939-7015-5_6.
38. Potter JL, Weisman RA. Cellulose synthesis by extracts of *Acanthamoeba castellanii* during encystment. Stimulation of the incorporation of radioactivity from UDP-

(^{14}C)glucose into alkali-soluble and insoluble beta-glucans by glucose 6-phosphate and related compounds. Biochim Biophys Acta 1976;428(1):240-52 doi: 10.1016/0304-4165(76)90125-2.

39. Derda M, Winiecka-Krusnell J, Linder MB, Linder E. Labeled *Trichoderma reesei* cellulase as a marker for *Acanthamoeba* cyst wall cellulose in infected tissues. Appl Environ Microbiol. 2009;75(21):6827-30 doi: 10.1128/AEM.01555-09.
40. Dudley R, Jarroll EL, Khan NA. Carbohydrate analysis of *Acanthamoeba castellanii*. Exp Parasitol. 2009;122(4):338-43 doi: 10.1016/j.exppara.2009.04.009.
41. Urbanowicz BR, Catalá C, Irwin D, Wilson DB, Ripoll DR, Rose JK. A tomato endo-beta-1,4-glucanase, SlCel9C1, represents a distinct subclass with a new family of carbohydrate binding modules (CBM49). J Biol Chem 2007;282(16):12066-74 DOI: 10.1074/jbc.M607925200.
42. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The Carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 2014;42 (Database issue):D490–D495 doi: 10.1093/nar/gkt1178.
43. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017 — beyond protein family and domain annotations. Nucleic Acids Res 2017;45(D1):D190-D199 doi: 10.1093/nar/gkw1107.
44. Samuelson J, Bushkin GG, Chatterjee A, Robbins PW. Strategies to discover the structural components of cyst and oocyst walls. Eukaryot Cell 2013;12(2):1578-87 doi: 10.1128/EC.00213-13.

45. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389-3402 doi: 10.1093/nar/25.17.3389
46. Das S, Van Dellen K, Bulik D, Magnelli P, Cui J, Head J, et al. The cyst wall of *Entamoeba invadens* contains chitosan (deacetylated chitin). Mol Biochem Parasitol 2006;148(1):86-92 doi: 10.1016/j.molbiopara.2006.03.002.
47. Aurrecoechea C, Barreto A, Brestelli J, Brunk BP, Caler EV, Fischer S, et al. AmoebaDB and MicrosporidiaDB: functional genomic resources for Amoebozoa and *Microsporidia* species. Nucleic Acids Res 2011;39:D612-9 doi: 10.1093/nar/gkq1006
48. Gonçalves IR, Brouillet S, Soulié MC, Gribaldo S, Sirven C, Charron N, et al. Genome-wide analyses of chitin synthases identify horizontal gene transfers towards bacteria and allow a robust and unifying classification into fungi. BMC Evol Biol 2016;16(1):252 doi.org/10.1186/s12862-016-0815-9.
49. Morozov AA, Likhoshway YV. Evolutionary history of the chitin synthases of eukaryotes. Glycobiology 2016;26(6):635-9 doi: 10.1093/glycob/cww018.
50. Demmerle J, Innocent C, North AJ, Ball G, Müller M, et al. Strategic and practical guidelines for successful structured illumination microscopy. Nat Protoc 2017;12(5):988-1010 doi: 10.1038/nprot.2017.019.
51. Monheit JE, Cowan DF, Moore Dg. Rapid detection of fungi in tissues using calcofluor white and fluorescence microscopy. Arch Pathol Lab Med 1984;108(8);616-8.
52. Shaw JA, Mol PC, Bowers B, Silverman SJ, Valdivieso MH, Durán A, et al. The function of chitin synthases 2 and 3 in the *Saccharomyces cerevisiae* cell cycle. J Cell Biol 1991;114(1):111-23 DOI: 10.1083/jcb.114.1.111.

53. Frisardi M, Ghosh SK, Field J, Van Dellen K, Rogers R, Robbins P, et al. The most abundant glycoprotein of amebic cyst walls (Jacob) is a lectin with five Cys-rich, chitin-binding domains. Infect Immun 2000;68(7):4217-24 DOI: 10.1128/IAI.68.7.4217-4224.2000.
54. Said-Fernández S, Campos-Góngora E, González-Salazar F, Martínez-Rodríguez HG, Vargas-Villarreal J, Viader-Salvadó JM. Mg2+, Mn2+, and Co2+ stimulate *Entamoeba histolytica* to produce chitin-like material. J Parasitol 2001;87(4):919-23 DOI: 10.1645/0022-3395(2001)087[0919:MMACSE]2.0.CO;2.
55. Wilhelmus KR, Osato MS, Font RL, Robinson NM, Jones DB. Rapid diagnosis of *Acanthamoeba* keratitis using calcofluor white. Arch Ophthalmol 1986;104(9):1309-12 doi:10.1001/archopht.1986.01050210063026.
56. Smith DB. Generating fusions to glutathione S-transferase for protein studies. Methods Enzymol 2000;326:254-70 doi: 10.1016/S0076-6879(00)26059.
57. Vanhecke D, Graber W, Studer D. Close-to-native ultrastructural preservation by high pressure freezing. Methods Cell Biol 2008;88:151-64 doi: 10.1016/S0091-679X(08)00409-3.
58. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. Nucleic Acids Res 2015;43(Database issue):D222-6. doi: 10.1093/nar/gku1221.
59. Eisenhaber B, Bork P, Eisenhaber F. Prediction of potential GPI-modification sites in proprotein sequences. J Mol Biol 1999;292(3):741-58 DOI: 10.1006/jmbi.1999.3069.

60. Krough A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305(3):567-80 doi.org/10.1006/jmbi.2000.4315.
61. Kapust RB, Waugh DS. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 1999;8(8):1668-74 DOI: 10.1110/ps.8.8.1668.
62. Peng Z, Omaruddin R, Bateman E. Stable transfection of *Acanthamoeba castellanii*. Biochim Biophys Acta 2005;1743(1-2):93-100 DOI: 10.1016/j.bbamcr.2004.08.014.
63. Bateman E. Expression plasmids and production of EGFP in stably transfected *Acanthamoeba*. Protein Expr Purif 2010;70(1):95-100. doi: 10.1016/j.pep.2009.10.008.
64. Xu Y, Liang X, Chen Y, Koehler TM, Hook M. Identification and biochemical characterization of two novel collagen binding MSCRAMMs of *Bacillus anthracis*. J Biol Chem. 2004;279(50):51760-8 DOI: 10.1074/jbc.M406417200.
65. Enguita FJ, Martins LO, Henriques AO, Carrondo MA. Crystal structure of a bacterial endospore coat component. A laccase with enhanced thermostability properties. J Biol Chem 2003;278(21):19416-2 DOI: 10.1074/jbc.M301251200.
66. Kiiskinen LL, Palonen H, Linder M, Viikari L, Kruus K. Laccase from *Melanocarpus albomyces* binds effectively to cellulose. FEBS Lett. 2004;576(1-2):251-5 DOI: 10.1016/j.febslet.2004.08.040.
67. Moon EK, Chung DI, Hong YC, Kong HH. Characterization of a serine proteinase mediating encystation of *Acanthamoeba*. Eukaryot Cell. 2008 Sep;7(9):1513-7 doi: 10.1128/EC.00068-08.

68. Corsaro D, Köhsler M, Montalbano Di Filippo M, Venditti D, Monno R, et al. Update on *Acanthamoeba jacobsi* genotype T15, including full-length 18S rDNA molecular phylogeny. Parasitol Res. 2017 Apr;116(4):1273-1284. doi: 10.1007/s00436-017-5406-1.
69. Chen L, Orfeo T, Gilmartin G, Bateman E. Mechanism of cyst specific protein 21 mRNA induction during *Acanthamoeba* differentiation. Biochim Biophys Acta 2004;1691(1):23-31 DOI: 10.1016/j.bbamcr.2003.11.005.
70. Vollmer AC, Bark SJ. Twenty-Five Years of Investigating the Universal Stress Protein: Function, Structure, and Applications. Adv Appl Microbiol. 2018;102:1-36 doi: 10.1016/bs.aambs.2017.10.001.
71. Chatterjee A, Ghosh SK, Jang K, Bullitt E, Moore L, Robbins PW, Samuelson J. Evidence for a "wattle and daub" model of the cyst wall of entamoeba. PLoS Pathog. 2009;5(7):e1000498 DOI: 10.1371/journal.ppat.1000498.
72. Chatterjee A, Carpentieri A, Ratner DM, Bullitt E, Costello CE, Robbins PW, et al. *Giardia* cyst wall protein 1 is a lectin that binds to curled fibrils of the GalNAc homopolymer. PLoS Pathog. 2010;6(8):e1001059 doi: 10.1371/journal.ppat.1001059.
73. Bushkin GG, Motari E, Magnelli P, Gubbels MJ, Dubey JP, Miska KB, et al. β-1,3-glucan, which can be targeted by drugs, forms a trabecular scaffold in the oocyst walls of *Toxoplasm*a and *Eimeria*. MBio 2012;3(5): e00258-12. doi: 10.1128/mBio.00258-12.
74. Van Dellen K, Ghosh SK, Robbins PW, Loftus B, Samuelson J. *Entamoeba histolytica* lectins contain unique 6-Cys or 8-Cys chitin-binding domains. Infect Immun 2002;70(6):3259-63 DOI: 10.1128/IAI.70.6.3259-3263.2002.

75. Vinayak S, Pawlowic MC, Sateriale A, Brooks CF, Studstill CJ, Bar-Peled Y, et al. Genetic modification of the diarrhoeal pathogen *Cryptosporidium parvum*. Nature 2015;523(7561):477-80 doi: 10.1038/nature14651.
76. Lee S, Ginese M, Beamer G, Danz HR, Girouard DJ, Chapman-Bonofiglio SP, et al. Therapeutic Efficacy of Bumped Kinase Inhibitor 1369 in a Pig Model of Acute Diarrhea Caused by *Cryptosporidium hominis*. Antimicrob Agents Chemother 2018;62(7): e00147-18. doi: 10.1128/AAC.00147-18.
77. Cherif M, Benhamou N, Blanger RR. Occurrence of cellulose and chitin in the hyphal walls of *Pythium ultimum:* a comparative study with other plant pathogenic fungi. Can J Microbiol 1993;39(2):212-223 doi: 10.1139/m93-030.
78. West CM. Comparative analysis of spore coat formation, structure, and function in Dictyostelium. Int Rev Cytol 2003;222:237-93 doi.org/10.1016/S0074-7696(02)22016-1.
79. West CM, Nguyen P, van der Wel H, Metcalf T, Sweeney KR, Blader IJ, et al. Dependence of stress resistance on a spore coat heteropolysaccharide in *Dictyostelium*. Eukaryot Cell 2009;8(1):27-36. doi: 10.1128/EC.00398-07.
80. Ghosh SK, Van Dellen KL, Chatterjee A, Dey T, Haque R, Robbins PW, et al. The Jacob2 lectin of the *Entamoeba histolytica* cyst wall binds chitin and is polymorphic. PLoS Negl Trop Dis 2010;4(7):e750 doi: 10.1371/journal.pntd.0000750.
81. Wang Y, Slade MB, Gooley AA, Atwell BJ, Williams KL. Cellulose-binding modules from extracellular matrix proteins of *Dictyostelium discoideum* stalk and sheath. Eur J Biochem. 2001;268(15):4334-45 doi: 10.1046/j.1432-1327.2001.02354.

82. CAZypedia Consortium. Ten years of CAZypedia: a living encyclopedia of carbohydrate-active enzymes. Glycobiology 2018;28(1):3-8. doi: 10.1093/glycob/cwx089.
83. Van Dellen KL, Chatterjee A, Ratner DM, Magnelli PE, Cipollo JF, Steffen M, et al. Unique posttranslational modifications of chitin-binding lectins of *Entamoeba invadens* cyst walls. Eukaryot Cell 2006;5(5):836-48 DOI: 10.1128/EC.5.5.836-848.2006.
84. Shen Z, Jacobs-Lorena M. A type I peritrophic matrix protein from the malaria vector *Anopheles gambiae* binds to chitin. Cloning, expression, and characterization. J Biol Chem 1998273(28):17665-70 doi: 10.1074/jbc.273.28.17665.
85. Maeda H, Ishida N. Specificity of binding of hexopyranosyl polysaccharides with fluorescent brightener. J Biochem 1967;62(2):276-8 .
86. Cabib E, Arroyo J. How carbohydrates sculpt cells: chemical control of morphogenesis in the yeast cell wall. Nat Rev Microbiol 2013;11(9):648-55 DOI: 10.1038/nrmicro3090.
87. Etienne-Manneville S, Hall A. Rho GTPases in cell biology. Nature 2002;420(6916):629-35 DOI: 10.1038/nature01148.
88. Stibbs HH. Monoclonal antibody-based enzyme immunoassay for *Giardia lamblia* antigen in human stool. J Clin Microbiol 1989;27(11):2582-8.
89. Spadafora LJ, Kearney MR, Siddique A, Ali IK, Gilchrist CA, Arju T, et al. Species-Specific Immunodetection of an *Entamoeba histolytica* Cyst Wall Protein. PLoS Negl Trop Dis 2016;10(5):e0004697. doi: 10.1371/journal.pntd.0004697.
90. Chaudhary PM, Tupe SG, Deshpande MV. Chitin synthase inhibitors as antifungal agents. Mini Rev Med Chem 2013;13(2):222-36 DOI : 10.2174/1389557511313020005.

91. Aqeel Y, Siddiqui R, Khan NA. Silencing of xylose isomerase and cellulose synthase by siRNA inhibits encystation in *Acanthamoeba castellanii*. Parasitol Res 2013;112:1221-7 DOI 10.1007/s00436-012-3254-6
92. Moon EK, Hong Y, Chung DI, Goo YK, Kong HH. Down-regulation of cellulose synthase inhibits the formation of endocysts in *Acanthamoeba*. Korean J Parasitol 2014;52:131-5 DOI:10.3347/kjp.2014.52.2.131.
93. Moon EK, Hong Y, Chung DI, Goo YK, Kong HH. Potential Value of Cellulose Synthesis Inhibitors Combined With PHMB in the Treatment of *Acanthamoeba* Keratitis. Cornea. 2015;34(12):1593-8 doi: 10.1097/ICO.0000000000000642.
94. Lakhundi S, Siddiqui R, Khan NA. Cellulose degradation: a therapeutic strategy in the improved treatment of *Acanthamoeba* infections. Parasit Vectors 2015;8:23 doi: 10.1186/s13071-015-0642-7.
95. Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat Biotechnol 2014;32(3):223-6 doi: 10.1038/nbt.2839

Figure legends.

Figure. 1. SIM shows purified cyst walls retain distinct ectocyst and endocyst layers, as well as ostioles. A. After six hr encystation, a GST-AcCBM49 in red and WGA in green each label dozens of small vesicles, which do not appear to overlap. CFW in blue is difficult to see, so overlaps with GST-AcCBM49 and WGA cannot be judged. B. After 12 hr, WGA is still prominently in vesicles, while all three probes bind to a single-layered cyst wall, which contains small, flat ostioles visible only with CFW. C. After 36 hr, GST-AcCBM49 predominantly labels the ectocyst layer; WGA diffusely labels both ectocyst and endocyst

layers and labels ostioles in a punctate manner; while CFW labels endocyst and ostioles. The ectocyst layer of intact (D) and purified *A. castellanii* cyst walls (E) labels with GST-AcCBM49; the edges of ostioles label with WGA; and the endocyst layer labels with CFW. Scale bars are 2 µm.

Figure 2. TEM also shows purified cyst walls retain endocyst and ectocyst layers and ostioles. A, B. Intact and purified *A. castellanii* cyst walls have an outer ectocyst layer (red arrows), an inner endocyst layer (blue arrows), and ostioles (green arrows) that connect the layers. Endocyst and ectocyst layers have the same appearance in intact cysts (C) and purified cyst walls (D). Purified cyst walls are missing amorphous material (purple arrow) between the wall and the plasma membrane of intact cysts. E. At the edge of the ostiole of an intact cyst, the endocyst layer bifurcates, and the outer branch (yellow arrow) meets the ectocyst layer. In the center of the ostiole, the ectocyst layer (pink arrow) forms a narrow cap over the inner branch of the endocyst layer (pale green arrow). Scale bars as marked on micrographs.

Figure 3. Representative proteins of abundant families of CWPs contain two CBM49s (Luke(2) lectin), two 8-Cys domains (Leo lectin), or one CAA domain (Jonah(1) lectin) . A representative Luke lectin has an N-terminal signal peptide (purple) and two CBM49s separated by short Ser- and Pro-rich spacers (light blue). The N-terminal CBM49 contains four Trp residues (red Ws), three of which are conserved in a C-terminal CBM49 of a tomato cellulase (larger font) and three of which are conserved in a single CBM49 of *Dictyostelium* spore coat proteins (underlined). The C-terminal CBM49 has two conserved Trp residues, which are also present in the middle and C-terminal CBM49s of Luke(3) lectins (S2 Fig). A 50-amino acid peptide used to immunize rabbits is underlined. A representative Leo

lectin has a signal peptide and two unique domains (dark blue) containing eight Cys residues each (red Cs). Leo(TKH) lectins have a Thr-,Lys- and His-rich domain between 8-Cys domain (S2 Fig). A 16-amino acid peptide used to immunize rabbits is underlined. A representative Jonah(1) has a signal peptide, a Thr-, Lys-, and Cys-rich domain (tan), and a single CAA domain (green). Jonah(3) lectins have three CAA domains, hydrophobic domains and Ser- and Pro-rich spacers (S2 Fig). A 50-amino acid peptide used to immunize rabbits is underlined. Localization of GFP-tagged CWPs expressed in transformed protists is shown in Figs 4 to 6. Binding of MBP-CWP fusions to encysting organisms and mature cysts and to microcrystalline cellulose and chitin beads are shown in Figs 7 and 8, respectively.

Fig. 4. SIM shows a representative Jonah(1) lectin tagged with GFP and expressed under its own promoter is made early during encystation and localizes to the ectocyst layer of mature cyst walls. A Jonah(1) lectin with a single CAA domain, which was tagged with GFP and expressed under its own promoter in transfected *A. castellanii*, is not expressed by trophozoites (data not shown). A. In contrast, after six hr encystation, Jonah(1)-GFP (green) is present in dozens of small encystation-specific vesicles, which do not overlap with WGA-labeled vesicles (red) containing chitin. B. After 15 hr encystation, Jonah(1)-GFP has a fibrillar appearance in the single-layered, primordial cyst wall, which also labels with WGA and CFW. Small, flat ostioles are labeled by CFW but are not visible with Jonah(1)-GFP or WGA. C. After 24 hr encystation, Jonah(1)-GFP remains with the ectocyst layer, while WGA and CFW predominantly label the endocyst layer and ostioles. D. Jonah(1)-GFP localizes to the ectocyst layer of the wall of mature cysts; WGA labels both ectocyst and endocyst layers; while CFW labels the endocyst layer and ostioles. E. Jonah(1)-GFP expressed under a GAPDH promoter also localizes to the ectocyst layer. F. Anti-GFP antibodies bind to 100% of

mature cysts expressing Jonah(1)-GFP. A to F. Scale bars are 2 µm. Methods and primers for making Jonah(1)-GFP and other constructs are shown in S1 Fig and S3 Excel file. Widefield micrographs of groups of cysts expressing Jonah(1)-GFP and other CWPs under their own promoters are shown in S3 Fig. RT-PCR of mRNAs of trophozoites and encysting organisms with primers to Jonah(1) and other CWPs are shown in S4 Fig. Western blots of rabbit anti-Jonah(1) and ant-Leo anti-antibodies to trophozoites and cyst proteins are shown in S5 Fig. Widefield and DIC micrographs of trophozoites and cysts expressing Jonah(1)-GFP, Luke(2)-GFP, and GFP alone under the GAPDH promoter are shown in S6 Fig.

Figure 5. A representative Luke(2) lectin is made later during encystation, marks small, flat ostioles in developing cyst walls, and is present in the endocyst layer and at the edges of conical ostioles in mature cyst walls. A representative Luke(2) lectin with two CBM49s, which was tagged with GFP and expressed under its own promoter in transfected *A. castellanii*, is not expressed by trophozoites and is difficult to visualize during the first 12 hr of encystation (data not shown). A, B. After 15 and 18 hr encystation, Luke(2)-GFP (green) sharply outlines a subset of small, flat ostioles, which are also visible with CFW (blue) but not with WGA (red). In a more developed cyst wall (lower right), Luke(2)-GFP also coats the endocyst layer. C. After 24 hr encystation, Luke(2)-GFP is diffusely present in the endocyst layer and ostioles, which are labeled with WGA and CFW. D. Luke(2)-GFP is diffusely present in the endocyst layer and sharply outlines conical ostioles of mature cysts. E. Luke(2)-GFP expressed under a GAPDH promoter is also present in the endocyst layer and sharply outlines conical ostioles of mature cysts. F. Luke(3) with three CBM49s, tagged with GFP and expressed under its own promoter, is also present in the endocyst layer and ostioles. A to F. Scale bars are 2 µm.

Figure 6. A representative Leo lectin is also made later during encystation and is present in the endocyst layer and at the edges of conical ostioles in mature cyst walls. A representative Leo lectin with two 8-Cys domains, which was tagged with GFP and expressed under its own promoter in transfected *A. castellanii*, is not expressed by trophozoites and is difficult to visualize during the first 12 hr of encystation (data not shown). A. After 15 hr encystation, Leo-GFP (green) is present in dozens of encystation-specific vesicles and is weakly present in the single-layered, primordial cyst wall, which is diffusely labeled red with WGA. In contrast, CFW alone labels small, flat ostioles. B. At 18 hr, Leo-GFP forms patches on the cyst wall, which are for the most part independent of ostioles. Only at 36 hr (C) does Leo-GFP accumulate at the edges of the ostioles, which is its location in mature cysts (D). Control cysts show punctate labeling of CSP21-GFP expressed under its own promoter (E) and GFP with a signal peptide from the N-terminus of Luke(2) lectin expressed under a GAPDH promoter (F). A to F. Scale bars are 2 µm.

Figure 7. MBP-CWP fusions show glycopolymers bound by Jonah(1) lectin are accessible in the ectocyst layer of mature cyst walls, while glycopolymers bound by Luke(2) and Leo lectins are much less accessible in the endocyst layer and ostioles. After 12 hr encysting, MBP-Jonah (A), MBP-Luke(2) (C), and MBP-Leo (E), each labels vesicles in nearly 100% of organisms, the primordial walls of which are labeled green with WGA and blue with CFW. While MBP-Jonah(2) labels the ectocyst layer of nearly 100% of mature cysts (B), MBP-Luke(2) (D) and MBP-Leo (F) each label the endocyst layer and ostioles of <10% of mature cysts. A to F. Scale bars are 2 µm. Widefield micrographs of groups of mature cysts labeled with MBP-CWP fusions are shown in S7 Fig.

Figure 8. Representative Western blots of pull-downs with microcrystalline cellulose and chitin beads show Luke, Leo, and Jonah lectins all bind cellulose, while binding to chitin is variable. Luke(2)-GFP, Jonah(1)-GFP, and GFP alone were each expressed under the GAPDH promoter in transfected *A. castellanii* (S1 and S6 Figs), extracted from lysed trophozoites, and incubated with microcrystalline cellulose (A) or chitin beads (B). Total proteins (T), bound proteins (B), and unbound proteins (U), as well as molecular weight markers (M), were run on SDS-PAGE, transferred to PVDF membranes, and detected with an anti-GFP reagent. Full-length products in total fractions are underlined in red. Luke(2)-GFP, which included some breakdown products, bound completely to microcrystalline cellulose (A) and partially to chitin (C). Jonah(1)-GFP, which also included some breakdown products, bound partially to microcrystalline cellulose but not at all to chitin. GFP alone (negative control) did not bind to cellulose or chitin. MBP-CWP fusions and MBP alone were made as recombinant proteins in the periplasm of bacteria and detected with an antibody to MBP. MBP-Leo partially bound to microcrystalline cellulose and bound weakly, if at all to chitin. MBP-Luke(2) and MBP-Jonah(1) each bound more completely to cellulose and partially to chitin, while MBP alone (negative control) did not bind to cellulose and chitin.

Supplemental files:

S1 Figure. Constructs made to localize CWPs and the glycopolymers to which they bind in encysting organisms and mature cysts and to determine their binding to microcrystalline cellulose and chitin beads. A. A representative Luke(2) lectin with two CBM49s was used to make GFP-, GST-, and MBP-fusions, and a representative Luke(3) lectin with three CBM49s was used to make GFP-fusions. B, C. Representative Leo and Jonah(1) lectins were made into GFP- and MBP-fusions. D. Vectors for expressing GFP-fusions in

transfected *A. castellanii* under its own promoter or a constitutive GAPDH promoter contained a neomycin resistance gene under a TATA-binding protein promoter [62, 63]. Primers for making constructs are listed in S3 Excel file.

S2 Figure. Sequences of representative candidate CWPs with three CBM49s (Luke(3) lectin), two 8-Cys domains separated by a Thr-, Lys-, and His-rich domain (Leo(TKH) lectin), or three CAA domains (Jonah(3) lectin). A representative Luke(3) lectin has an N-terminal signal peptide (purple) and three CBM49s separated by short Ser- and Pro-rich spacers (light blue). The CBM49s contain conserved Trp (red Ws) present in Luke(2) lectin (Fig 3). B. A representative Leo(TKH) lectin has a signal peptide, two domains containing eight Cys residues each (red Cs), and a Thr-,Lys- and His-rich spacer (brown). C. A representative Jonah(3) has three CAA domains (green), hydrophobic regions (tan), and short Ser- and Pro-rich spacers (light blue). The representative Luke(3) lectin, expressed with a GFP tag under its own promoter, localizes to the endocyst layer and ostioles (Fig 5).

S3 Figure. Widefield micrographs of mature cysts expressing GFP-labeled CWPs show Jonah(1) lectin is accessible to anti-GFP antibodies, while Luke(2) and Leo-GFP are poorly accessible. A to C. Cysts expressing Jonah(1)-GFP, Luke(2)-GFP, and Leo-GFP (green), each under its own promoter, were labeled with WGA (red) and CFW (blue). D to F. Anti-GFP antibodies (red) show that Jonah(1)-GFP, which localizes to the ectocyst layer by SIM (Fig 4), is accessible in nearly 100% of organisms. In contrast, Luke(2)-GFP and Leo-GFP, which localize to the endocyst layer and ostioles, were accessible in 3% and 2% of cysts, respectively. A to F. Scale bars are 5 µm.

S4 Figure. RT-PCR shows mRNAs of representative Luke(2), Leo, and Jonah(1) lectins, as well as those of cellulose synthase, are encystation-specific. DNA and total

RNA was extracted from trophozoites and organisms encysting for one to three days. RT-PCRs were performed with primers specific for segments of each CWP mRNA, as well as primers specific for segments of mRNAs for GAPDH and cellulose synthase (see S3 Excel file). PCR with DNA was used as a positive control, while omission of reverse-transcriptase (-RT) was used as a negative control. Messenger RNAs encoding CWPs and cellulose synthase are absent or nearly absent in trophozoites but are easily detectable in encysting organisms. In contrast, mRNAs for GAPDH are expressed by trophozoites and encysting organisms. These results support encystation-specific expression of GFP-tagged CWPs under their own promoters (Figs 4 to 6).

S5 Figure. Western blots with rabbit antibodies to peptides of Leo and Jonah(1) lectins show each CWP is absent in trophozoites but is easily detected in mature cysts. Proteins of lysed trophozoites and cysts (Coomassie blue stain in A), as well as MBP alone, MBP-Jonah(1), or MBP-Leo, were separated on SDS-PAGE, transferred to PVDF membranes, and incubated with rabbit antibodies to a 50-amino acid peptide of Jonah(1) (B) or to a 16-amino acid peptide of Leo (C). Full-length products in lysed trophozoites and cysts are underlined in red. Both rabbit antibodies bound to MBP-CWP fusions but not to MBP alone and to proteins from cysts but not from trophozoites. These results confirm encystation-specific expression of Jonah(1) and Leo (Figs 4 and 6). Rabbit antibodies to a 50-amino acid peptide of Leo failed to bind to trophozoites or cysts. None of the rabbit antibodies bound to encysting organisms or mature cyst walls when examined by widefield microscopy.

S6 Figure. Widefield and differential interference contrast (DIC) micrographs of unfixed trophozoites (A) and cysts (B) expressing GFP-tagged CWPs under a GAPDH promoter. Luke(2)-GFP, Jonah(1)-GFP, and GFP alone (green), which were used for pull-

downs with glycopolymers (Fig. 7), all express well in trophozoites, which maintain an amoeboid appearance with numerous acanthopods. Luke(2)-GFP and Jonah(1)-GFP are located in punctate vesicles of trophozoites and in the endocyst layer (Luke(2)-GFP) and ectocyst layer (Jonah(1)-GFP) of mature cyst walls. CFW labels the endocyst layer of mature cyst walls. In contrast, GFP alone remains in the cytosol of trophozoites and cysts. A, B. Scale bars are 5 µm.

S7 Figure. Widefield microscopy shows glycopolymers bound by MBP-Jonah(1) are accessible in the ectocyst layer of mature cyst walls, while glycopolymers bound by MBP-Luke(2) and MBP-Leo are, for the most part, inaccessible in the endocyst layer and ostioles. MBP-CWP fusions made in the periplasm of bacteria and purified on amylose resins were used to label mature cyst walls, which were also labeled with WGA and CFW. Groups of organisms were visualized in order to show that the Jonah(1) lectin binds to nearly 100% of cysts, while Luke(2) and Leo lectins bind to at most one cyst per visual field. SIM of MBP-fusions binding to individual encysting cells or mature cysts are shown in Fig. 7. A to C. Scale bars are 5 µm.

S8 Figure. Contrasting use of CBM49 by *A. castellanii* and *D. discoideum*. CBM49, which was first shown to be a cellulose-binding domain at the C-terminus of tomato cellulase, is repeated two or three times in Luke lectins of *A. castellanii* and is also present at the N-terminus of a metalloprotease. In contrast, CBM49 is present in a single copy in the majority of *D. discoideum* proteins and as three copies in rare proteins. CBM49 is also present in two copies in a *D. discoideum* cysteine protease and as a single copy in a GH5 glycoside hydrolase.

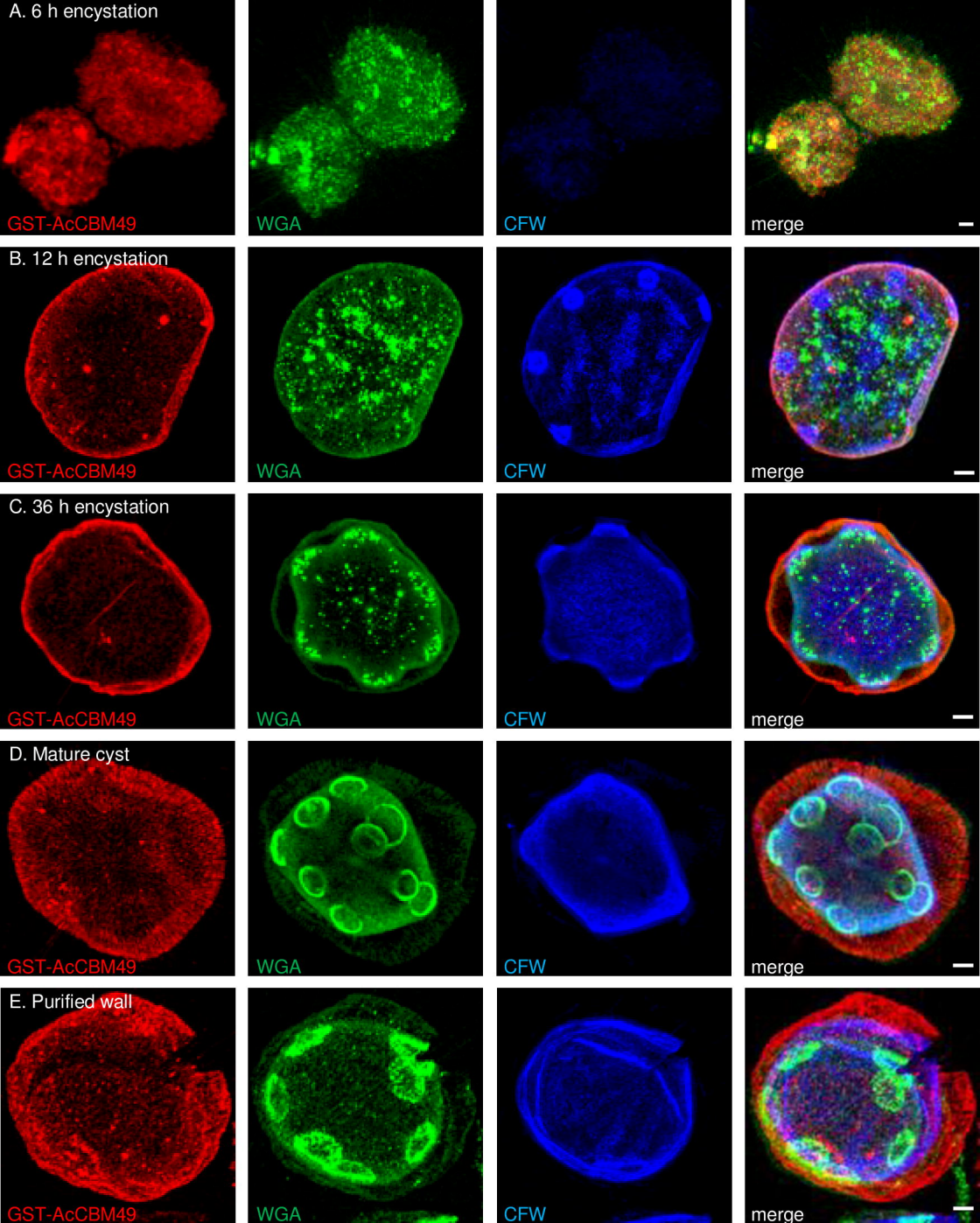
S9 Figure. *A. castellanii* Leo lectins and *E. histolytica* Jacob lectins have common structures even though they share no common ancestry (convergent evolution). Abundant cyst wall proteins of *A. castellanii* (Leo lectins) and *E. histolytica* (Jacob lectins) have unique 8-Cys or 6-Cys domains, respectively, that bind cellulose or chitin. In each protist, some of the lectins lack spacers, while others have spacers rich in Thr, Lys, and His (*A. castellanii*) or Ser and Thr (*E. histolytica*).

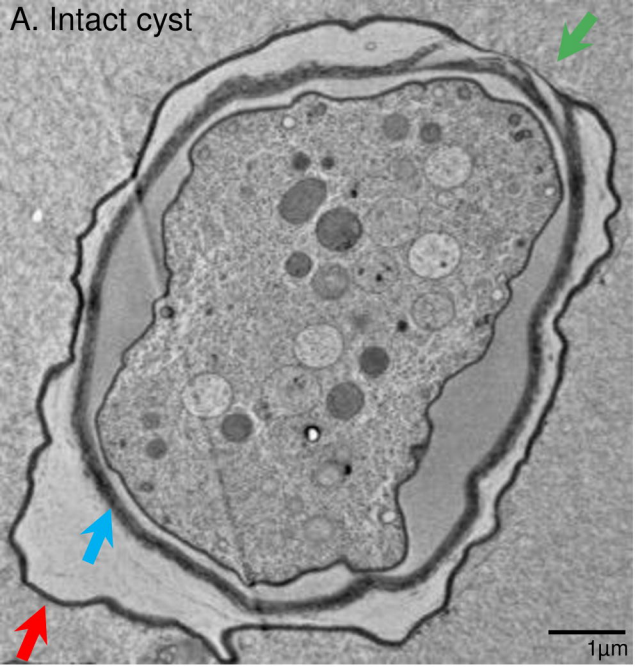
S10. Contrasting use of choice of anchor A (CAA) domains in *A. castellanii*, oomycetes, and bacteria. Jonah lectins, which are abundant in cyst walls of *A. castellanii*, have one or three CAA domains. The former are preceded by Thr-, Lys-, and Cys-rich sequences (gray), while the latter are separated by Ser- and Pro-rich spacers (blue) and hydrophobic domains (tan). CAA domains of oomycetes (*Pyromyces* or *Neocallimastix*) have three to five CAA domains, while the spore coat protein of *Bacilllus* has a single CAA domain attached to a collagen-binding domain, which is absent in *A. castellanii*.

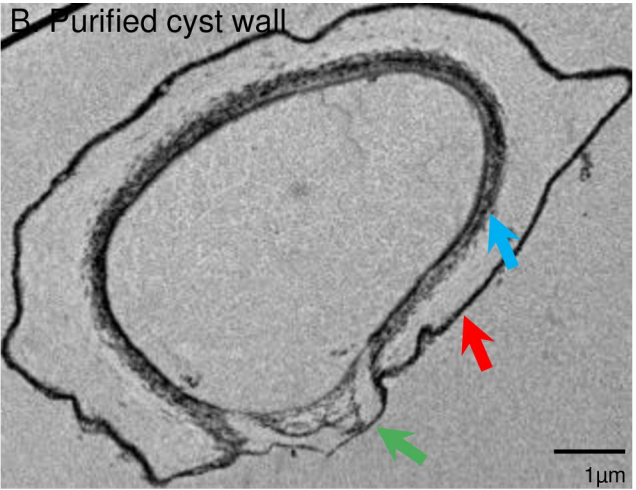
S1 Excel file. Excel file listing the most abundant proteins identified by mass spectrometry of cyst walls purified on the Percoll gradient and retained on a membrane containing 8-µm pores. Proteins with <7 unique peptides have been removed, because they are dominated by cytosolic contaminants. Luke, Leo, and Jonah lectins have been highlighted in orange. Other candidate CWPs are marked in yellow.

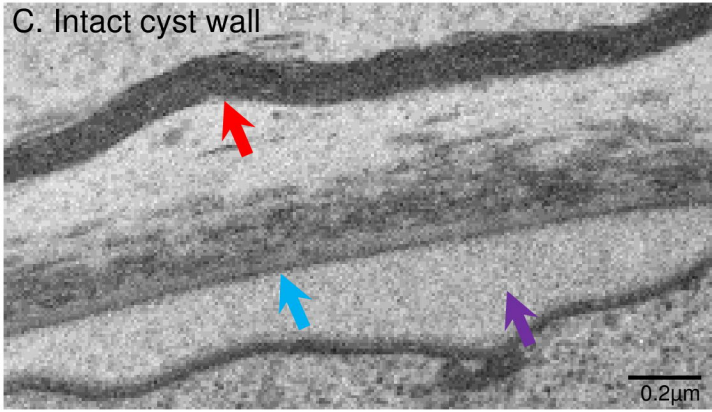
S2 Excel file. Complete list of proteins identified by mass spectrometry including a cyst wall preparation that was heavily contaminated with cytosolic proteins, because the Percoll gradient and porous membrane were not used during their purification. We have included proteins with at least two unique peptides.

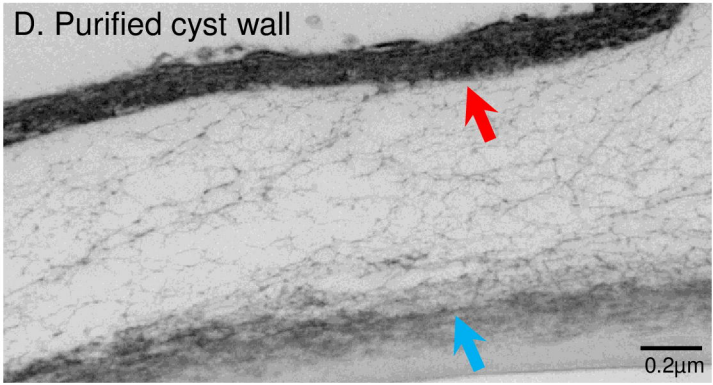
S3 Excel file. Primers used for construction of GST-, GFP-, and MBP-fusions.

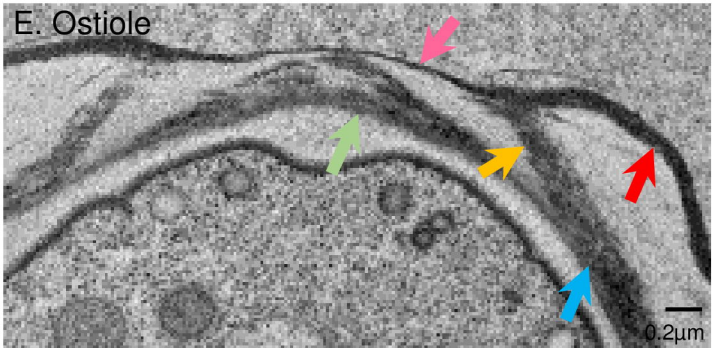

A. 6 h encystation
GST-AcCBM49
WGA
CFW
merge
B. 12 h encystation
GST-AcCBM49
WGA
CFW
merge
C. 36 h encystation
GST-AcCBM49
WGA
CFW
merge
D. Mature cyst
GST-AcCBM49
WGA
CFW
merge
E. Purified wall
GST-AcCBM49
WGA
CFW
merge


A. Intact cyst
1µm


B. Purified cyst wall
1µm


C. Intact cyst wall
0.2µm


D. Purified cyst wall
0.2µm


E. Ostiole
0.2µm

Luke(2) lectin (ACA1_377670)

MKLLFCLTVLAALIFSCASA
QSCTASLNQTRQASWVDSEQFPRSLWTVEIRNTGQQAVTNVLLSIQGSINQIWEVVLVN
DSLYRLPDWRLQVGGIPAGQSHVFGFIVNNNSTAAPVSLVSVQC
GNVTVPSPSPSASASSVPSPSASASAAPSPSAGASPSSSPRPSPSGG
CSLLASQVARSGAGGSWSDNSNRFQIYDITLNNNGASRLTQASLTIAIAADQAIAQFWNL
ERVDATNTFNIESFLLPAPGGSQSGLGYVLQTPLNSTSDGSSIGAVNFRC

CBM49

CBM49

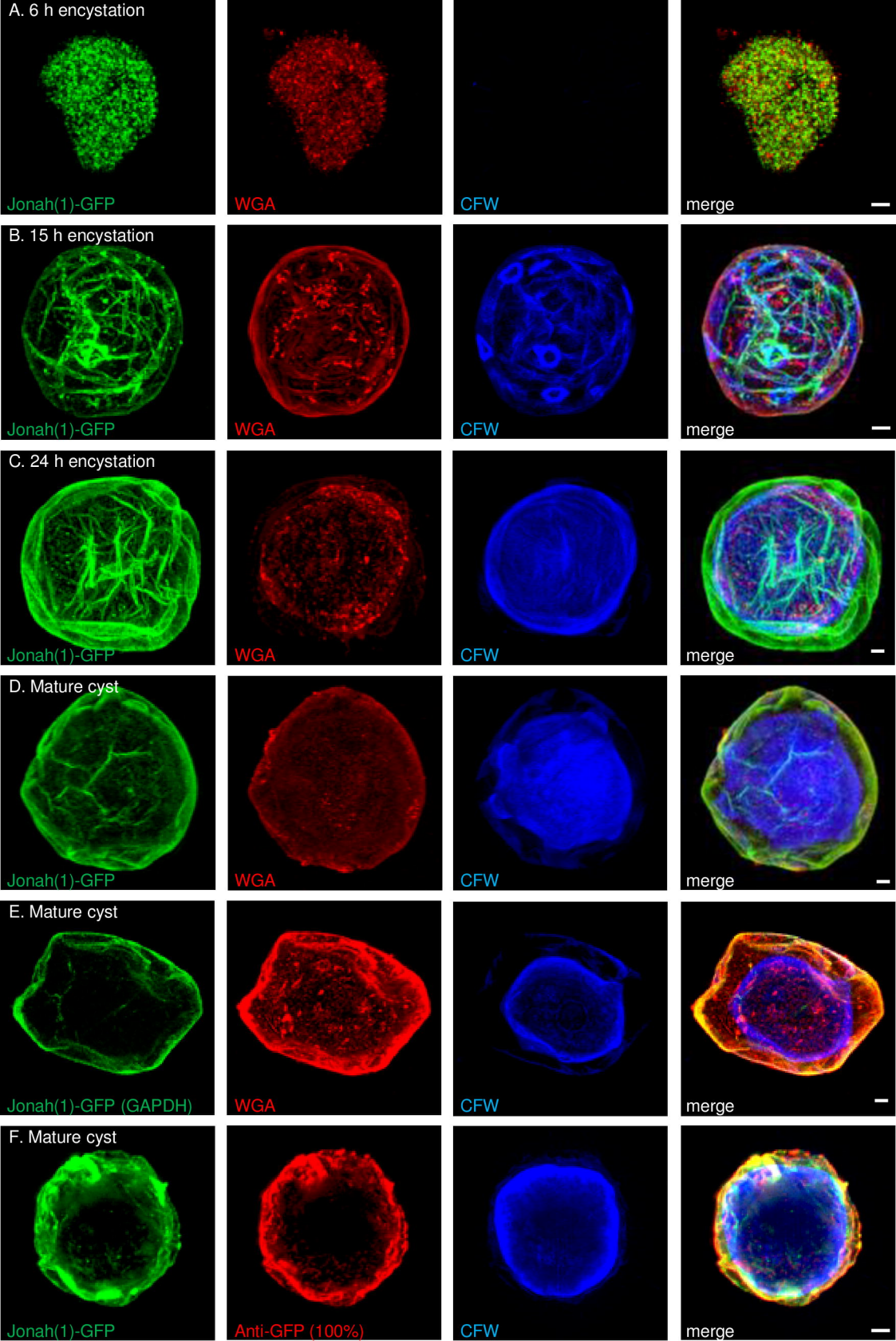
Leo lectin (ACA1_074730)

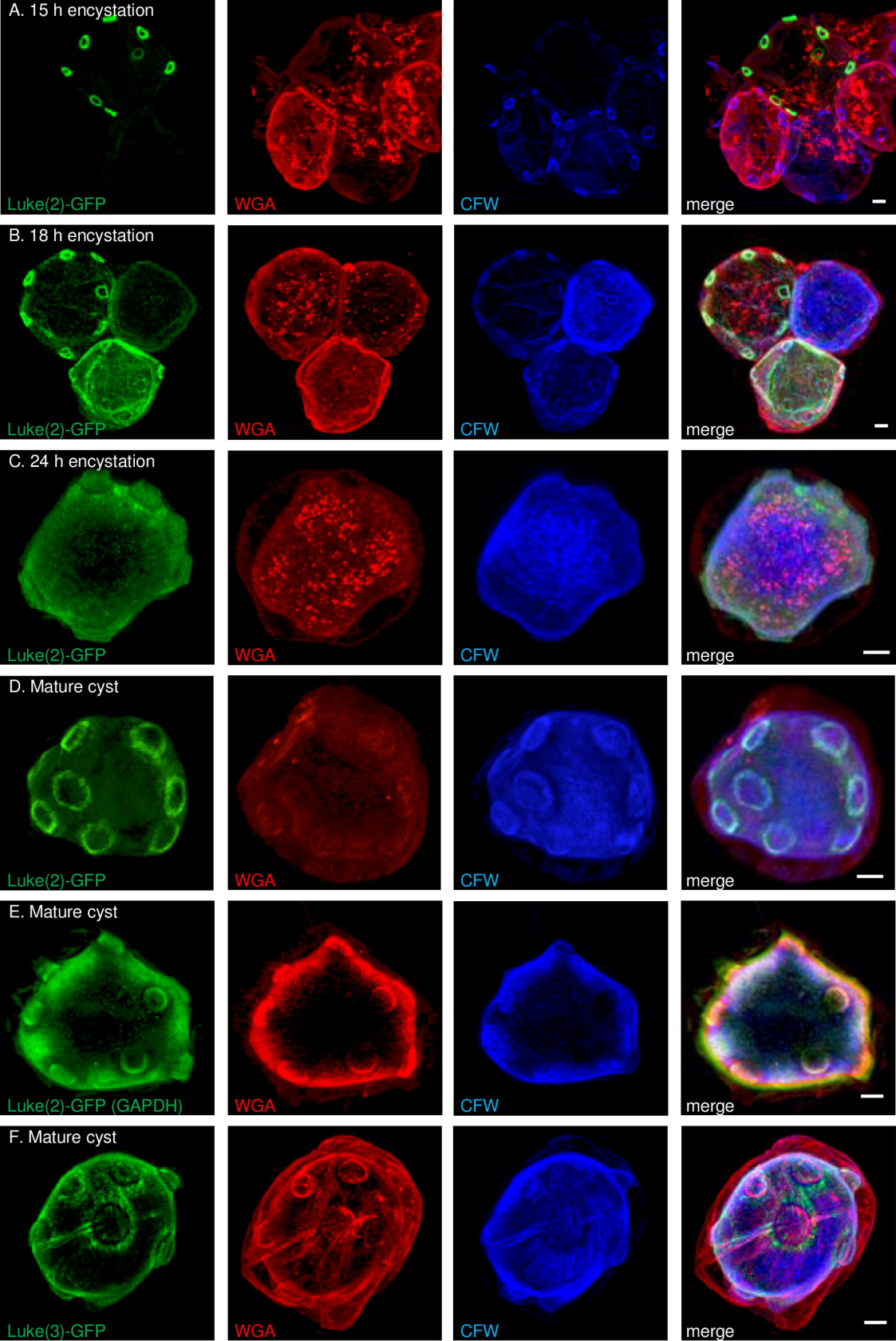
MNRFALCFFVVLCAASAAMG QR
CPGSDFFPDQRGCCPVVVNSVPFYADAQGCFPRIISGVAFYADAGGCYPSSAGVYRSAGAVCPADRVCEPECPDMAIVVAEQQGGARL
CANGFFSDVNGCCPRVLVDNVPSYRDAQGCYPISIASVIFYADVGRCYPNSSGVYRPAGAVCPAGQQCVVSCDRNYA

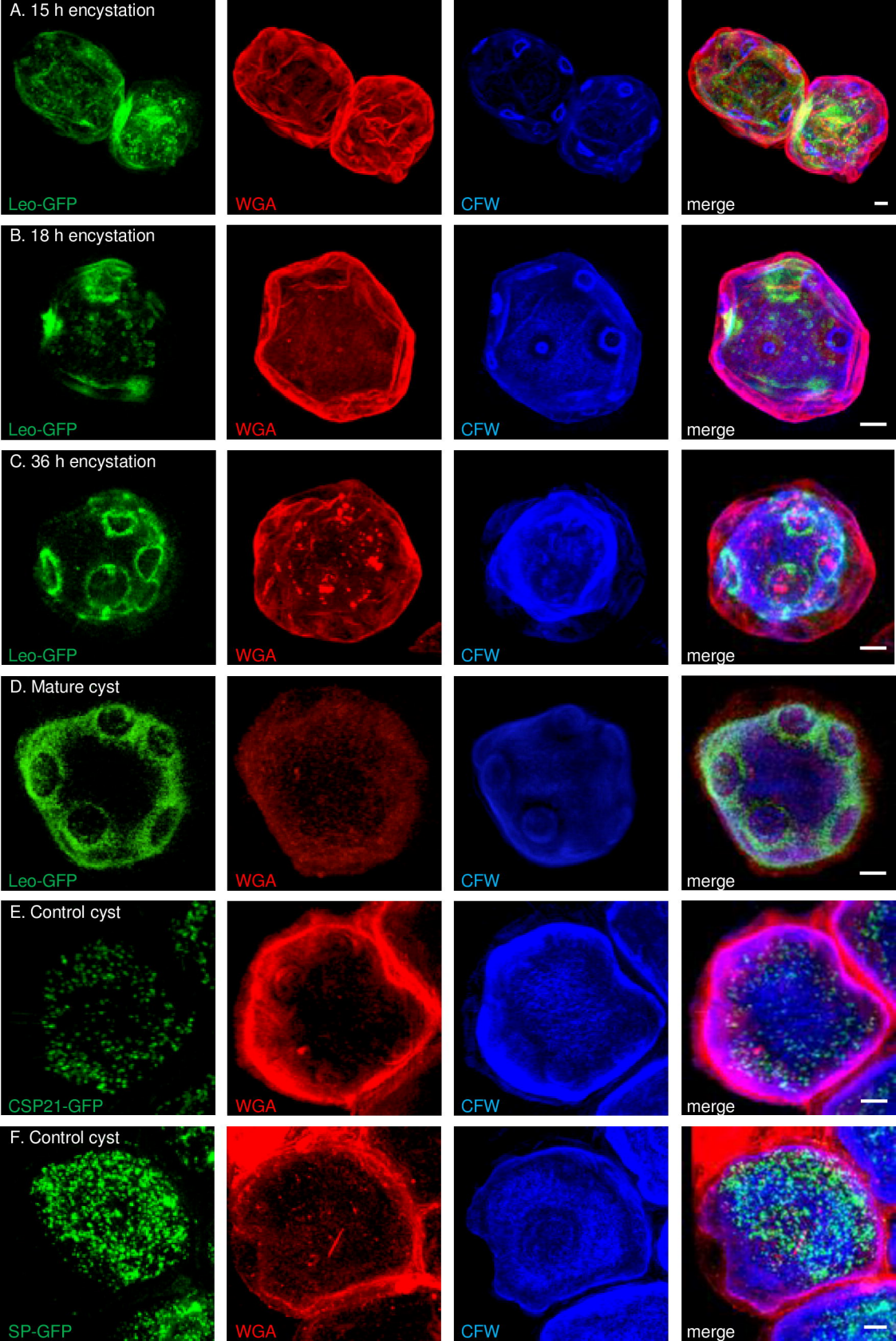
Jonah(1) lectin (ACA1_164810)

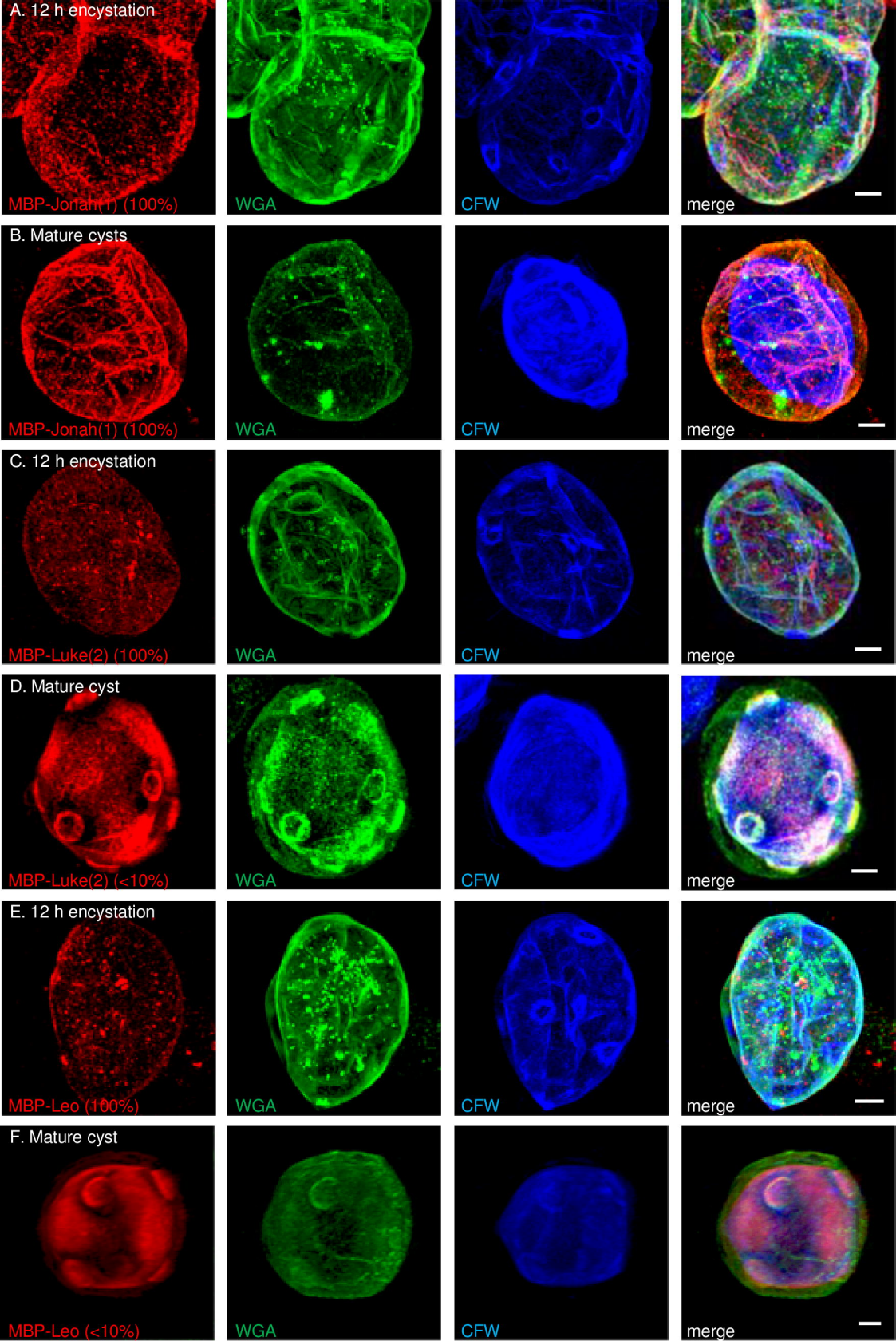
MKTTYLLLALAALLFCAVVAASA
DDDDRRGSDMNHWDDDDEGHEWDNEWEDEEDGEGRGRGQWDDGREKCRTTTKKCRTTTKKCGTTTKHCPTTTTKRCG
TTTKDCGTTTTKSCGTTTKDCGTTTMKPCTTTSKPVEKCDFAGRDNCLKASEDKFEECVDRKKCAFTVCELEAEEKF
KKCLRGRDGRDGGRENPFEREGENDRQGNERFDQCKKEQLEDKKQCLIQRKASHAGCVAAKECRDFKCLAFIKDCKP
CPTPAATTTTVRVTTH
PPTVGSSCPAEDFFGPAADLNVLVTGSDKLIYDLDMVNGDVEGRVAVNGGFRVKSFGTAQAYSCPDTANYASMFNLI
VNGKMDYSNGQLFCGSSISLSSMNDPSLRQKDFGATNARVAAGKTIKDVTGFDFVATAGYLQGVSNYFSKYAPGTHS
HFRVTEVSPADVNKLNIEGGDYVQVINIKGSGDINFSNFEIPTALNPSQVIYNVVGNNNIKISGFGLKGHLLALNSD
VVLENGHVAGNVYVRSLVGQGSGQVNLAPCP

CAA

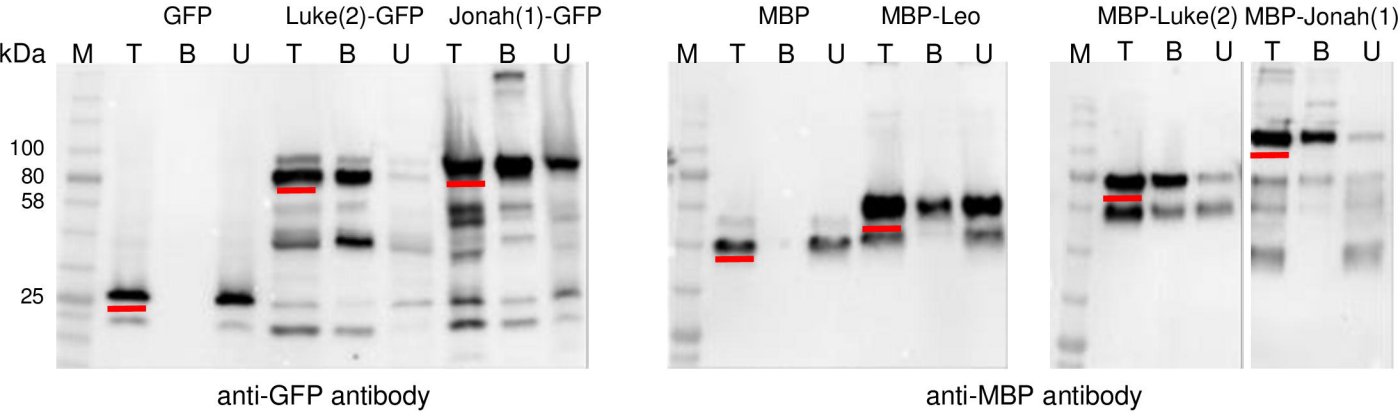
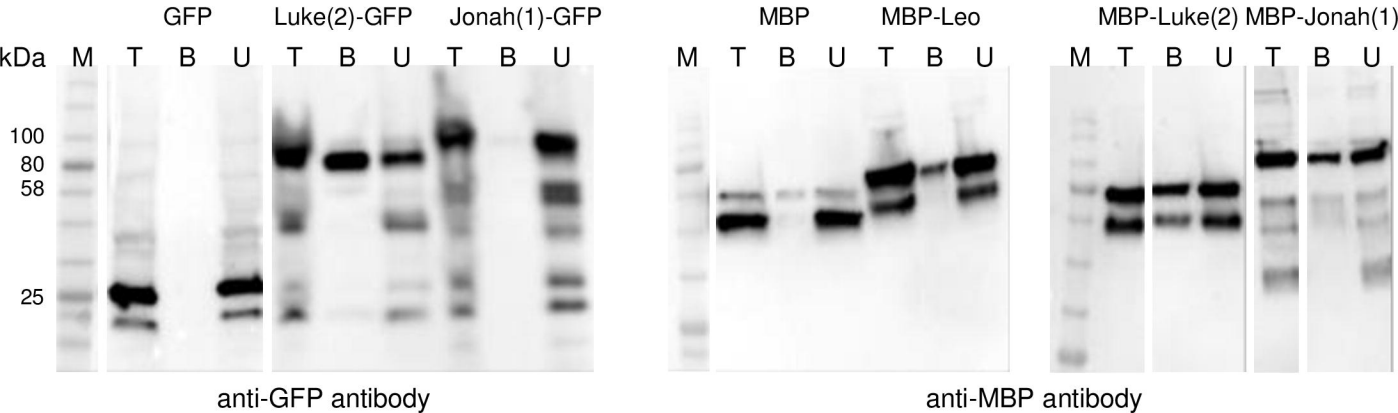

A. 6 h encystation
Jonah(1)-GFP
WGA
CFW
merge
B. 15 h encystation
Jonah(1)-GFP
WGA
CFW
merge
C. 24 h encystation
Jonah(1)-GFP
WGA
CFW
merge
D. Mature cyst
Jonah(1)-GFP
WGA
CFW
merge
E. Mature cyst
Jonah(1)-GFP (GAPDH)
WGA
CFW
merge
F. Mature cyst
Jonah(1)-GFP
Anti-GFP (100%)
CFW
merge


A. 15 h encystation
Luke(2)-GFP
WGA
CFW
merge
B. 18 h encystation
Luke(2)-GFP
WGA
CFW
merge
C. 24 h encystation
Luke(2)-GFP
WGA
CFW
merge
D. Mature cyst
Luke(2)-GFP
WGA
CFW
merge
E. Mature cyst
Luke(2)-GFP (GAPDH)
WGA
CFW
merge
F. Mature cyst
Luke(3)-GFP
WGA
CFW
merge


A. 15 h encystation
Leo-GFP
WGA
CFW
merge
B. 18 h encystation
Leo-GFP
WGA
CFW
merge
C. 36 h encystation
Leo-GFP
WGA
CFW
merge
D. Mature cyst
Leo-GFP
WGA
CFW
merge
E. Control cyst
CSP21-GFP
WGA
CFW
merge
F. Control cyst
SP-GFP
WGA
CFW
merge


A. 12 h encystation
MBP-Jonah(1) (100%)
WGA
CFW
merge
B. Mature cysts
MBP-Jonah(1) (100%)
WGA
CFW
merge
C. 12 h encystation
MBP-Luke(2) (100%)
WGA
CFW
merge
D. Mature cyst
MBP-Luke(2) (<10%)
WGA
CFW
merge
E. 12 h encystation
MBP-Leo (100%)
WGA
CFW
merge
F. Mature cyst
MBP-Leo (<10%)
WGA
CFW
merge

A. Microcrystalline cellulose



B. Chitin beads