Functional analysis of the COBRA-like family

Nadav Sorek, Andrew Ryan, Andrew Carroll, Chris Somerville

The COBRA protein is found only in higher plants and algae, and has been implicated in cellulose synthesis, but its specific function remains unknown. To shed light on COBRA function we tested the ability of putative Arabidopsis COBRA orthologs to complement mutations in the COBRA gene. We expanded the previously published phylogenetic analysis and found that COBRA genes can be grouped into five clades. Representative COBRA genes from the different clades were expressed in a cobra mutant background. Surprisingly, none of the COBRA-LIKE genes was able to complement cobra, including the closest homolog which we found to bind cellulose with the same avidity as COBRA. We hypothesize that the lack of ability of the COBRA-LIKE genes to complement the cobra mutant is due to an unidentified binding interaction that is sensitive to the structure of the COBRA-LIKE proteins. However, membrane-based yeast two hybrid assays failed to identify positive interactors for COBRA.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

Functional analysis of the COBRA-like family

Nadav Sorek[1,2]*, Andrew Ryan[1,3], Andrew Carroll[4] and Chris Somerville[1,2]

[1]Energy Biosciences Institute, University of California, Berkeley CA 94720, USA

[2] Plant and Microbial Biology Department, University of California, Berkeley CA 94720, USA

[3] Division of Biological Sciences, University of California, San Diego CA 92093, USA

[4]DNAnexus, Mountain View, CA 94040, USA

*Corresponding author - nadavsorek@berkeley.edu

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

Abstract

The COBRA protein is found only in higher plants and algae, and has been implicated in cellulose synthesis, but its specific function remains unknown. To shed light on COBRA function we tested the ability of putative Arabidopsis *COBRA* orthologs to complement mutations in the *COBRA* gene. We expanded the previously published phylogenetic analysis and found that *COBRA* genes can be grouped into five clades. Representative *COBRA* genes from the different clades were expressed in a *cobra* mutant background. Surprisingly, none of the *COBRA-LIKE* genes was able to complement *cobra*, including the closest homolog which we found to bind cellulose with the same avidity as COBRA. We hypothesize that the lack of ability of the *COBRA-LIKE* genes to complement the *cobra* mutant is due to an unidentified binding interaction that is sensitive to the structure of the COBRA-LIKE proteins. However, membrane-based yeast two hybrid assays failed to identify positive interactors for COBRA.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

Introduction

Cellulose synthesis in plants is catalyzed at the plasma membrane by the multimeric cellulose synthase (CESA) complex (McFarlane et al., 2014). The CESA gene family has eleven members in Arabidopsis (Richmond and Somerville, 2000). The stoichiometry of the CESA complex remains unknown, but expression pattern and mutant phenotypes suggest that CESA4, CESA7 and CESA8 compose the CESA complex for cellulose synthesis in secondary cell walls (Carroll et al., 2012; Li et al., 2014), while CESA1, CESA3 and four partially redundant CESA6-like genes are the core genes for cellulose synthesis in primary cell walls (Beeckman et al., 2002). Several CESA genes have unique expression patterns, and were shown to be involved in cellulose synthesis in a particular tissue, e.g. CESA2 in seed mucilage (Mendu et al., 2011). Along with the cellulose synthase complex, there is also a number of accessory proteins that are required for normal cellulose synthesis (Persson et al., 2005). Most of the accessory proteins have unknown roles in cellulose synthesis. An exception is cellulose synthase-interactive protein 1 (CSI1) (Gu et al., 2010) which was found to connect the cellulose synthase complexes to cortical microtubules (Li et al., 2012).

One of the genes that is highly co-expressed with the cellulose synthase complex and has been implicated to be involved in cellulose crystallization is *COBRA* (Roudier et al., 2005). *cobra* mutants were originally identified in a screen for root defects (Benfey et al., 1993), and were shown to have cell wall defects (Roudier et al., 2005; Schindelman et al., 2001). COBRA is localized at the plasma membrane and binds individual glucan chains, as well as cellulose (Liu et al., 2013). However, except for a carbohydrate binding domain, the COBRA protein lacks recognizable motifs and the mechanism by which COBRA facilitates cellulose formation remains unknown. In Arabidopsis, the COBRA family contains eleven members (Roudier et al., 2002), and additional *COBRA-LIKE* genes (*COBL*) have been identified in several others dicots and monocots (Brady et al., 2007; Cao et al., 2012; Sindhu et al., 2007). Phylogenetic analysis of the plant *COBRA* genes has been done with a small number of species and limited only to monocots and dicots (Niu et al., 2015; Zhang et al., 2010).

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

Materials and Methods

Sequence identification

Sequences were obtained from Phytozome v10.2 by blasting the AtCOBRA (At5g60920.1) sequence against the following genomes: Physcomitrella patens genome V3.0 , Selaginella moellendorffii genome v1.0, Populus trichoderma genome v3.0 Prunus persicum genome v1.0, Solanum lycopersicum genome iTAG2.3, Arabidopsis thaliana genome TAIR10, Oryza sativa genome v 7.0, Zea mays genome 6a, Musa acuminata early release genome v1, Amborella trichopoda early release genome v1.0. Picia abies sequences were identified using a pBLAST search of the AtCOBRA sequence against the Picia abies 1.0 genome (Z4006) database of high confidence genes at congenie.org. Phoenix dactylifera sequences were taken from the Phoenix dactylifera Annotation Release 100 from the NCBI Eukaryotic genome pipeline. Charophyte sequences were obtained by creating a BLAST database using shotgun transcriptome assemblies from NCBI : GBSK01000000(Mesostigma viride NIES995), GBSO01000000 (Klebsormidium flaccidum UTEX321), GBSL01000000 (Coleochaete orbicularis UTEX2651), GBST01000000 (Nitella mirabilis S040), and GBSM01000000 (Spirogyra pratensis UTEX928) and then using the AtCOBRA sequence to tBLASTn against each assembly. No sequences with an E value <=0 were found in the Mesostigma viride assembly (Ju et al., 2015).

Phylogenetic analysis

COBRA family sequences lacking the CCVS motif or otherwise severely truncated were removed before alignment. An alignment was done using the MEGA6.06 program and its MUSCLE alignment algorithm, under the default settings. A maximum likelihood tree was then constructed using default settings, and a phylogeny test of 100 bootstrap replicates was performed. This was repeated multiple times and the tree shown is representative of the trees produced.

Molecular cloning

All plasmids are listed in Table S1, Gateway Cloning was performed according to the manufacturer's instructions (Invitrogen). *COBRA* and *COBL1* were cloned into pDES15 to create *GST-COBRA* (pNS110) and *GST-COBL1* (pNS111). cDNA of *COBRA, COBL1, COBL5, COBL9, COBL11* and Physcomitrella patens *COBL1* were subcloned into pDONOR P2R-P3, and then used to create the expression vector p35S::COB:NOS in pH7 mu34GW (plasmids NS97 and NS90-949) by multi gateway reaction.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

Plant growth

Arabidopsis ecotype Columbia seeds and various mutant lines were sterilized and germinated on Murashige and Skoog (MS) plates (one-half-strength MS salts, 0.8% agar, and 0.05% MES, pH 5.7). Seedlings were then grown vertically at 22°C under continuous light or in dark for 5 days. For phenotype characterization 50 seedlings from each line were measured for either root length or hypocotyl length.

Protein purification and blotting

E. coli DH5α was used for DNA propagation. For protein expression, pNS110 and pNS111 were transformed into E. coli BL21 CodonPlus DE3 RIL (Stratagene) cells. Cells were grown to OD600 of 0.6. Protein expression was induced by adding 0.2% (W/V) L-Arabinose. Protein extraction and purification was carried out with AKTA-prime (GE Healthcare, UK) using GSTrap HP columns according to the manufacturer's instructions (GE Healthcare, UK).

Anti-GST antibodies (Sigma G7781) were used at a dilution of 1:3,000 together with blotting grade horseradish peroxidase-conjugated goat anti-rabbit immunoglobulin G (Bio-Rad, Hercules, CA). Detection was performed with SuperSignal West Pico Chemiluminescent Substrate (Pierce, cat#15168).

Avicel binding assay

Experiments were done based on (Walter and Schrempf, 2003). Proteins (100 ug) were incubated with 1mg/ml of Avicel PH-101 (Sigma 11365), or as a negative control with 1 mg/ml Polyvinylpolypyrrolidone (PVPP - Sigma 77627) in 1 ml of 20 mM Tris-Cl buffer (pH 8) and 150 mM NaCl for 120 min at 22°C with slow shaking. Avicel recovered by centrifugation was washed 3 times with 20 mM Tris-Cl buffer (pH 8) and 150 mM NaCl. Matrix-bound proteins were released by heating in SDS sample buffer for 5 min at 100 °C and subjected to SDS-PAGE.

qRT-PCR

Total RNA was extracted from homogenized tissue frozen in liquid nitrogen and digested with DNAse (Cat# 79254, Qiagen), and 1 µg RNA/20 µL reaction was used to generate first-strand cDNA using Superscript II Reverse Transcriptase (Invitrogen) following the manufacturer's protocol. For qRT-PCR experiments, cDNA was obtained as described above and 1 ul was used to analyze gene expression using SYBR greenER qPCR supermix (Life Technologies) and the following PCR conditions: 50°C for 2min, 95°C for 10 min, 40 cycles of (95°C for 15 s, 59°C for 30 s and 68°C for 45 s, followed by a fluorescence reading). Housekeeping control ribosomal RNA 60S (Walley et al., 2007) was amplified in parallel on each plate for normalization. "No template" controls and melting curves were examined to ensure against contamination and

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

primer-dimer formation. The relative starting quantities of each gene were determined by the ΔΔCT method, as described in (Hietala et al., 2003). Unless otherwise noted, primers were designed using online tool ATRTPrimers (Han and Kim, 2006).

Yeast two hybrid

Yeast two hybrid was done using the *COBRA* coding sequence lacking the signal peptide sequence and the GPI anchor sequence. Yeast two hybrid screen was performed using cDNA library from Arabidopsis seedlings with the split ubiquitin method. The screen was performed by Hybrigenics (http://www.hybrigenics.com)

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

Results

COBRA gene family

To further explore the plant COBRA family, we created a phylogenetic tree (figure 1) by blasting *AtCOBRA* against the translation products of the genomes of the following organisms: Division Charophyta: Mesostigma viridae, Klebsormidium flaccidum, Coleochaete orbicularis, Nitella mirabilis, Spirogyra pratensis. Division Bryophyta: Physcomitrella patens. Division Lycopodiophyta: Selaginella mollendorfi. Gymnosperms: Picia abies. Basal flowering Plants: Amborella trichopoda. Clade Eudicots: Arabidopsis thaliana, Populus trichocarpa, Solanum lycopersium, Prunus persica. Clade Monocots: Oryza sativa, Zea mays, Phoenix dactylifera, Musa acuminata.

COBRA genes can be identified in the charophyte lineage of algae. While the most ancestral charophyte, Mesostigma viride, does not seem to express any *COBRA-LIKE* genes (Chuanli Ju, 2014), a pair of *COBRA* genes appears in Klebsormidium, which expands to three *COBRA-LIKE* genes by the divergence of Spirogyra. In order to determine the diversity of the COBRA family, we constructed a large phylogenetic tree of all the COBRA amino acid sequences from the foregoing list of species. Before alignment, sequences that were abnormally truncated or of extremely different lengths were removed. Since this included COBL5, we checked whether there was any 'truncated COBRA-LIKE' grouping the shorter sequences aligned to, but as none was found, they were considered as products of pseudogenes or poor annotations. Sequences were aligned in the MEGA program using the MUSCLE algorithm under standard conditions, and a maximum likelihood tree was constructed with 100 bootstraps.

There are five major clades of COBRA-LIKE gene products (figure 1), with every flowering plant species having at least one representative in each. Group I contains AtCOBRA, AtCOBL1, and AtCOBL2. Group II contains AtCOBL4. Group III contains AtCOBL6, and Group IV contains AtCOBL7,8,9. Group V contains AtCOBL10 and AtCOBL11. These results are in agreement with previous work, showing that the COBRA family consists of five clades (Zhang et al., 2010). Our analysis shows that expanding the tree with more representatives of monocots and dicots, as well as to include more basal organisms, maintains the five clade distributions.

Complementation tests of the Arabidopsis COBRA-LIKE genes

Different members in the COBRA family have a unique expression pattern, with very little overlapping expression (Brady et al., 2007; Cao et al., 2012; Gao et al., 2013; Zhang et al., 2010). To assess whether the dramatic phenotype of the *cobra* mutant can be rescued by members of the *COBRA-LIKE* (*COBL*) family, cDNA sequences of *COBL* genes were placed under transcriptional control of the 35S promoter to overcome tissue specificity. Since COBRA sequences have a secretion peptide at the N-terminus and a GPI anchoring motif at the C-

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

terminus, all transgenes were expressed with no additional tags. In all cases, the same plasmid backbone was used to express the transgene. We tested AtCOBL1 since it is the closest Arabidopsis homolog to COBRA. We also tested AtCOBL5 due to its novel truncated structure amongst the COBRA family. To assess the diversity in the lineage, we tested the two most divergent Arabidopsis homologs, AtCOBL9 and AtCOBL11. As a representative of the basal COBRA, we tested PpCOBL8 from Physcomitrella patens. We also tested AtCOBRA as a positive control for complementation.

Complementation was determined by measuring root length in light grown seedlings (Figure 2, upper panel) and hypocotyl length in dark grown seedlings (Figure 2, lower panel). At least ten transgenic lines were recovered and tested for each construct. qRT-PCR was performed on at least three transgenic lines for each construct. Expression of *AtCOBRA* under transcriptional control of the 35S promoter successfully complemented the *cobra* (*cob-6*) phenotype, restoring both root length and hypocotyl length to WT levels. Expression of *AtCOBL1* complemented the *cobra* phenotype slightly (supplementary figure 1 for statistical analysis). Expression of the other *COBRA-LIKE* genes, *AtCOBL5, AtCOBl9, AtCOBL11* and *PpCOBL8* failed to complement the *cobra* phenotype. To verify the expression of the transgenes we analyzed the expression levels by qPCR (supplementary figure 1). In all lines the transcript levels of the transgene increased significantly compared to the expression of the endogenous gene in the *cobra* background (supplementary figure 1). Since the cDNAs of the transgene were cloned with no tags, it was not possible to verify protein expression.

COBRA-cellulose binding

The results of the complementation experiments suggest that the sequence variation within the COBRA family in Arabidopsis has functional significance. COBRA was previously reported to harbor a cellulose binding motif (Sato et al., 2010). To determine if the inability of *AtCOBL1* to fully complement cobra phenotype is related to cellulose binding capacity, we performed a cellulose binding assay for AtCOBRA and AtCOBL1 (Figure 3). GST-COBRA and GST-COBL1 were expressed in E. coli and were partially purified. Binding to avicel (crystalline cellulose powder) was tested based on a pull down assay. Protein samples were incubated with avicel for 2 hours at 22°C. The sample was then washed three times and eluted with 1% SDS at 95°C for 5 minutes. Gels were blotted using anti-GST antibodies, and affinity was estimated based on the ratio between the intensity of the band with correct size in the total fraction compared to that in the pellet fraction. Both COBRA and COBL1 showed approximately the same affinity. This correlates with high sequence similarity of the putative cellulose binding motif (CBM) of AtCOBRA with AtCOBL1 CBM (91% similarity). More generally, the CBM in the COBRA gene family is located in the N-terminus of the gene, a region that is highly conserved within the family tree.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

No COBRA interacting proteins can be found

COBRA has no known interacting proteins. It is still unclear whether COBRA directly interacts with the cellulose synthase complex, or if other proteins are essential for COBRA function. We employed the company Hybrigenics to perform a split-Ubiquitin-based membrane Two-Hybrid screen (both LexA and Gal4) in yeast, in an attempt to identify proteins that physically interact with COBRA. We used the COBRA coding sequence, excluding the signal peptide and GPI anchor motif, as a bait, and cDNA library from Arabidopsis seedlings. Despite using the split-Ubiquitin-based membrane Two-Hybrid screen, and a cDNA library that had been validated in other screens by the company, no interactors were found except for AT5G08720, which is a known artifact of the system. COBRA was shown to be N-glycosylated (Roudier et al., 2005). It is possible that this post translational modifications that are missing in the split-Ubiquitin-based membrane Two-Hybrid screen are important for proper folding of the protein or for protein-protein interactions.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

Discussion

Phylogenetic analysis

Our extended phylogenetic analysis confirmed that the COBRA family consists of five clades (Zhang et al., 2010). Our analysis is the most comprehensive analysis of the COBRA family so far, based on 17 species that range from unicellular algal charophytes to monocots and dicots. Therefore, we believe that the division of the COBRA family into five distinct clades will not change by adding more organisms. We did not find *COBRA-LIKE* genes in the unicellular algae Mesostigma viride, but did find *COBRA-LIKE* genes in all the multicellular charophytes analyzed from Klebsormidium to Spirogyra (figure 1).

COBRA primary structure

COBRA protein structure is highly conserved. With few exceptions, COBRA-LIKE proteins have a signal peptide at the N terminus of the gene and GPI anchoring motif at the C terminus, both of which presumably target the protein into the outer leaflet of the plasma membrane. It is hard to predict whether the *COBRA* genes that encode proteins lacking any of these motifs are functional. The protein N terminus contains a CBM motif, which shares high sequence similarity within the different clades. In many cases the GPI anchor motif is cleaved after the protein has been targeted to the outer leaflet of the plasma membrane (Butikofer et al., 2001). Therefore, it is possible that the ability to bind cellulose is important for proper localization of COBRA at the plasma membrane-cell wall connection.

There is a short (~20 amino acids) variable region between the conserved CBM motif at the N terminus and the conserved C terminal region. It is not known whether this short motif has an independent function or whether it simply serves as a linker between the N and C termini motifs. A transgenic line with a mCherry inserted in the variable region failed to complement *cobra* mutant (supplementary figure 2). Although this result does not help to distinguish if this variable region has a specific function or not, it might suggest that the CBM and the C terminus motif needs to be in close proximity.

The C terminus is also conserved within the five different clades. However, there is no assigned function for the C terminus. There is a short CCVS motif at the C terminus, which was suggested to be involved in metal ion binding or in formation of disulfide bonds (Roudier et al., 2002). However, there is no sequence similarity that can help predict the domain function.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

Interpretation of the failure to complement

Taken at face value, the inability of the *COBL* genes to complement a cobra mutation suggests that the various *COBL* genes have divergent functions. In view of the high degree of primary sequence conservation between COB and some COBL proteins it seems unlikely that the COBL proteins are catalyzing different chemical reactions. It seems more likely that the functional divergence is related to the binding of COB and the COBL proteins to structurally diverse ligands. Since the N-terminal part of the COBL proteins appears to be a CBM, and since at least one of the COBL CBMs and COB bind cellulose, it seems possible that the functional divergence of the COBL proteins is mostly or entirely due to the C-terminal part of the proteins.

The obvious caveat is that we were unable to test whether the COBL proteins are produced in the recombinant plants because no protein tags were used so as to not risk impairing the function of the COBL proteins. The mRNA was produced, so the *COBL* genes were expressed in the plants used for the complementation tests. Thus, although some uncertainty remains, we suggest that future efforts to establish the function of the COBRA family of proteins should entertain the idea that the different members of the family have distinct functions.

The search for COBRA interactors

Since cobra is thought to act at the plasma membrane outer leaflet, it is possible that cobra interactors (if they exist) may be tightly bound to other cell wall polymers and/or plasma membrane as well. Therefore, a problem with trying to identify interactors through a pull-down assay, may be solubilizing the relevant fraction. Abundance is also an issue, as COBRA and CESA proteins are not readily detected by proteomics methods (Borner et al., 2003; Nuhse et al., 2004). In light of these potential problems, we performed a membrane based yeast two hybrid assay. However, the screen failed to identify positive interactors. Toxicity tests for cobra in yeast scored negative. Therefore, it might be worthwhile to consider the use of a membrane based yeast two hybrid assay. Since the biochemical function of the CBM is known, it might be useful to design a screen with only the C terminus of the protein, or other short peptides at the N terminus which are not part of the CBM.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

Conclusions

Constitutive expression of *COBRA-LIKE* genes, including *COBL1* which encodes a protein that is 83% similar to COBRA, failed to complement the *cobra* phenotype in transgenic plants. The inability of *COBL* genes to complement the *cobra* mutants was unexpected. The high similarity in the CBM sequence across the COBRA family, together with the results shown in figure 3 demonstrating that COBL1 binds cellulose as well as COBRA, suggests that the inability of the COBL members to complement the cobra phenotype is probably not due to differences in the ability to bind cellulose. It seems likely that the apparent functional diversity is related to some sort of binding interaction that is relatively sensitive to the structure of the COBL proteins. Since COBL1 binds cellulose with similar avidity as COB, we infer that the COBL proteins also bind something else which exhibits structural variability. However, a two-hybrid screen failed to identify candidates, suggesting that specialized screens may be necessary to identify the putative COBRA interaction partners.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

Acknowledgements

We thank Heidi Szemenyei, Trevor Yeats and Tamara Vellosillo for helpful advice and for their comments on the manuscript. We thank Valerie Lavenburg for excellent technical support. This work was supported, in part, by the Energy Biosciences Institute and the Philomathia Foundation. NS was the recipient of Postdoctoral Award number FI-434-2010 from the Binational Agricultural Research and Development Fund.

Figures and figure legends

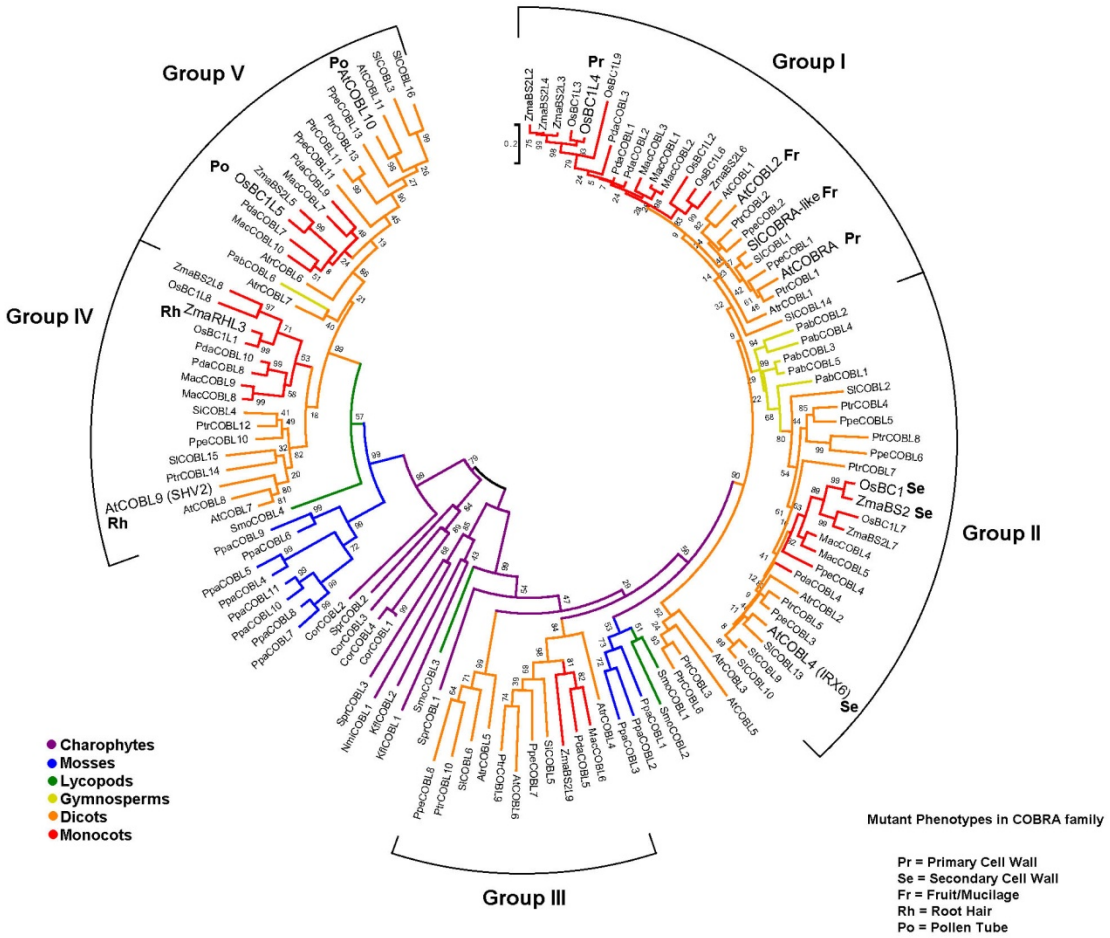

Figure 1 – Phylogenetic tree of the COBRA family. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions with less than 90% site coverage were eliminated. Evolutionary analyses were conducted in MEGA6 (Tamura et al., 2013). Group I contains two genes whose absence causes defects in primary cell wall assembly, *AtCOBRA* in Arabidopsis thaliana (Roudier et al., 2005), and *OsBC1L4* in rice (Dai et al., 2011). The patterns of expression of both of these genes are highly correlated with those of primary wall-forming CESAs. In Group I there are also two genes that have been shown to have an effect on the seeds and fruits of the plant, *AtCOBL2* in Arabidopsis, which causes defects in seed mucilage due to malformation of cellulosic rays (Ben-Tov et al., 2015), and *SlCOBRA-LIKE* in Tomato, which is required for proper fruit development, and, when overexpressed, improves the longevity of the tomato fruit (Cao et al., 2012). In

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

Group II, there are three genes known to be involved in secondary cell wall synthesis, *AtCOBL4* (Taylor-Teeples et al., 2015), *OsBC1* (Li et al., 2003), and *ZmBS2* in maize (Sindhu et al., 2007). Group IV contains *AtCOBL9*, also known as *SHAVEN2*, which is defective in root hair growth (Jones et al., 2006), as well as *ZmRHL*, another *COBRA* family gene that fails to develop root hairs (Hochholdinger et al., 2008). Group V has two genes required for pollen tube growth, *AtCOBL10* (Li et al., 2013) and *OsBC1L5* (Dai et al., 2009).

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

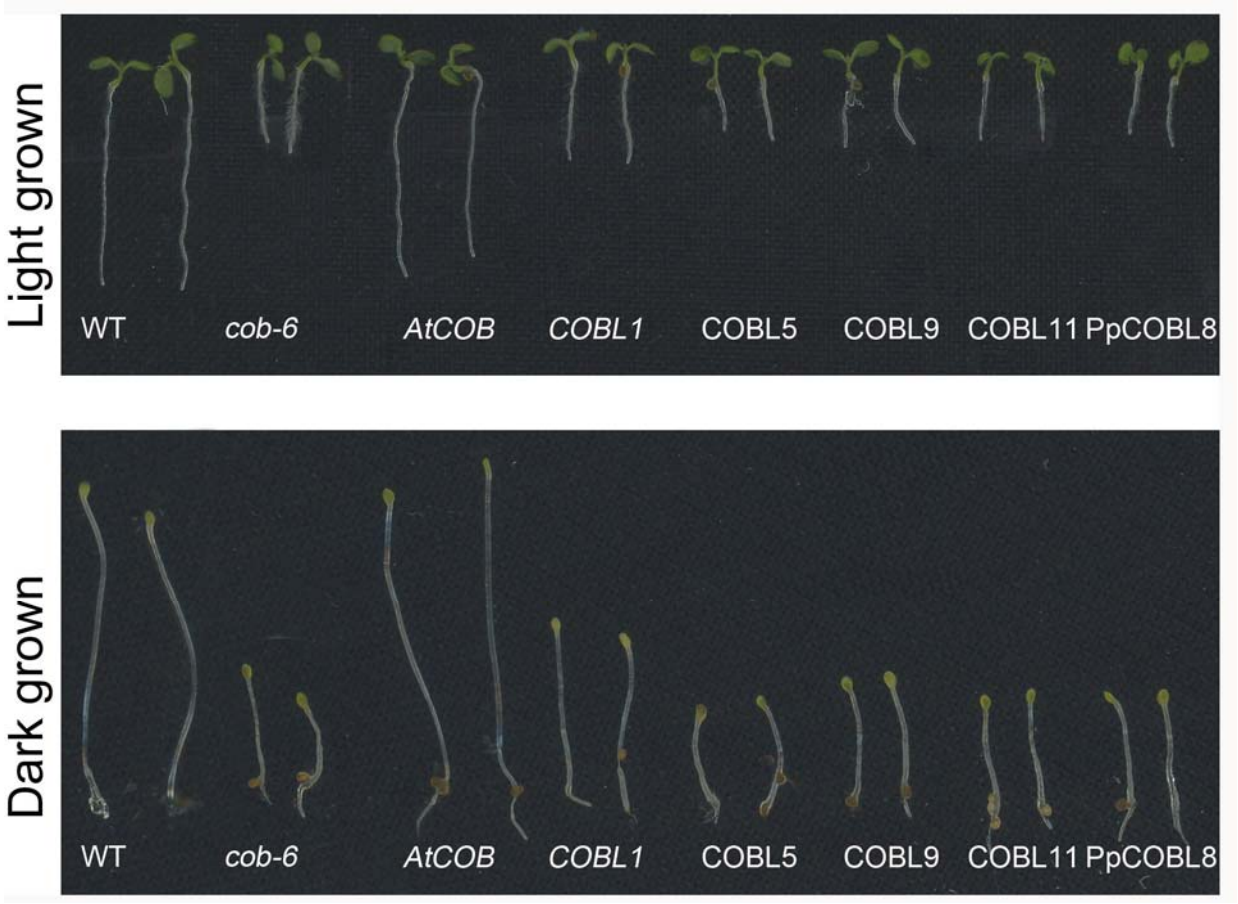

Figure 2 – Overexpression of COBRA family members in cobra background. For all genes, the cDNA was expressed under the 35S promoter. Expression of *AtCOBRA* (*AtCOB*) successfully complemented the *cobra* (cob-6) phenotype. Expression of *AtCOBL1* slightly complements *cobra*, while all other *COBL* genes from Arabidopsis and from Physcomitrella patens (*PpCOBL8*) failed to complement cobra. See supplementary figure 1 for statistical analysis.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

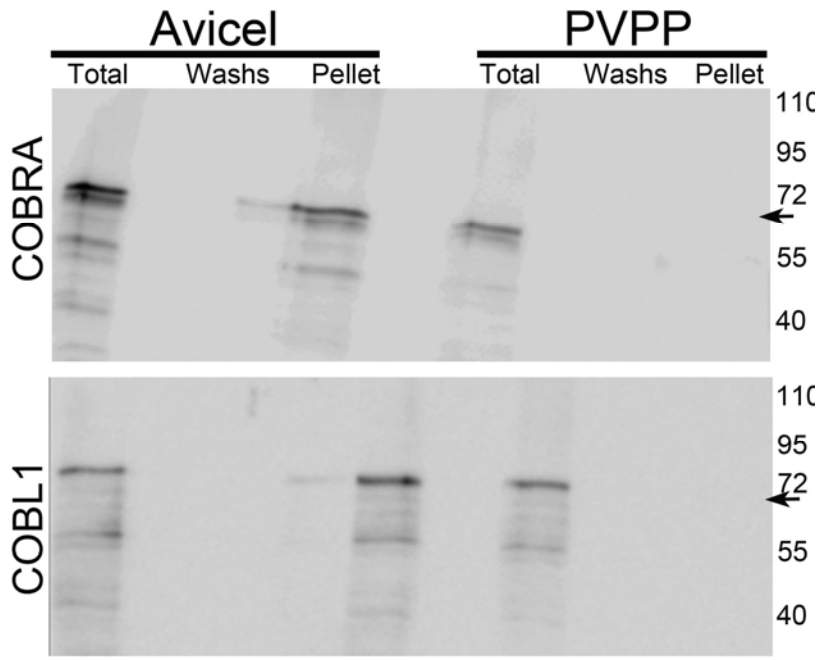

Figure 3 – Cellulose binding assay for COBRA and COBL1. GST-COBRA and GSTCOBL1 were tested for binding to Avicel (crystalline cellulose powder). Assays were based on pull down. Both COBRA and COBL1 show the same estimated affinity for cellulose binding. As a negative control we used Polyvinylpolypyrrolidone (PVPP) to test for nonspecific hydrophobic interactions.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

References

1. Beeckman, T., Przemeck, G.K., Stamatiou, G., Lau, R., Terryn, N., De Rycke, R., Inze, D., and Berleth, T. (2002). Genetic complexity of cellulose synthase a gene function in Arabidopsis embryogenesis. Plant Physiol *130*, 1883-1893.
2. Ben-Tov, D., Abraham, Y., Stav, S., Thompson, K., Loraine, A., Elbaum, R., de Souza, A., Pauly, M., Kieber, J.J., and Harpaz-Saad, S. (2015). COBRA-LIKE2, a member of the glycosylphosphatidylinositol-anchored COBRA-LIKE family, plays a role in cellulose deposition in arabidopsis seed coat mucilage secretory cells. Plant Physiol *167*, 711-724.
3. Benfey, P.N., Linstead, P.J., Roberts, K., Schiefelbein, J.W., Hauser, M.T., and Aeschbacher, R.A. (1993). Root development in Arabidopsis: four mutants with dramatically altered root morphogenesis. Development *119*, 57-70.
4. Borner, G.H., Lilley, K.S., Stevens, T.J., and Dupree, P. (2003). Identification of glycosylphosphatidylinositol-anchored proteins in Arabidopsis. A proteomic and genomic analysis. Plant Physiol *132*, 568-577.
5. Brady, S.M., Song, S., Dhugga, K.S., Rafalski, J.A., and Benfey, P.N. (2007). Combining expression and comparative evolutionary analysis. The COBRA gene family. Plant Physiol *143*, 172-187.
6. Butikofer, P., Malherbe, T., Boschung, M., and Roditi, I. (2001). GPI-anchored proteins: now you see 'em, now you don't. Faseb Journal *15*, 545-548.
7. Cao, Y., Tang, X., Giovannoni, J., Xiao, F., and Liu, Y. (2012). Functional characterization of a tomato COBRA-like gene functioning in fruit development and ripening. BMC Plant Biol *12*, 211.
8. Carroll, A., Mansoori, N., Li, S., Lei, L., Vernhettes, S., Visser, R.G., Somerville, C., Gu, Y., and Trindade, L.M. (2012). Complexes with mixed primary and secondary cellulose synthases are functional in Arabidopsis plants. Plant Physiol *160*, 726-737.
9. Chuanli Ju, B.V.d.P., Endymion D. Cooper, James H. Thierer, Theodore R. Gibbons, Charles F. Delwiche & Caren Chang (2014). Conservation of ethylene as a plant hormone over 450 million years of evolution. Nature Plants.
10. Dai, X.X., You, C.J., Chen, G.X., Li, X.H., Zhang, Q.F., and Wu, C.Y. (2011). OsBC1L4 encodes a COBRA-like protein that affects cellulose synthesis in rice. Plant Mol Biol *75*, 333-345.
11. Dai, X.X., You, C.J., Wang, L., Chen, G.X., Zhang, Q.F., and Wu, C.Y. (2009). Molecular characterization, expression pattern, and function analysis of the OsBC1L family in rice. Plant Mol Biol *71*, 469-481.
12. Gao, Y., Bian, L., Shi, J., Xu, J., Xi, M., and Wang, G. (2013). Expression of a conifer COBRA-like gene ClCOBL1 from Chinese fir (Cunninghamia lanceolata) alters the leaf architecture in tobacco. Plant Physiol Biochem *70*, 483-491.
13. Gu, Y., Kaplinsky, N., Bringmann, M., Cobb, A., Carroll, A., Sampathkumar, A., Baskin, T.I., Persson, S., and Somerville, C.R. (2010). Identification of a cellulose synthase-associated protein required for cellulose biosynthesis. Proc Natl Acad Sci U S A *107*, 12866-12871.
14. Han, S., and Kim, D. (2006). AtRTPrimer: database for Arabidopsis genome-wide homogeneous and specific RT-PCR primer-pairs. BMC Bioinformatics *7*, 179.
15. Hietala, A.M., Eikenes, M., Kvaalen, H., Solheim, H., and Fossdal, C.G. (2003). Multiplex real-time PCR for monitoring Heterobasidion annosum colonization in Norway spruce clones that differ in disease resistance. Appl Environ Microbiol *69*, 4413-4420.
16. Hochholdinger, F., Wen, T.J., Zimmermann, R., Chimot-Marolle, P., Silva, O.D.E., Bruce, W., Lamkey, K.R., Wienand, U., and Schnable, P.S. (2008). The maize (Zea mays L.) roothairless3 gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield. Plant Journal *54*, 888-898.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

17. Jones, M.A., Raymond, M.J., and Smirnoff, N. (2006). Analysis of the root-hair morphogenesis transcriptome reveals the molecular identity of six genes with roles in root-hair development in Arabidopsis. Plant J *45*, 83-100.
18. Ju, C.L., Van de Poel, B., Cooper, E.D., Thierer, J.H., Gibbons, T.R., Delwiche, C.F., and Chang, C.R. (2015). Conservation of ethylene as a plant hormone over 450 million years of evolution. Nature Plants *1*.
19. Li, S., Bashline, L., Lei, L., and Gu, Y. (2014). Cellulose synthesis and its regulation. Arabidopsis Book *12*, e0169.
20. Li, S., Ge, F.R., Xu, M., Zhao, X.Y., Huang, G.Q., Zhou, L.Z., Wang, J.G., Kombrink, A., McCormick, S., Zhang, X.S.*, et al.* (2013). Arabidopsis COBRA-LIKE 10, a GPI-anchored protein, mediates directional growth of pollen tubes. Plant J *74*, 486-497.
21. Li, S., Lei, L., Somerville, C.R., and Gu, Y. (2012). Cellulose synthase interactive protein 1 (CSI1) links microtubules and cellulose synthase complexes. Proc Natl Acad Sci U S A *109*, 185-190.
22. Li, Y., Qian, Q., Zhou, Y., Yan, M., Sun, L., Zhang, M., Fu, Z., Wang, Y., Han, B., Pang, X.*, et al.* (2003). BRITTLE CULM1, which encodes a COBRA-like protein, affects the mechanical properties of rice plants. Plant Cell *15*, 2020-2031.
23. Liu, L., Shang-Guan, K., Zhang, B., Liu, X., Yan, M., Zhang, L., Shi, Y., Zhang, M., Qian, Q., Li, J.*, et al.* (2013). Brittle Culm1, a COBRA-like protein, functions in cellulose assembly through binding cellulose microfibrils. PLoS Genet *9*, e1003704.
24. McFarlane, H.E., Doring, A., and Persson, S. (2014). The cell biology of cellulose synthesis. Annu Rev Plant Biol *65*, 69-94.
25. Mendu, V., Griffiths, J.S., Persson, S., Stork, J., Downie, A.B., Voiniciuc, C., Haughn, G.W., and DeBolt, S. (2011). Subfunctionalization of cellulose synthases in seed coat epidermal cells mediates secondary radial wall synthesis and mucilage attachment. Plant Physiol *157*, 441-453.
26. Niu, E., Shang, X., Cheng, C., Bao, J., Zeng, Y., Cai, C., Du, X., and Guo, W. (2015). Comprehensive Analysis of the COBRA-Like (COBL) Gene Family in Gossypium Identifies Two COBLs Potentially Associated with Fiber Quality. PLoS One *10*, e0145725.
27. Nuhse, T.S., Stensballe, A., Jensen, O.N., and Peck, S.C. (2004). Phosphoproteomics of the Arabidopsis plasma membrane and a new phosphorylation site database. Plant Cell *16*, 2394-2405.
28. Richmond, T.A., and Somerville, C.R. (2000). The cellulose synthase superfamily. Plant Physiol *124*, 495-498.
29. Roudier, F., Fernandez, A.G., Fujita, M., Himmelspach, R., Borner, G.H., Schindelman, G., Song, S., Baskin, T.I., Dupree, P., Wasteneys, G.O.*, et al.* (2005). COBRA, an Arabidopsis extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation. Plant Cell *17*, 1749-1763.
30. Roudier, F., Schindelman, G., DeSalle, R., and Benfey, P.N. (2002). The COBRA family of putative GPI-anchored proteins in Arabidopsis. A new fellowship in expansion. Plant Physiol *130*, 538-548.
31. Sato, K., Ito, S., Fujii, T., Suzuki, R., Takenouchi, S., Nakaba, S., Funada, R., Sano, Y., Kajita, S., Kitano, H.*, et al.* (2010). The carbohydrate-binding module (CBM)-like sequence is crucial for rice CWA1/BC1 function in proper assembly of secondary cell wall materials. Plant Signal Behav *5*, 1433-1436.
32. Schindelman, G., Morikami, A., Jung, J., Baskin, T.I., Carpita, N.C., Derbyshire, P., McCann, M.C., and Benfey, P.N. (2001). COBRA encodes a putative GPI-anchored protein, which is polarly localized and necessary for oriented cell expansion in Arabidopsis. Genes Dev *15*, 1115-1127.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016

33. Sindhu, A., Langewisch, T., Olek, A., Multani, D.S., McCann, M.C., Vermerris, W., Carpita, N.C., and Johal, G. (2007). Maize Brittle stalk2 encodes a COBRA-like protein expressed in early organ development but required for tissue flexibility at maturity. Plant Physiol *145*, 1444-1459.
34. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol *30*, 2725-2729.
35. Taylor-Teeples, M., Lin, L., de Lucas, M., Turco, G., Toal, T.W., Gaudinier, A., Young, N.F., Trabucco, G.M., Veling, M.T., Lamothe, R.*, et al.* (2015). An Arabidopsis gene regulatory network for secondary cell wall synthesis. Nature *517*, 571-575.
36. Walley, J.W., Coughlan, S., Hudson, M.E., Covington, M.F., Kaspi, R., Banu, G., Harmer, S.L., and Dehesh, K. (2007). Mechanical stress induces biotic and abiotic stress responses via a novel cis-element. PLoS Genet *3*, 1800-1812.
37. Walter, S., and Schrempf, H. (2003). Oligomerization, membrane anchoring, and cellulose-binding characteristics of AbpS, a receptor-like Streptomyces protein. J Biol Chem *278*, 26639-26647.
38. Zhang, D.Q., Yang, X.H., Zhang, Z.Y., and Li, B.L. (2010). Expression and nucleotide diversity of the poplar COBL gene. Tree Genet Genomes *6*, 331-344.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1740v1 | CC-BY 4.0 Open Access | rec: 10 Feb 2016, publ: 10 Feb 2016