Ancient origin of fucosylated xyloglucan in charophycean green algae

Mikkelsen, Maria Dalgaard; Harholt, Jesper; Westereng, Bjørge; Domozych, David; Fry, Stephen C; Johansen, Ida Elisabeth; Fangel, Jonatan U.; yk, Mateusz; Feng, Tao; Nancke, Louise

Total number of authors:
13

Published in:
Communications Biology

Link to article, DOI:
10.1038/s42003-021-02277-w

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Mikkelsen, M. D., Harholt, J., Westereng, B., Domozych, D., Fry, S. C., Johansen, I. E., Fangel, J. U., yk, M., Feng, T., Nancke, L., Mikkelsen, J. D., Willats, G. T., & Ulvskov, P. (2021). Ancient origin of fucosylated xyloglucan in charophycean green algae. Communications Biology, 4(1), [754]. https://doi.org/10.1038/s42003-021-02277-w
Ancient origin of fucosylated xyloglucan in charophycean green algae

Maria Dalgaard Mikkelsen1,2, Jesper Harholt3, Bjørge Westereng4, David Domozych5, Stephen C. Fry6, Ida Elisabeth Johansen1, Jonatan U. Fangel1, Mateusz Łężyk7, Tao Feng7, Louise Nancke1, Jørn D. Mikkelsen7, William G. T. Willats1,8 & Peter Ulvskov1

The charophycean green algae (CGA or basal streptophytes) are of particular evolutionary significance because their ancestors gave rise to land plants. One outstanding feature of these algae is that their cell walls exhibit remarkable similarities to those of land plants. Xyloglucan (XyG) is a major structural component of the cell walls of most land plants and was originally thought to be absent in CGA. This study presents evidence that XyG evolved in the CGA. This is based on a) the identification of orthologs of the genetic machinery to produce XyG, b) the identification of XyG in a range of CGA and, c) the structural elucidation of XyG, including uronic acid-containing XyG, in selected CGA. Most notably, XyG fucosylation, a feature considered as a late evolutionary elaboration of the basic XyG structure and orthologs to the corresponding biosynthetic enzymes are shown to be present in Mesotaenium caldarium.
The group of green algae that is most closely related to land plants is the charophycean green algae (CGA or basal streptophytes). Phylogenetic analyses also indicate that land plants (or embryophytes) evolved from an ancestral taxon of the CGA class, Zygnematophyceae5–8 approximately 500+ million years ago7. The subsequent global radiation of land plants profoundly changed the biogeochemistry and natural history of the planet, and has consequently resulted in Earth’s highly diverse terrestrial ecosystems5.

The ancestral CGA that colonized land evolved traits that enabled them to survive in a terrestrial habitat, including specialized cell wall features9. This supposition has been supported by recent genomic sequencing of CGA leading to significant insights into the evolutionary origin of many land plant characteristics5,9.

Extant CGA produce many cell wall polymers that are also found in land plants11–13. This has been recently corroborated by establishing orthology between genes involved in the biosynthesis of cell wall polymers in flowering plants and various CGA taxa5,14,15. One such polymer type is xyloglucan (XyG)16. XyG, a major wall component, has been well-characterized in non-commelinid angiosperm primary cell walls, but also refers to Type-I cell walls17. It has been hypothesized that XyG exhibits a close physical affinity for cellulose, forming tethers between microfibrils and, in turn, contributing to the load-bearing capacity of the wall18. While this model has been challenged19–21, the significance of XyG in cell expansion and cell wall architecture is generally accepted. XyG consists of a β-(1,4)-glucan (designated with the one-letter code G) backbone, with sidechains starting with α-(1,6) linked xylose (X). The xylose can be further decorated with different glycosyl groups, such as β-(1,2) linked galactose (L) followed by α-(1,2) linked fucose (F) or β-(1,2) glucuronic acid (Y) followed by β-(1,4) galactose (P)22–24. XyG is synthesized in the Golgi apparatus via the action of glycosyl transferases derived from different gene families, several of which have been verified through functional characterization16,25,26.

Due to the absence of biochemical evidence for XyG being present in CGA, a prominent hypothesis is still that XyG originated in land plants12. Furthermore, fucosylation of XyG was thought to have arisen in the common ancestor of flowering plants based on its absence in the cell walls of early divergent land plants27. Fucosylated XyG has previously been considered a hallmark feature of seed plants (spermatophytes). However, fucosylated XyG has been found in hornworts28, while XyG-like fucosylated XyG has been found in land plants11.

The phylogeny of XyG biosynthesis

**Fig. 1 Phylogeny of the green plant lineage.** Three later diverged CGA classes are situated as the closest living relatives to land plants; the Charophyceae, Coleochaetophyceae and the Zygnematophyceae, while the Klebsormidophyceae, Chlorokybophyceae and Mesostigmatophyceae are classes of earlier evolved CGA. The Charophytes and Prasinophytes are ancestral to CGA. Green algal species of interest here are illustrated by pictures. The Rhodophytes are used as ancestral group for the green plants.

**Results**

**Phylogeny of XyG biosynthesis.** In the green plant lineage (Fig. 1) cellulose is nearly ubiquitous. Recent gene sequencing of several CGA taxa5,10,33–36 has allowed the identification of CGA orthologs to the cellulose synthase-like C (CslC) clade of CAZy glycosyl transferase family 2 (GT2), containing the Glc β-1-4 XyG-backbone synthase. On the basis of these studies, we have identified sequences from Mesotaenium, Spirogyra, and Cylindrothates where Spiroplloea sequence 80 is on the unresolved borderline between the clades and 78–79 squarely in clade C (Fig. S1).

GTs of GT47 clade A adds the next sugar moiety, either galactose or arabinose (D-chain)24,41,42. To investigate the genetic potential of CGA galactosyl transferase capability we cloned two full-length Coleochaete orbicularis sequences, CoGT47A1 and 2 (Genbank accession numbers: MW149251 and MW149252, respectively), which are positioned in a subclade subtending the clade of Arabidopsis thaliana XLT2 (a galactosyltransferase) and tomato XST1&2 (arabinosyltransferases) (Fig. S2). The Pentium margaritaceum genome was published3 after the work for this paper was completed, and is not included in our analysis except for the XyG backbone using homologous enzymes to those of land plants. Orthologs required for sidechain biosynthesis (GT34, 37, and 47), were also identified, suggesting that late divergent CGA have a full biosynthetic inventory for production of complex XyGs. XyG sidechain synthesis is initiated by xylosylation by xyloglucan xylosyltransferases (XXTs) belonging to the GT34 family (Fig. S1)37–39. The GT34 B-clade comprising the galactomannan galactosyltransferases is very poorly resolved, and it will require analyses with more species included to ascertain whether the A and C-clades37,40 are stable. Phycosinella does not have sequences in either, yet is able to make the L-chain and more elaborate XyG sidechains (but not the F-chain)28. Selaginella is not represented in the C-clade but features two sequences in the A-clade and is able to make the F-chain40. Similar sequences are identified in Mesotaenium, Spirogyra, and Cylindrothates whereas Spiroplloea sequence 80 is on the unresolved borderline between the clades and 78–79 squarely in clade C (Fig. S1).

GTs of GT47 clade A adds the next sugar moiety, either galactose or arabinose (D-chain)24,41,42. To investigate the genetic potential of CGA galactosyl transferase capability we cloned two full-length Coleochaete orbicularis sequences, CoGT47A1 and 2 (Genbank accession numbers: MW149251 and MW149252, respectively), which are positioned in a subclade subtending the clade of Arabidopsis thaliana XLT2 (a galactosyltransferase) and tomato XST1&2 (arabinosyltransferases) (Fig. S2). The Pentium margaritaceum genome was published3 after the work for this paper was completed, and is not included in our analysis except for the XyG backbone using homologous enzymes to those of land plants. Orthologs required for sidechain biosynthesis (GT34, 37, and 47), were also identified, suggesting that late divergent CGA have a full biosynthetic inventory for production of complex XyGs. XyG sidechain synthesis is initiated by xylosylation by xyloglucan xylosyltransferases (XXTs) belonging to the GT34 family (Fig. S1)37–39. The GT34 B-clade comprising the galactomannan galactosyltransferases is very poorly resolved, and it will require analyses with more species included to ascertain whether the A and C-clades37,40 are stable. Phycosinella does not have sequences in either, yet is able to make the L-chain and more elaborate XyG sidechains (but not the F-chain)28. Selaginella is not represented in the C-clade but features two sequences in the A-clade and is able to make the F-chain40. Similar sequences are identified in Mesotaenium, Spirogyra, and Cylindrothates whereas Spiroplloea sequence 80 is on the unresolved borderline between the clades and 78–79 squarely in clade C (Fig. S1).
that Supplementary data 1 provides a list of the genomic sequences of the GT47 A-clade. We note that no P. margaritaceum sequence is closely related to MUR3/Katamari and that Supplementary data 1 provides a list of the genomic sequences, i.e. XyG occurrence by comprehensive microarray polymer profiling (CoMPP)47, using two different XyG specific antibodies LM15 and LM25, see heatmaps in Fig. 3 and S4. The early divergent green algae Chlorokybus and Klebsormidium did not show any labeling, whereas a low xyloglucanase (E-XEGP) digestible signal was observed in Mesostigma viride (Fig. 3 and S4). Positive labeling was observed in many species of the late divergent CGA. The XyG recognized by LM25 was susceptible to E-XEGP treatment, whereas LM15 specific XyG showed an intriguing recalcitrance to E-XEGP treatment, normally not observed in higher plants48. Interestingly, XyG labeling and E-XEGP susceptibility varied between different species and isolates of the same genus (Fig. 3 and S4). The recalcitrance towards xyloglucanase was further investigated using a xyloglucanase (XcXGHA)48 with a unique specificity towards substituted XyG, cleaving between two X-chains, which is in contrast to the commonly used xyloglucanase, which cleaves between a substituted and a non-substituted glucosyl residue in the xyloglucan backbone. The results, however, showed a similar recalcitrance, suggesting that is likely not caused by a lack of un-substituted residues (Fig. S5). The specificities of LM15 and LM25 are not identical and neither are their sensitivities, with LM25 being substantially more sensitive. LM15 was raised against an XXXG conjugate and cross-reacts to some extent with XLXG, while LM25 was raised against an XLXG/XLG conjugate49 and cross-reacts to some extent with XXXG. We suggest that LM25 detects...
a separate and smaller XyG population and there are a number of reasons why catalytically active xyloglucanases may cleave XyG yet only reduce LM15 binding to a minor extent. If the product oligosaccharides are sufficiently large, they will likely remain immobilized to the nitrocellulose. Similarly, if they are covalently attached to other macromolecules they may remain immobilized.

In this regard, it is worth-noting that endo-transglycosidases in CGAs are rather donor substrate tolerant and can graft mannan, for example, onto XyG. It is also possible that certain levels and/or patterns of substitution may reduce the accessibility of a xyloglucanase to, or its activity against, a portion of XyG that nonetheless remains detectible by LM15. This effect has been noted previously for other substrate/enzyme/mAb systems. For example, Willats et al. reported that despite extensive degradation of pectin by a highly active endo-polygalacturonase, there was a residue of recalcitrant homogalacturonan (HG) that was still detectable by anti-HG mAbs and this was most likely due to differential branching patterns.

The defining substructure of XyG is X, i.e., the disaccharide isoprimeverose and the release of isoprimeverose by Driselase digestion is thus diagnostic of XyG. Mesotaenium caldariorum and Netrium digitus were subjected to Driselase digestion and isoprimeverose was detected, although it was not detectable in Coleochaete orbicularis (Fig. S6a), nor had it been detected earlier by identical methods in Coleochaete scutata, Chara and Klebsormidium, indicating that it is not a quantitatively major component of CGA polysaccharides. Based on these results and our labeling studies (see below) we propose that XyG is developmentally dynamic in the cell walls of the CGA studied, possibly, due to involvement in cell division and wall expansion, and that this may be the reason for the variable results.

**Immunohistochemistry.** In order to identify the cellular localization of XyG in CGA, immunolabeling with the XyG-binding LM15 monoclonal antibody, was performed on live and fixed/resin-embedded algae (Fig. 4 and Figs. S7–S9). XyG epitopes were detected in all species investigated. Labeling was found within specific cell domains, especially in expanding or dividing cells (Fig. 4 and Figs. S7–S9). To further explore the possible role of XyG in cell division or cell expansion, synchronized Mesotaenium caldariorum cells were labeled with LM15 (Fig. 5). Labeling was noted from the beginning of the production of the new "primary cell wall" at the isthmus, i.e., the site of cell expansion and division (Fig. 5a, b). It should be noted that this is not analogous to "primary cell walls" of land plants. Daughter cells retained labeling in the zones near the division zone (Fig.5c–f). After the daughter cells stopped expanding, the primary wall was shed, which is a typical wall developmental event in Mesotaenium species (Fig. 5g–h). These results indicate that XyG, as highlighted by LM15, is produced in expanding walls during semi-cell morphogenesis but not in non-expanding mature cell walls.

**Fig. 3 Heatmap of Comprehensive Microarray Polymer Profiling (CoMPP) of a range of CGA species.** Samples were sequentially extracted with CDTA and NaOH. Spotted arrays were probed with XyG-specific antibodies LM15 and LM25 with or without pre-digestion with XEGP xyloglucanase.
composition not architecture, and we did not pursue investigations of possible polymer masking. It is important to note that our in vivo labeling depicts epitopes that are accessible to the antibody. However, uniform labeling of the walls was never observed, and this lack of labeling in mature walls may be the result of incorporation of XyG into wall constructs that render it inaccessible to the antibody. Likewise, it may be the result of XyG metabolism in different developmental stages. While these two scenarios are not mutually exclusive and labeling appears to correlate with division and/or expansion, future work using systematic unmasking protocols is required to elucidate the presence of XyG at different developmental stages.

XyG fine structure mapping using MALDI-ToF. We selected CGA taxa representing the late divergent CGA classes, Coleochaetophyceae and Zygnematophyceae, which showed high labeling and substantial E-XEGP susceptibility in COMPP and analyzed XyG composition in their cell walls using MALDI-ToF (OLIMP) (Fig. 6). Peaks matching known XyG oligomers (XGO) and with mass differences in series of related XGOs, hexoses (m/z of 162), pentoses (m/z 132), uronic acids (m/z 176), and deoxy-hexoses (m/z 146) could be observed in all species analyzed. In Coleochaete orbicularis XGOs with mass differences corresponding to hexoses and xyloses were found, strongly suggesting the presence of non-fucosylated XyG (Fig. 6). In Cylindrocystis brebissonii m/z peaks corresponding to XGOs could be identified corresponding to differences of uronic acids (Fig. 6). XyG with galacturonic acid (GalA) has previously been found in root hairs of A. thaliana and in P. patens and Marchantia polymorpha, while glucuronic acid has never been found in XyG to date. Coleochaete orbicularis and Cylindrocystis brebissonii feature sparsely branched XyG, with the suggested structures of LGGGG (1146), PGGGG (1322), LLGGG (1440), and LPGGG (1616), forming a new series of XGOs (Figs. 6 and 7). Furthermore, there were indications of a XyG specific acetylation (1616->1658 a mass difference of 42), although this needs further investigation. In Netricum digitus the m/z differences found correspond to hexose, pentose and uronic acid (Fig. 6), suggesting XyG with a complex decoration with both hexoses, pentoses and uronic acids, but generally with smaller XGO masses.

The yield of enzymatically extractable XGOs in Mesotaenium caldarium was high compared to the yield from the other species. The m/z differences implied the presence of hexoses, pentoses, uronic acids, and very interestingly also deoxy-hexoses, (Fig. 6) identified as fucose, see below. The latter have, to date, only been identified in land plants. XyG structure can vary substantially between tissue and cell types as observed in a study of monocots,57 of fucosylated XyG in tobacco58 and of XyG featuring uronic acid XyG78,54,56–58. Fucosylated XyG, featuring the F-chain, is the dominating type in A. thaliana while the galacturonic acid (GaLA) containing the Y sidechain (Fig. 7a) has been found only in root hairs.56

To verify the findings of fucosylated and galactosylated XyG in Mesotaenium caldarium, xyloglucanase digestion was supplemented with either α-1,6-galactosidase (E-BGLAN, Megazyme) or α-,1-fucosidase (MFucS).59 After fucosidase treatment, m/z 1099 and 1393 disappeared, while m/z 1555 was partially degraded (Figs. 7c and 8c). The predicted m/z after degradation of these peaks was 953, 1247, and 1409, respectively. These m/zs were already present before fucosidase treatment, e.g., no new masses were produced, although the peaks all appeared to increase in relative abundance. The m/zs of these three fucosidase degradable XGOs corresponds to the masses of XFG, XXFG, and XLFG, respectively. The presence of XXFG was further supported by TLC analysis where a distinct band co-migrated with the standard XXFG (Fig. 8b), though the corresponding band was very faint in Netricum digitus, undetectable in Coleochaete orbicularis (Fig. 8b) and had not been detected by identical methods in Chara. In addition, CoMPP analysis using antibodies specific for fucosylated XyG, CCRC-M1, and CCRC-M39, revealed antibody labeling in Mesotaenium caldarium which was diminished after treatment with fucosidase. Labeling of fucosylated XyG was undetected in Netricum digitus, Coleochaete orbicularis and Cylindrocystis brebissonii supporting the MALDI-ToF analysis (Fig. S10).

Based on the MALDI-ToF analysis, galactosylated XGOs are abundant in Mesotaenium caldarium and galactosidase digestion shows that m/z 953, 1115, 1247, 1409, and 1555 were susceptible to galactosidase digestion, suggesting that the peaks are XLG, GLXL/GLXG/LLG, XXLG/XLXG, XLLG, and XLFG, respectively (Fig. 8b). Some of these findings were also supported by TLC analysis, where XGO co-migrated with XXLG/XLXG and another XGO possibly co-migrating with XLLG (Fig. 8b). Additionally, it was observed that the series of related XGOs with m/z of 1129, 1423, 1445, and 1585 were susceptible to galactosidase degradation, resulting in the appearance of m/z of 967, 1261, and a less abundant m/z of 1283 (Fig. 8b). Tandem MS, performed on the XGOs m/z 1129 and 1423 (Fig. 8d, e), both of which are galactosidase sensitive, revealed that they differ by an extra X in 1423, 1129 comprising one terminal pentose, a terminal hexose (likely the terminal galactose, determined by galactosidase treatment), and a terminal uronic acid. These findings suggest the presence of sidechain P. The branching order in these XGOs is generally unknown so possible structures...
comprise XPG/PXG and XXPG/PXXG, respectively. BP GG is an alternative structure of XXPG shown in Fig. 8b. We consider this unlikely as the XGOs of Fig. 8 would then no longer belong to the same series.

Taken together, these data show that Mesotaenium caldariorum produces XyG featuring sidechains previously thought to have evolved in land plants. The emergence of XyG fucosylation among later-evolving CGA raises the question of how early the origin of XyG itself should be sought. We searched the Chlorokybus genome for sequences related to ClsC or CslK that putatively could encode XyG backbone synthases and found none. However, we found a CESA gene that appears to fulfill the criteria for being rosette forming, suggesting that plant cell wall biosynthesis, as we know it from land plant studies, has its origin in a common ancestor of Chlorokybus and land plants.

Discussion

The discussion of XyG sidechains presented by Schultink et al., suggests that the evolution of XyG structure is one of gradual simplification, such that X-, L-, and F-chains eventually become the predominant sidechains. It is thus not surprising that the repertoires of XyG sidechains vary within the CGA, with the observation of a new series of negatively charged and sparsely branched XGOs in Coleochaete and Cylindrocystis as striking examples. Here, we extend our understanding of the evolutionary history of XyG biosynthetic machinery, and propose that the origin of XyG roughly coincided with the emergence of the phragmoplast. Furthermore, we contribute the identification of CGA Csl sequences that are unambiguously assigned to the CslC clade and provide Coleochaete full-length sequences orthologous to XLT2 and XST1&2 (GT47, Fig. S2).

The P and Q sidechains, which are unique in containing galactosyl bound to the O-4 position together with a galacturonate linked to O-2 on the same xylose, have been identified in liverworts and mosses and not in hornworts, whose cell walls feature GalA-containing XyG motifs that are indistinguishable from those of flowering plants. The related linear Y-chain has been found in A. thaliana roots and the phylogenetic analysis of the galacturonosyltransferase that creates the Y-chain showed that it, AtXUT1, is a GT47 clade A protein belonging to a subclade different from MUR3 and with several Physcomitrella members, suggesting that these GalA-containing XyG sidechains are homologous rather than results of convergent evolution. Knocking out the galacturonosyltransferase in A. thaliana resulted in shorter root hairs, suggesting a role of AtXUT1 and GalA containing XyG in root hair elongation. Mutations in the XyG-specific galacturonosyltransferase (MpXUT1) gene in the liverwort M. polymorpha resulted in very short rhizoids that eventually burst, suggesting a role of the GalA containing XyG in the tip-growing rhizoid, but the exact function is still unclear. These GT47 members belong to clade A of GT47 and all transfer glycosyl residues to the O-2 position of xylose in XyG.

Fig. 5 XyG production in Mesotaenium caldariorum during semicell morphogenesis. Immunolabelling of cell cycle synchronized Mesotaenium caldariorum cells during semicell morphogenesis. a, b LM15 binding was found in the isthmus when cells were starting to divide. c, d LM15 is evident in the ends of daughter cells near the division zone (while chlorophyll is also evident and colored green in the middle of the two daughter cells as indicated by arrows) e, f cell expansion is almost completed of the new Mesotaenium caldariorum cells and LM15 labeling is very clear in the newly formed “primary cell wall”. g, h Cell division and expansion is completed and the “primary cell wall” is cast off as a single piece depicted here, including the LM15 labeling. Top panel: LM15 (TRITC) (green), lower panel: DIC light image. c Chlorophyll autofluorescence (green), in the middle of the cells. Scalebars as indicated in the DIC images: 20 µm.
The F-chain has been found in hornworts but is absent from mosses. This absence is corroborated by sequence comparisons of moss GT37 proteins with those of vascular plants supporting the conclusion, that the clade responsible for L-chain fucosylation has no moss taxa. Clade structure analysis is not without complications in family GT37, yet available evidence suggests that XyG fucosylation is lost in mosses; and there may be little selection pressure for preserving XyG fucosylation due to the type of conductive tissues in mosses. The functional relevance of fucosylation to vasculature has been demonstrated by Brennan et al. In an investigation of cell type specific occurrence of fucosylated XyG in commelinid and non-commelinid monocots. The authors found that XyG is not fucosylated in most tissues of non-commelinid monocots. However, with few exceptions, they were able to detect fucosylated XyG in cells associated with the vasculature in a range of commelinids. A similar investigation would be highly relevant to Solanales, which also generally feature non-fucosylated XyG. In tomato (Solanaceae) two XLT2 galactosyltransferases were found to be neofunctionalised to transfer L-Ara rather than D-Gal thus generating an S-chain rather than an L-chain and at the same time eliminating the F-chain. However, synthesis of fucosylated XyG has been demonstrated in pollen tubes of tobacco also belonging to Solanales, corroborating that fucosylation was not acquired multiple times during evolution but rather lost in tissues where no selection pressure favors its conservation. It appears that fucosylated XyG has functional properties that sets it apart from XyG in general but it is not yet clear what precisely is the decisive property.

**Methods**

**Bioinformatics.** Genome sequences from the CGA Mesostigma viride and Chlorokybus atmosphitus, Chara brauni, Klebsormidium flaccidum, Spiroloce muscicola gen. nov., Mesotaenium endlicherianum, and CGA transcriptomic data was obtained from 1KP dataset. In short a local BLAST database consisting of the glycosyltransferases of CAZy (http://www.cazy.org) with the addition of the glycosyltransferases from the land plants A. thaliana, Oryza sativa, S. moeller-dorffii, and P. patens, acquired from Genbank, using Harholt et al. as guide, was generated. The proteome of each taxon was blasted against this data with a cut-off of 10−25 and quality controlled. Phylogenetic analysis was done as in Harholt et al. using Phylogeny.fr (http://phylogeny.lirmm.fr/phylo/cgi/index.cgi). Penium maritimum GT47 phylogenies were estimated by alignment using MUSCLE trimming in TrimAl with gt=0.05, and trees were generated using PhyML. Keys to trees are provided in Supplementary data 1.

**Cloning of Coleochaete orbiculare genes.** Cells from a fresh culture of Coleochaete orbiculare were harvested by centrifugation at 4000×g, resuspended in 2 mg/ml Driselase (Sigma-Aldrich) and incubated with gentle shaking for 30 min at room temperature. Protoplasts were harvested at 4000×g. App. 0.5 g cells were transferred to a mortar, ground in liquid nitrogen and RNA was extracted with Spectrum Plant Total RNA kit (Sigma-Aldrich). cDNA was synthesized with SuperScript III First strand SuperMix (Invitrogen). Two putative and partial Coleochaete orbiculare GT47A sequences of 447 and 661 bp were identified in the 1 Kp data set by alignment to land plant sequences (Selaginella, Physcomitrella, Amborella, Vitis and Cucumis). Primers were constructed for the sequences (FW1 and RV; Table S1) and PCR using LA Taq polymerase (Takara) was performed on Coleochaete orbiculare cDNA. PCR resulted in gene fragments of the expected sizes and were confirmed by sequencing.

To obtain the 5′ end of the CoGT47A1 gene, a forward primer (FW2) was placed in the conserved regions in the 3′ end of a Coleochaete scutata sequence (found in the 1KP dataset: 10200_VQBJ_GT47_MUR3/1409). This primer was used with the CoGT47A1 reverse (RV) primer and resulted in the expected product and was cloned and sequenced and found to have perfect overlap with the CoGT47A1 FW + RV PCR product.

The 3′ end of CoGT47A1 was amplified by a nested PCR using first CoGT47A1 FW3n and an anchored oligo-DT primer followed by nested PCR reaction CoGT47A1 FW4n and anchored oligo-DT primer. The obtained product where cloned and sequenced.

To obtain the 5′ end of both CoGT47A genes, 5′ RACE was performed using SMARTer RACE5′ kit (ClonTech/Takara) according to manufacturer’s instructions using the GSP1 and NGSP1 primers for the respective genes (Table S1). The resulting PCR products were cloned and sequenced. The full-length coding sequences were amplified by PCR using the primers FLFW and FLRV for...
Fig. 7 Tentative assignments of peaks in Coleochaete orbicularis and Cylindrocystis brebissonii and suggested XyG structures found in Mesotaenium caldariorum. a Relevant XyG structures including standard one-letter codes for the XyG sidechains represented in b, c sodium adducts. b Putative-related XGOs are released after xyloglucanase treatment in Coleochaete orbicularis and Cylindrocystis brebissonii. Theoretical masses are indicated here while Fig. 6 reports recorded masses. c Tentative monosaccharide composition and candidate structures in Mesotaenium caldariorum are provided along with the enzyme sensitivities of the proposed oligosaccharides. Pen pentose, Hex, hexose, dHex deoxy-hexose, Uron uronic acid.
CoGT47A1 and CoGT47A2, respectively (Table S1). The full-length genes were uploaded to Genbank with the following accession numbers: MW149251 for CoGT47A1 and MW149252 for CoGT47A2.

**Plant materials, cultures, and media.** An Anthoceros caucasicus (hornwort) cell-culture was kindly donated by Dr Maike Petersen, University of Marburg, Germany; rose, spinach, maize cell-cultures were as reported by Popper and Fry.\(^{72}\) CGA for CoMPP and TLC were grown for 10–14 days and harvested by centrifugation 800×g for 1 min.\(^{12}\) Chlorella was grown at the Skidmore college Greenhouse wet wall. The CGA for MALDI-ToF analysis were grown as follows: Netrium digitus (LB599 University of Texas Algal Culture Collection or UTEX culture collection) and Cyndrocladae breibisoni (UTEX 1922) were cultured in 250 mL flasks containing 125 mL of sterile WHM medium supplemented with soil extract (10% final concentration) at 20 °C, 16 h light/8 h dark photodicy with 2000 lux of cool white fluorescent light, Coleochaete orchidaris (UTEX LB2651) was grown in petri dishes containing sterile liquid WHM supplemented with peat extract (5% final concentration) under the conditions described above but with the temperature reduced to 18 °C. Mesotaenium caldariorum (UTEX 41) was grown 10–14 days in WHM medium. All CGA material wash with fresh sterile medium and frozen in liquid N\(_2\) and lyophilized.

**Preparation of AIR from CGA.** Lysophilized CGA samples were ground using liquid N\(_2\) and agate mortar. Seventy percentage of EtOH was added 3–5 times and incubated RT for 10 min and 100 rpm. When no further green color was extracted from the EtOH extraction, 96% EtOH was added and incubated at RT for 10 min at 100 rpm. chloroform:methanol was added in an 1:1 ratio and 100 rpm for 10 min. the sample was centrifuged 10 min at 4000 rpm after each treatment and the supernatant discarded. Acetone was added in five volumes and incubated 5 min at 100 rpm. Samples were dried in the fume hood at RT overnight.

**CoMPP.** Different monoclonal antibodies were used to probe CoMPP arrays: LM15\(^{54}\), LM23\(^{59}\) (purchased from Plantprobes) and the CCRC-M1 and
CCRC-M39 (purchased from CarboSource). The antibodies CCRC-M173 and M39 are both specific towards fucosylated XyG. Algal AIR was extracted with EDTA and NaOH, printed on arrays and blocked with skim milk solution. After blocking, arrays were treated with enzymes; xyloglucanase E-XEGP (GHS) from Paenibacillus sp. (Megazyme, Bray, County Wicklow, Ireland) in 100 mM sodium acetate buffer pH 5.5 at concentrations of 0.1 M U/mL or MFruc5 in 20 μM citrate-phosphate buffer pH 6 at concentrations of 0.02 μM U/mL and incubated overnight at 30 °C and 80 rpm. With arrays, the xyloglucanase XcXGHA was used at a concentration of 0.5 mg/mL. Arrays were treated with 1% (w/v) Driselase in PyAW (pyridine/acetic acid/water, 1:1:98, pH 4.7, containing 0.5% chlorobutanol) at 20 °C for 3 days, then only the ethanol-treated with 1% (w/v) Driselase in PyAW (pyridine/acetic acid/water, 1:1:98, pH 4.7, containing 0.5% chlorobutanol) at 20 °C for 3 days, then only the ethanol-soluble digestion products (oligosaccharides) were resolved by TLC on silica-gel in butan-1-ol/acetic acid/water, by vortexing and spinning at 15,000 rpm for 10 min. Buffer (5 mM sodium acetate buffer pH 6.5) at 20 °C and 80 rpm. Arrays were washed three times and stained with thymol/H2SO4.

MALDI-Tof and tandem MS analysis. Two microliters of a 9 mg mL−1 solution of 2,5-dihydroxybenzoic acid (DHB) in 30% (v/v) acetonitrile was applied on a target plate (MTP 384 target plate ground steel TF from Bruker Daltonics). One microliter sample was then mixed into the DHB droplet followed by drying under a stream of hot air. The analysis was run on an Ultraflex MALDI-Tof-Tof (Bruker Daltonics, GmbH, Bremen, Germany) with a Nitrogen 337 nm laser beam. The instrument was operated in positive acquisition mode and controlled by the FlexControl 3.0.184.0 software package. The acquired spectra were analyzed on FlexAnalysis software (version 3.096.0). The acquisition range was used from m/z 200 to 4000. The data were collected from averaging 1000 laser shots, with the lowest laser energy necessary to obtain sufficient signal to noise ratios.

Mass spectrometry of XyG oligo-saccharides, direct injection MS. For direct injection, oligosaccharides were analyzed using an LTQ-Velos Pro linear ion trap mass spectrometer (Thermo Scientific, San Jose, CA, USA) connected to an Ultimate 3000RS HPLC (Dionex, Sunnyvale, CA, USA). The setup was used for direct injection without a column; the pump delivered 200 µL min−1 70% acetonitrile containing 1 mM formic acid, and data was acquired for 24 s after injection. For the MS, the capillary voltage was set to 3.5 kV and the scan range was m/z 800–2000 using two micro scans. The automatic gain control was set to 10,000 charges and the maximum injection time was 20 ms. For fragmentation of selected precursor ions by MS/MS, the normalized collision energy was set to 35 and two-four micro scans were used.

Statisics and reproducibility. Three extractions were performed for CoMPP analysis each spotted in four dilutions resulting in 12 spots per sample. CoMPP is a semiquantitative analysis and suited for assessing relative differences only, hence the choice of heatmaps for data presentation. Biological variability, where observed, is discussed in the main text.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The data that support the findings of this study are available from the corresponding author upon upon request. Sequence data that support the findings of this study have been deposited in GenBank: MW149251 for CoG747A1 and MW149252 for CoG747A2.

Received: 25 January 2021; Accepted: 28 May 2021; Published online: 17 June 2021

References

1. Wickett, N. J. et al. Phylogenomic analysis of the origin and early diversification of land plants. Proc. Natl Acad. Sci. USA 111, E4859–E4868 (2014).
2. Kenrick, P. & Crane, P. R. The origin and early evolution of plants on land. Phil. Trans. R. Soc. B 365, 1969–1983 (2010).
3. Jiao, C. et al. The Penium margaritaceum genome: hallmarks of the origins of land plants. Cell 181, 1097–1111 (2020).
4. Morris, J. L. et al. The timescale of early land plant evolution. Proc. Natl Acad. Sci. USA 115, E2274–E2283 (2018).
5. Kenrick, P. & Crane, P. R. The origin and early evolution of plants on land. Nature 389, 33–39 (1997).
6. Bower, F. O. The origin of a land flora. Z. Indukt. Abstamm. Vererbungsl. 7, 263–265 (1908).
7. Harlott, J., Moestrup, Ø. & Ulvskov, P. Why plants were terrestrial from the beginning. Trends Plant Sci. 21, 96–101 (2016).
8. Wang, S. et al. Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. Nat. Plants 6, 95–106 (2020).
9. Popper, Z. A. Evolution and diversity of green plant cell walls. Curr. Opin. Plant Biol. 11, 286–292 (2008).
10. Popper, Z. A. & Fry, S. C. Primary cell wall composition of bryophytes and charophytes. Annu. Bot. 91, 11–12 (2003).
11. Sorensen, I. et al. The charophycean green algae provide insights into the early origins of plant cell walls. Plant J. 68, 201–211 (2016).
12. Jensen, J. K. et al. Identification of an algal xylan synthase indicates that there is functional orthology between algal and plant cell wall biosynthesis. New Phytol. 218, 1049–1060 (2018).
13. Mikkelsen, M. D. et al. Evidence for land plant cell wall biosynthetic mechanisms in charophyte green algae. Annu. Bot. 114, 1217–1236 (2014).

Enzymatic digestion of CGA AIR for MALDI-Tof and tandem MS analysis. CGA AIR samples (2.5 mg per experiment) were washed three times in ice-cold water, by vortexing and spinning at 15,000 rpm for 10 min. Buffer (5 mM sodium acetate buffer pH 6) was added and the mixture was incubated 10 min at RT, vortexed vigorously and spun 10 min at 10,000 rpm. The supernatant was discarded, leaving only a small amount of fluid over the pellet. Buffer was added 5 mM sodium acetate buffer pH 6 with or without xyloglucanase E-XEGP 0.1 U/mL and incubated at 40 °C overnight and 1000 rpm. Samples were vortexed vigorously and spun 10 min at 10,000 rpm. The supernatants were collected and spun additionally for 10 min at 10,000 rpm resulting in the XGO mixture. Further enzymatic degradation with fucosidase MFruc5 or galactosidase (Megazyme; β-galactosidase E-BGLAN from Aspergillus niger) was performed by addition of 0.2 mg/mL of MFruc5 from Anaerobiosmesa thermoI or 16 U/mL galactosidase to the XGO mixture and incubation overnight at 30 °C and 100 rpm.
16. Pauly, M. et al. Hemicellulose biosynthesis. *Planta* **238**, 627–642 (2013).
17. Carpta, N. C. & Gibeaut, D. M. Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J.* **3**, 1–30 (1993).
18. McCann, M. C., Wells, B. & Roberts, K. Direct visualization of cross-links in the primary plant cell wall. *J. Cell Sci.* **96**, 323–334 (1990).
19. Cosgrove, D. J. Diffuse growth of plant cell walls. *Plant Physiol.* **176**, 16–27 (2017).
20. Scheller, H. V. & Ulvskov, P. Hemicelluloses. *Annu. Rev. Plant Biol.* **61**, 263–289 (2010).
21. Thompson, D. S. How do cell walls regulate plant growth? *J. Exp. Bot.* **56**, 2275–2285 (2005).
22. Fry, S. C. et al. An unambiguous nomenclature for xyloglucan-derived polysaccharides. *Phytochemistry* **74**, 1–9 (2013).
23. Tuomivaara, S. T., Yaoi, K., Orell, M. A. & Yorke, W. S. Generation and structural validation of a library of diverse xyloglucan-derived oligosaccharides, including an update on xyloglucan nomenclature. *Carbohydr. Res.* **402**, 56–66 (2015).
24. Schultink, A., Liu, Z., Zhu, L. & Pauly, M. Structural diversity and function of hemicellulose side chains. *Planta* **3**, 526–542 (2014).
25. Lerouxel, O., Cavaleri, D. M., Liepan, A. H. & Keegstra, K. Biosynthesis of plant cell wall polysaccharides—a complex process. *Curr. Opin. Plant Biol.* **9**, 621–630 (2006).
26. Scheible, W. R. & Pauly, M. Glycoconyltransferases and cell wall biosynthesis: Novel pathways and insights. *Curr. Opin. Plant Biol.* **7**, 285–295 (2004).
27. Del Ben, L. E. V. & Vincenz, M. G. A. Evolution of xyloglucan-related genes in green plants. *EMBO J.* **10**, 341 (2010).
28. Peña, M. J., Darwin, A. G., Eberhard, S., York, W. S. & O’Neill, M. A. Moss and liverwort xyloglucans contain galacturonic acid and are structurally distinct from the xyloglucans synthesized by hornworts and vascular plants. *Glycobiology* **18**, 1104–1109 (2008).
29. Ikekaya, H. et al. Presence of xyloglucan-like polysaccharide in Spirogyra and possible involvement in cell–cell attachment. *Physiol. Res.* **56**, 216–222 (2008).
30. Domoych, D. S., Sorensen, I. & Willats, W. G. T. The distribution of cell wall polymers during antheridium development and spermatogenesis in the Charophycean green alga. *Chlorella* *cornutula*. *Ann. Bot.* **104**, 1045–1056 (2009).
31. Fry, S. C., Mohler, K. E., Neisleder, B., H. W. A. & Franková, L. Mixed linkage β-glucan: Xyloglucan endotransglucosylase, a novel wall-modelling enzyme from *Equisetum* (horsetails) and charcoal plant algae. *J. Cell Sci.* **55**, 240–252 (2008).
32. Van Sandt, V. S. T., Steiperæare, H., Guize, Y., Verbeelen, J. P. & Vissenberg, K. XET activity is found near sites of growth and cell elongation in bryophytes and some green algae: New insights into the evolution of primary cell wall elongation. *Ann. Bot.* **99**, 51–57 (2001).
33. Leebens-Mack, J. H. et al. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
34. Horik, R. et al. *Klebsormidium flacccatum* genome reveals primary factors for plant terrestrialization. *Nat. Commun.* **9**, 417 (2018).
35. Nishiyama, T. et al. The *chara* genome: secondary complexity and implications for plant terrestrialization. *Cell* **174**, 448–464 (2018).
36. Cheng, S. et al. Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. *Cell* **179**, 1057–1067 (2019).
37. Zabotina, O. A. et al. *Arabidopsis* XX5 gene encodes a putative α-L-xylosyltransferase that is involved in xyloglucan biosynthesis. *Plant J.* **56**, 101–115 (2011).
38. Ade, C. P., Bemm, F., Dickson, M. J. M., Walter, C. & Harris, P. J. Family 34 glycosyltransferase (GT34) genes and proteins in *Pinus radiata* (radiata pine) and *Pinus taeda* (lobolly pine). *Plant J.* **78**, 305–318 (2014).
39. Vuittonchaikhi, S. et al. *Arabidopsis* GT34 family contains five xyloglucan α-1,6-xylosyltransferases. *New Phytol.* **195**, 585–595 (2012).
40. Harbott, J. et al. The glycosyltransferase repertoire of the euprymorphia Selaginella moellendorffii and a comparative study of its cell wall. *PLoS ONE* **7**, e35846 (2012).
41. Madson, M. et al. The MUR3 gene of Arabidopsis encodes a xyloglucan galactosyltransferase that is evolutionarily related to animal exostosins. *Plant Cell* **15**, 1662–1670 (2003).
42. Jensen, J. K., Schultink, A., Keegstra, K., Wilkerson, C. G. & Pauly, M. RNA-seq analysis of developing nasticum seeds (*Tropaeolum majus*): identification and characterization of an additional galactosyltransferase involved in xyloglucan biosynthesis. *Mol. Plant* **5**, 984–992 (2012).
43. Perria, R. M. et al. Xyloglucan fucosyltransferase, an enzyme involved in plant cell wall biosynthesis. *Scientific Reports* **284**, 1976–1979 (2019).
44. Vanzin, G. F. et al. The mur2 mutant of *Arabidopsis thaliana* lacks fucosylated xyloglucan because of a lesion in fucosyltransferase ATFTU1. *Proc. Natl Acad. Sci. U.S.A.* **284**, 1976–1979 (1999).
45. Tryfona, T. et al. Characterisation of a series of transglycanases in charophyte algae. *Mol. Plant Biol.* **104**, 699–710 (1994).
74. Pattathil, S. et al. A comprehensive toolkit of plant cell wall glycan-directed monoclonal antibodies. *Plant Physiol.* **153**, 514–525 (2010).
75. Frankova, L. & Fry, S. C. Phylogenetic variation in glycosidases and glycanases acting on plant cell wall polysaccharides, and the detection of transglycosidase and trans-β-xylanase activities. *Plant J.* **67**, 662–681 (2011).
76. Jork, H., Funk, W., Fischer, W., Wimmer, H. & Burns, D. T. Thin-layer chromatography. Reagents and detection methods. Physical and chemical detection methods: fundamentals, reagents I. Volume 1a. *Anal. Chim. Acta* https://doi.org/10.1016/s0003-2670(00)83962-8. (1990).
77. Cocuron, J. C. et al. A gene from the cellulose synthase-like C family encodes a β-1,4 glucan synthase. *Proc. Natl Acad. Sci. USA* **104**, 8550–8555 (2007).
78. Goubet, F. et al. Cell wall glucomannan in Arabidopsis is synthesised by CSLA glycosyltransferases, and influences the progression of embryogenesis. *Plant J.* **60**, 527–538 (2009).

Acknowledgements
J.H. was supported by the Carlsberg Foundation.

Author contributions
M.D.M., J.H., B.W., S.C.F., I.E.J., T.F., L.N., and P.U.: Performed all experiments and analyzed results. M.D.M., J.H., P.U., S.C.F., and D.D.: Data visualization, data curation and formal analysis. W.G.T.W., P.U., and M.D.M.: Project administration. W.G.T.W., P.U., J.H., and J.D.M.: Experimental ideas and supervision M.D.M., J.H., and P.U.: Paper writing with help from all co-authors.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-021-02277-w.

Correspondence and requests for materials should be addressed to P.U.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Caitlin Karniski. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021