Title: Cryogenian Origin and Subsequent Diversification of the Plant Cell-Wall Enzyme XTH Family

Running head: Origin and diversification of the XTH family
(44 characters, including spaces)

Authors for correspondence:

N. Shinohara
Department of Biology, Kanagawa University,
Tsuchiya 2946, Hiratsuka, 259-1293, Kanagawa, JAPAN
Tel: +81 463594111
FAX: +81 463596277
Email: naoki-s@yd5.so-net.ne.jp

K. Nishitani
Department of Biology, Kanagawa University,
Tsuchiya 2946, Hiratsuka, 259-1293, Kanagawa, JAPAN
Tel: +81 463594111
FAX: +81 463596277
Email: nishitani@kanagawa-u.ac.jp

(Subject Areas) This manuscript is being submitted for consideration for publication in the PCP Special Focus Issue entitled “The mechanics and biology of plant cell walls” edited by Taku Demura, Toshi Kotake, Jenny Mortimer, and Misato Ohitani.

No. of Figures: 8 (Figs 1–8 in color)
No. of Tables: none
No. of Supplementary Data files: 2
1: Suppl_Data.pdf (including Figs S1–S7 and Table S1)
2: Dataset_S1.xlsx

© The Author(s) 2021. Published by Oxford University Press.
Title: Cryogenian Origin and Subsequent Diversification of the Plant Cell-Wall Enzyme XTH Family

Running head: Origin and diversification of the XTH family

Authors: Naoki Shinohara * and Kazuhiko Nishitani *

Addresses: Department of Biology, Kanagawa University, Tsuchiya 2946, Hiratsuka, 259-1293, Kanagawa, JAPAN

*Corresponding authors:
N. Shinohara
    Email: naoki-s@yd5.so-net.ne.jp
    FAX: +81 463596277
K. Nishitani
    Email: nishitani@kanagawa-u.ac.jp
    FAX: +81 463596277

ORCID
Naoki Shinohara: https://orcid.org/0000-0002-4029-239X
Abstract (244 words)

All land plants encode large multigene families of xyloglucan endotransglycosylase/hydrolases (XTHs), plant-specific enzymes that cleave and reconnect plant cell-wall polysaccharides. Despite the ubiquity of these enzymes, considerable uncertainty remains regarding the evolutionary history of the XTH family. Phylogenomic and comparative analyses in this study traced the non-plant origins of the XTH family to Alphaproteobacteria ExoKs, bacterial enzymes involved in loosening biofilms, rather than Firmicutes licheninases, plant biomass digesting enzymes, as previously supposed. The relevant horizontal gene transfer (HGT) event was mapped to the divergence of non-swimming charophycean algae in the Cryogenian geological period. This HGT event was the likely origin of charophycean EG16-2s, which are putative intermediates between ExoKs and XTHs. Another HGT event in the Cryogenian may have led from EG16-2s or ExoKs to fungal CRHs, enzymes that cleave and reconnect chitin and glucans in fungal cell walls. This successive transfer of enzyme-encoding genes may have supported the adaptation of plants and fungi to the ancient icy environment by facilitating their sessile lifestyles. Furthermore, several protein evolutionary steps, including coevolution of substrate-interacting residues and putative intra-family gene fusion, occurred in the land plant lineage and drove diversification of the XTH family. At least some of those events correlated with the evolutionary gain of broader substrate specificities, which may have underpinned the expansion of the XTH family by enhancing duplicated gene survival. Together, this study highlights the Precambrian evolution of life and the mode of multigene family expansion in the evolutionary history of the XTH family.

Keywords: Alphaproteobacteria • Cryogenian • Glycoside hydrolase family 16 • Horizontal gene transfer • Streptophyta • Xyloglucan endotransglycosylase/hydrolase
Introduction

Early speculation on the mechanisms underlying plant cell-wall extension proposed the existence of a hypothetical enzyme involved in cleavage and reconnection of cell-wall polysaccharides cross-linking cellulose microfibrils (Albersheim 1975). This prediction was later supported by the experimental detection of such enzymatic activities (Nishitani and Tominaga 1991; Smith and Fry 1991; Farkas et al. 1992; Fry et al. 1992) and substantiated by molecular cloning and characterization of the enzymes, named endo-xyloglucan transferases (Nishitani and Tominaga 1992; Okazawa et al. 1993; Purugganan et al. 1997; Campbell and Braam 1999a). The enzyme sequences demonstrated mutual homology and also exhibited homology with xyloglucan endo-β(1,4)-glucanase of nasturtium (de Silva et al. 1993). This multifaceted similarity initiated an analysis of the relationships between sequence and activity in the enzyme family (Nishitani 1997; Campbell and Braam 1999b). Subsequently, the first discovered enzyme class was renamed xyloglucan endotransglycosylase (XET), the xyloglucan glucanase class was renamed xyloglucan endohydrolase (XEH), and the larger encompassing enzyme family was termed the xyloglucan endotransglycosylase/hydrolase (XTH) family (Rose et al. 2002).

Crystallographic and mutational analyses revealed that both XET and XEH enzymes cleaved substrate xyloglucan chains via a catalytic glutamic acid residue nucleophile at the active center (Johansson et al. 2004; Baumann et al. 2007; Mark et al. 2009). In the reaction model of XEH, a cleaved substrate chain is transferred to a water molecule at the completion of the hydrolysis reaction (Baumann et al. 2007; Mark et
In the reaction model of XET, the cleaved substrate chain is transferred to the reducing end of another xyloglucan molecule at the completion of the transglycosylation reaction (Johansson et al. 2004; Baumann et al. 2007; Mark et al. 2009). In this context, the substrate molecule being cut is termed the donor substrate, and the connecting substrate is termed the acceptor substrate. On a single XTH molecule, the donor and acceptor binding sites are delimited by the catalytic nucleophile and together compose the substrate-binding cleft (Johansson et al. 2004; Baumann et al. 2007; Mark et al. 2009).

The biological significance of the XTH family is underscored by their prolific and highly conserved nature, and each land plant genome encodes 20–60 XTH genes (Yokoyama and Nishitani 2001; Eklöf and Brumer 2010; Yokoyama et al. 2010). Ongoing sequence-based classification of XTH genes has been conducted over the last two decades. Early attempts classified 23 cDNA sequences from nine angiosperms and the 33 genes (AtXTH1-AtXTH33) from the Arabidopsis thaliana genome into three major groups (hereafter I, II, and III) on the basis of phylogram topology (Nishitani 1997; Rose et al. 2002). With some modifications, including additional groups, this classification is widely accepted. Group III was divided into groups III-A and III-B and delineated XET and XEH activities, the latter of which is confined to group III-A (Baumann et al. 2007; Kaewthai et al. 2013). In contrast to this neat delineation, the incorporation of 29 members (OsXTH1-OsXTH29) of the rice XTH family into the phylogeny of the 33 A. thaliana XTHs introduced ambiguity into the boundary between groups I and II, and composite group I/II was introduced (Yokoyama et al. 2004). A small group named Ancestral was later separated from the rest of the composite group I/II (Baumann et al. 2007). As its name suggests, Ancestral group members form an early-diverging clade in
phylograms (Baumann et al. 2007; Yokoyama et al. 2010; Shinohara et al. 2017); however, surprisingly, genomic datasets of bryophytes (early-diverging land plants) do not include members of the Ancestral group (Yokoyama et al. 2010; Shinohara et al. 2017). Thus, the XTH family requires further research to clarify its classification.

Some XTHs have additional transglycosylation activities alongside their XET activity (Hrmova et al. 2007; Simmons et al. 2015; Shinohara et al. 2017; Stratilová et al. 2019; Stratilová et al. 2020). For example, EfHTG of *Equisetum fluviatile* (water horsetail) catalyzes two additional types of transglycosylation: cleavage of mixed-linkage β(1,3;1,4)-glucan (or MLG) and reconnection to xyloglucan, and cleavage of cellulose and reconnection to xyloglucan. This broader substrate specificity is not found in three closely related members (EfXTH-A, EfXTH-H, and EfXTH-I) of the *E. fluviatile* XTH family (Holland et al. 2020). Despite this variation in substrate specificity, all four XTHs belong to composite group I/II (Simmons et al. 2015; Holland et al. 2020). The classification of composite group I/II may therefore benefit from revision to delineate additional enzymatic activities of group members. The evolutionary implications of the broader substrate specificities of some XTH family members remain to be explained.

The recent rapid expansion of genome and transcriptome datasets has promoted phylogenomics-based classification of XTHs in the land plant lineage (Eklöf and Brumer 2010; Del-Bem and Vincentz 2010). Additional data were provided by a census of plant glycoside hydrolase family 16 (GH16), which included
more than 15,000 XTH sequences (Behar et al. 2018), and a subgrouping project of the whole GH16 family, including a large number of prokaryotic and eukaryotic enzymes (Viborg et al. 2019). These data resources are highly valuable for in-depth analysis of the evolutionary history of intra-family and inter-family relationships in the XTH family.

Considerable uncertainty remains regarding the evolutionary emergence of the XTH family. One report concluded that the first XTHs emerged in Zygnemataceae (Del-Ben and Vincentz 2010), a sister group of the land plants (Embryophyta). Another study found XTH-encoding transcripts in transcriptome datasets of non-charophycean green algae, remote relatives of land plants (Behar et al. 2018). Therefore, it remains unclear which algal lineage first acquired XTHs. Uncertainty also remains regarding the non-plant origins of XTHs. Bacterial licheninases, which digest plant biomass to provide nutrients to the bacteria (Stülke and Hillen 2000), were long considered as the non-plant ancestors of the XTH family because of their sequence similarities (Barbeyron et al. 1998; Michel et al. 2001). However, bacterial evolutionary gain of a plant digestive function presupposes the existence of plants for the bacteria to feed upon and, in this context, the plant-degrading function is unlikely to fully reflect the ancient function of the non-plant origin of XTHs. To address these questions, phylogenomic and comparative analyses were performed to provide insights regarding first, the evolutionary emergence of XTHs; second, the non-plant origin of the XTH family; and third, the molecular evolutionary events underlying diversification of the XTH family.
Results

Overview of the evolutionary history of the XTH family

Phylogenetic and comparative approaches were used to examine evolutionary relationships among XTH family members. The main conclusions are graphically summarized in Figures 1 and 2. Figure 1 outlines the inter-family relationship of XTHs and other related GH16 enzymes. As shown in Figure 1, Alphaproteobacteria ExoKs may have been adopted by plants and lost their extended loops to become plant EG16-2s, followed by gain of the acceptor-side extension to become the XTH family. Figure 2 includes a geographical timeline and plant evolutionary pathway. Intra-family relationships of XTHs are outlined, with their multiple diversification events shown relative to protein evolution.

Emergence of the XTH family

Available plant transcriptome datasets with XTH-encoding sequences include non-charophycean green algae species *Dunaliella salina*, *D. tertiolecta*, and *Pseudoscurfieldia mariana* (Behar et al. 2018), which are remotely related to land plants. Phylogenetic analysis (Fig. S1) revealed that these sequences were more closely related to angiosperm XTHs than to liverwort XTHs. The incongruence between the protein tree (Fig. S1) and the currently accepted species tree (Wickett et al. 2014) may be explained by variable sampling procedures for their transcriptome datasets or by multiple horizontal gene transfer (HGT) events.
from land plants to algae. However, in either case, the XTHs are unlikely to be direct descendants of the first XTHs to emerge.

As previously reported (Del-Bem and Vincentz 2010), an early lineage of the XTH family was found in freshwater-dwelling charophycean algae belonging to Zygnematophyceae, a sister group of land plants that is a late-diverging clade of Charophyta (Timme et al. 2012). In a prior census (Behar et al. 2018), nine of the 46 charophycean species examined had XTH transcripts, and all nine belonged to Zygnematophyceae. In addition, two recently sequenced Zygnematophyceae genomes (Spirogloea muscicola and Mesotaenium endlicherianum) harbored multiple XTH genes (Cheng et al. 2019; Fig. 3), and the genome of Klebsormidium nitens, a charophycean alga that diverged earlier than Zygnematophyceae, contained no XTH genes (Hori et al. 2014; Bowman et al. 2017; Fig. 4). Accordingly, the emergence of the XTH family was mapped to the last common ancestor of Zygnematophyceae and land plants (Fig. 2).

Ancestry of the XTH family in the plant kingdom

Common shared ancestry between plant XTHs and bacterial licheninases was proposed because of their sequence similarities (Barbeyron et al. 1998; Michel et al. 2001). Licheninases, which have alternative names in the literature (e.g., lichenases or MLGases), hydrolyze linear polysaccharides found in lichens and cereals. These polysaccharides, termed MLGs, are composed of β(1,3)- and β(1,4)-glucosyl residues,
and licheninases specifically cleave the β(1,4)-glucosyl linkage subsequent to the β(1,3)-glucosyl residue. Evolutionary intermediate groups have now been recognized between plant XTHs and bacterial licheninases (Eklöf et al. 2013; McGregor et al. 2017; Behar et al. 2018; Viborg et al. 2019; Fig. 1).

Although EG16s were previously included in the XTH family (Yokoyama et al. 2010), they now have their own group name (Eklöf et al. 2013; McGregor et al. 2017; Behar et al. 2018), belonging to the GH16 subgroup closest to the XTH family (Viborg et al. 2019). EG16s have licheninase-like endo-glycosidase activity with broad substrate specificity that cleaves both MLG and xyloglucan (Eklöf et al. 2013; McGregor et al. 2017). Together with their shared protein structural features, this activity prompted the proposal that EG16s were licheninase-to-XTH intermediates (Eklöf and Brumer 2010; McGregor et al. 2017). Recently, another plant enzyme group, termed EG16-2, was identified (Behar et al. 2018). EG16-2s are structurally and phylogenetically close to EG16s but remain biochemically uncharacterized (Behar et al. 2018). In a subgrouping analysis of the GH16 family, the XTH family was assigned its own subgroup number, GH16_20, but no such number has yet been assigned to the EG16 group (Viborg et al. 2019).

The plant origin of the composite clade of EG16-2s, EG16s, and XTHs was explored. K. nitens encodes six EG16-2s, but no EG16s or XTHs (Hori et al. 2014; Bowman et al. 2017; Fig. 4). Two charophycean algae that diverged earlier than K. nitens, Mesostigma viride and Chlorokybus atmophyticus, did not encode any EG16-2s, EG16s, or XTHs (Behar et al. 2018; Del-Bem 2018; Wang et al. 2020; Fig. 4). Unlike XTHs, EG16s lack signal peptides for apoplastic secretion (Behar et al. 2018). However, five of the
six EG16-2s in the *K. nitens* genome v1.1 dataset (Hori et al. 2014) were predicted to have conventional signal peptides (Figs 1, 4), suggesting that apoplastic secretion was the ancestral form in the composite clade of EG16-2s, EG16s, and XTHs. EG16-2s exhibited early emergence and lacked derived structural characteristics found in EG16s and XTHs, namely the absence of a signal peptide in EG16s and the presence of an acceptor-side extension in XTHs (Fig. 1), and EG16-2s therefore form the putative ancestor of both EG16s and XTHs. Furthermore, two charophycean algae belonging to Coleochaetales have *EG16* transcripts but no *XTH* transcripts (Behar et al. 2018). Coleochaetales resides phylogenetically between the Klebsormidiophyceae and Zygnematophyceae (Wickett et al. 2014), indicating that the emergence of EG16s predated the emergence of XTHs (Figs 1, 2).

No EG16 sequences were attributed to non-charophycean green algae in an earlier census (Behar et al. 2018), but some EG16-2 sequences were found in the algae species *Botryococcus braunii* and *Interfilum paradoxum* (Behar et al. 2018). The latter of these was originally classified as a non-charophycean green alga (Chodat 1922) but was subsequently reclassified as a charophycean alga closely related to *Klebsormidium* (Mikhailyuk et al. 2008). Therefore, of the two species, only *B. braunii* was a non-charophycean green alga. Furthermore, when the two fragment EG16-2 sequences identified from *B. braunii* (Behar et al. 2018) were subjected to BLAST searches against the JGI *B. braunii* v2.1 genome dataset, no corresponding genes were found. The fragment sequences also lacked the signature residues common to other EG16-2s and were therefore relegated to an unclassified category (Dataset S1). The emergence of EG16-2s was thus mapped to the origin of Klebsormidiophyceae (Fig. 2).
Out of the plant kingdom

Accumulating evidence indicates that HGT events are common among virtually all the phylogenetic branches of life (Husnik and McCutcheon 2018), and HGT is a potential explanation for the non-plant origin of XTHs. In a GH16 subgrouping study (Viborg et al. 2019), the plant-specific composite clade of EG16s and XTHs is closely related to the GH16_21 group, which is enriched with bacterial enzymes including licheninases, and the composite fungal enzyme clade of GH16_18 and GH16_19. A phylogram of representative sequences from the groups was constructed, and charophycean EG16-2s were found to be closest to the bacterial enzymes of the GH16_21 group (Fig. 5). This topology supported the notion that the non-plant origin of XTHs was a bacterial gene, as proposed previously (Barbeyron et al. 1998; Michel et al. 2001; Eklöf et al. 2013; McGregor et al. 2017; Viborg et al. 2019).

The fungal origin of the composite group of GH16_18 and GH16_19 was explored further, and group members were found in Zoopagomycota and later diverging fungi (Fig. S2). The common ancestral gene of GH16_18 and GH16_19 may therefore have transferred to the common ancestor of terrestrial fungi, whose emergence was associated with the loss of flagella and the gain of hyphal growth (Naranjo-Ortiz and Gabaldón 2019). Together, GH16_18 and GH16_19 are termed the CRH family. The CRH family derives its name from yeast knockout mutants with increased sensitivity to cell-wall-interfering compounds (e.g., Congo red and calcofluor white), termed Congo Red Hypersensitive mutants.
(Rodríguez-Peña et al. 2000). GH16_18 and GH16_19 both include biochemically characterized members (Fig. S2), all of which mediate cleavage of a polysaccharide chain of chitin, β-(1,4)-linked N-acetylglucosamine homopolymer, and reconnection of the cleaved chain to another molecule of chitin, β-(1,3)-glucan, or β-(1,6)-glucan (Arroyo et al. 2016; Fang et al. 2019). This reaction resembles that of XTHs in that both are transglycosylation reactions involving the endolytic cleavage of cellulose analogs, as previously noted (Hrmova et al. 2009). The protein tree topology (Fig. 5) and the structural features (Fang et al. 2019; Fig. 1) suggest that the CRH family may have originated from plant EG16-2s or bacterial GH16 enzymes.

The bacterial origin of the XTH family

Bacterial licheninases (Hahn et al. 1995; Furtado et al. 2011), long-standing candidates for the origin of plant XTHs (Barbeyron et al. 1998; Michel et al. 2001), were recently classified into the GH16_21 group (Viborg et al. 2019). To revisit bacterial origins, a protein tree of GH16_21 members and charophycean EG16-2s was constructed and superimposed on the taxonomy of source organisms and the occurrence of known enzymatic activity (Fig. 6). This analysis revealed that bacterial GH16_21 proteins are encoded by single-copy genes, contrasting with the dozens of XTH genes in land plant genomes. The GH16_21 proteins were divided into two major clades that reflected differences in bacterial taxonomic groups and enzymatic activity (Fig. 6). One clade comprised Firmicutes enzymes, including well-characterized
licheninases of *Bacillus* and *Clostridium* species. The other clade comprised Alphaproteobacteria enzymes, including ExoKs. Charophycean EG16-2s were nested within the latter clade (Fig. 6).

An *ExoK* gene was first discovered in a gene cluster of *Ensifer meliloti* (formerly *Rhizobium meliloti* or *Sinorhizobium meliloti*; Glucksmann et al. 1993), a nitrogen-fixing bacterium dwelling in root nodules of leguminous hosts (Corbin et al. 1983). The gene cluster is involved in the biosynthesis, export, and modification of succinoglycan (Mendis et al. 2013), a branched acidic polysaccharide found in biofilms of Proteobacteria (Halder et al. 2017). ExoKs cleave the β(1,4)-glucosyl linkage subsequent to the β(1,3)-galactosyl residue of succinoglycan (Staehelin et al. 2006) and liberate oligosaccharides that facilitate symbiosis between bacteria and host plants (Mendis et al. 2013). However, succinoglycan occurs in bacteria that have no clear symbiotic relationships with plants (Halder et al. 2017), and conservation of *ExoK* genes in such bacteria (Fig. 6) suggest a more general function for ExoKs. Recently, an *ExoK* knockout mutant of *Agrobacterium* sp. ATCC 31749 was reported to produce a firmer biofilm (Gao et al. 2020). Accordingly, biofilm-loosening ExoK-like enzymes are the likely bacterial origin of plant XTHs (Figs 1, 2).

Alphaproteobacteria ExoKs and Firmicutes licheninases exhibited similar folding structures (Fig. 1) but also had notable variations in amino acid residues at the aligned sites corresponding to Q46, S67, and H83 of mature PttXET16-34 (Fig. 7). The Q46 and H83 sites were located in the substrate-binding cleft, but the Q and H residues of ExoKs provided more space than did the R and W residues of licheninases (Fig. 7).
Considered alongside their different substrate specificities (Planas 2000; Staehelin et al. 2006), this structural variation supports the notion that ExoKs and licheninases are functionally differentiated. The deep branches in the protein tree (Fig. 6) suggest that divergence occurred as early as the last common ancestor of Alphaproteobacteria and Firmicutes, which can be traced to more than three billion years ago (Moore et al. 2017). The distinct differences in cell-wall structure and composition between those bacterial groups are widely recognized and visualized by Gram staining (Battistuzzi and Hedges 2009) and may underlie the self/nonself targeting nature of these enzyme groups. Firmicutes may have gained licheninase-like enzymes during evolution to digest other bacteria in the ancient prokaryotic world, where such bacterial cells were postulated to be almost the only substrates for heterotrophs (Schönheit et al. 2016).

GH16_21 comprises numerous bacterial enzymes and a small number of eukaryotic enzymes (Viborg et al. 2019). The eukaryotic enzymes were nested within the clade of Firmicutes licheninases and confined to Neocallimastigomycota (Fig. 6), a group of anaerobic fungi dwelling in the gastrointestinal tracts of herbivorous animals (Hibbett et al. 2007). HGT events of plant-degrading enzyme genes from bacteria are rampant in this fungal lineage and are postulated to coincide with or postdate the evolutionary habitat transition to the guts of herbivores (Murphy et al. 2019). This transition is estimated to have occurred close to 70 million years ago (Wang et al. 2019), much later than the emergence of XTHs in plants (~600 million years ago; Figs 1, 2). The differences in their putative bacterial donors and the estimated timings of the
relevant HGT events indicate that the GH16_21 proteins of Neocallimastigomycota are essentially unrelated to the non-plant origin of the XTH family.

*Early divergence of XTHs in the land plant lineage*

Phylogenetic analysis revealed that the diversification of XTH groups I and III predated the emergence of land plants (Figs 3, S3), but it was unclear which group emerged first. Genome datasets from bryophytes (Rensing et al. 2008; Bowman et al. 2017; Zhang et al. 2020), the basal lineage of the land plants (Wickett et al., 2014), included only these two groups (Dataset S1; Fig. S3), and the bryophyte XTHs were equidistant from the algal XTHs (Fig. S3). Group III XTHs include a Group III-specific extension in loop 2 (Eklöf and Brumer 2010). This extension was absent in algal XTHs and was confined to group III throughout the land plant lineage (Dataset S1; Fig. S3). This indicates that group I was the ancestral lineage from which group III arose (Fig. 2). Differences in enzymatic activity between groups I and III remain unknown, and both groups have XET activity (Eklöf and Brumer 2010).

*Divergence of group II XTHs in euphyllophytes*

A previous protein tree of the *A. thaliana* XTH family suggested that group II members formed a distinct monophyletic clade (Rose et al. 2002). However, incorporation of the *Oryza sativa* XTH family into the protein tree introduced ambiguity into the group II boundary and led to the introduction of composite
group I/II (Yokoyama et al. 2004). The signature residues of group II *A. thaliana* XTH family members were therefore examined in this context, and H102 and Q116 were identified (Fig. 8). Q116 was previously reported as the signature residue of group II (Eklöf and Brumer 2010), and the two corresponding residues (Q102 and R116) of PttXET16-34 (group I) were adjacent to the catalytic nucleophile E89, allowing interaction with the β-(1,4)-glucan backbone of the xyloglucan substrate (Johansson et al. 2004; Fig. 8). Examination of amino acid variation at the two sites in an evolutionary context revealed a coevolutionary link between the sites (Figs 3, S3). Most of the examined XTH sequences had either the Q102-R116 or the H102-Q116 combination, the former of which was predominant and was the ancestral form (Figs 3, S3). The evolutionary shift to the H102-Q116 combination occurred at least five times: once in a charophycean alga, twice in bryophytes, once in *Selaginella moellendorffii*, and once in the euphyllophyte lineage that includes ferns and seed plants (Figs 3, S3). Similar to convergent evolution of proteins (Uchiyama et al. 2020), the first four events were instances of convergent coevolution of amino acid residues, but the last event led to the formation of group II (Fig. 3). In this circumscription, group II of the *O. sativa* XTH family comprised 15 members, two of which (OsXTH4 and OsXTH17) harbored secondary mutations (Fig. 3). These secondary mutations may underlie the previous fusion of groups I and II.

Within the group of horsetail (*E. fluviatile*) XTHs, EfHTG has broader substrate specificities than EfXTH-A, EfXTH-H, or EfXTH-1 (Simmons et al. 2015; Holland et al. 2020). Similarly, nasturtium (*Tropaeolum majus*) TmXET6.3 has a broad substrate specificity that has not been found in other
nasturtium XTHs (Stratilová et al. 2019). In the aforementioned circumscription, both EfHTG and TmXET6.3 belonged to group II (Dataset S1; Fig. 3), and EfXTH-A, EfXTH-H, and EfXTH-1 belonged to group I (Dataset S1). Group I was ancestral to composite group I/II (Fig. 3), and the broad substrate specificities of EfHTG and TmXET6.3 can therefore be viewed as an evolutionary gain caused by the coevolution of the substrate-interacting residues.

**Divergence of group III-A in seed plants**

By comparison with XTHs in composite group I/II, group III XTHs contain an extension in loop 2 (Eklöf and Brumer 2010), and this group is divided into groups III-A and III-B (Baumann et al. 2007). XTHs in Group III-A predominantly have XEH activity (Baumann et al. 2007; Kaewthai et al. 2013). This activity is a derived characteristic attributed to the loop 2 extension, which is longer than the group III-B loop extension (Baumann et al. 2007; Eklöf and Brumer 2010). Multiple-alignment of land plant XTH sequences showed that the III-A-type loop extension was confined to seed plants, being found in conifers, cycads, and angiosperms, but not in ferns, lycophytes, or bryophytes (Table S1). TmNXG1, the first identified member of group III-A, is postulated to be involved in the mobilization of seed-storage xyloglucan (de Silva et al. 1993), and some other members of group III-A are localized to rapidly growing tissues (Tabuchi et al. 2001; Kaewthai et al. 2013; Hara et al. 2014). Therefore, it is likely that the emergence of group III-A is correlated with the evolutionary development of seeds and rapid seedling growth.
Divergence of the Ancestral group of XTHs in core angiosperms

Four members (AtXTH1, AtXTH2, AtXTH3, and AtXTH11) of the *A. thaliana* XTH family, first reported as members of group I (Rose et al. 2002), were reclassified as a distinct small group within the expanded collection of XTH sequences (Baumann et al. 2007; Miedes and Lorences 2007). This small group is called the “Ancestral” group (Baumann et al. 2007) or “group IV” (Miedes and Lorences 2007). As the name “Ancestral” suggests, this group displays deep phylogram branching (Baumann et al. 2007; Yokoyama et al. 2010; Shinohara et al. 2017; Fig. 3). However, early-diverging land plants such as bryophytes and lycopods encode no members of the Ancestral group (Yokoyama et al. 2010; Shinohara et al. 2017; Figs 3), and this constitutes an incongruity between the protein and species trees.

Signature residues of the Ancestral group were examined to address the protein/species discrepancy (Fig. S4). A comparison between the Ancestral group and other XTH groups identified no signature residues, but a comparison between the Ancestral group and the remaining members of composite group I/II identified two signature residues in the N-terminal quarter region (Fig. S4). The two residues were found preferentially in both the Ancestral group and group III, suggesting that an intra-family gene fusion between groups I and III led to the emergence of the Ancestral group. Gene fusions are known to occur between promoter and coding regions as well as between different coding regions, presumably in a random fashion, and a small percentage of genes have undergone traceable inter-protein fusion events.
(Durrens et al. 2008). In addition, many examples of circularly permuted gene pairs have been identified whose evolutionary origins are explained by intra-family gene fusion (Bliven and Prlić 2012). These observations support the hypothesis that intra-family gene fusion events may have occurred and descended over generations at non-negligible levels. The deep branch of the Ancestral group can be explained by the early divergence of their parental groups, but is not explained by early emergence of this group per se.

An early lineage of plants encoding members of the Ancestral group was examined. Amborella genome and Austrobaileyales transcriptome datasets did not include any group members, but transcriptome datasets of Peperomia fraseri and Houttuynia cordata, both early-diverging core angiosperms, did include some Ancestral group members (Fig. S5). The origin of the Ancestral group was thus traced to core angiosperms (i.e., Mesangiospermae). Some monocots (e.g., O. sativa and Zea mays) also encoded some Ancestral XTHs (Figs S5, S6), but others, namely Spirodea polyrhiza (common duckmeat) and Dioscorea alata (purple yam), did not (Figs S5, S6). The eudicot tree Eucalyptus grandis (flooded gum) encoded a relatively large number of Ancestral group members (five; Fig. S6). Thus, both lineage-specific loss and expansion of the Ancestral group occurred in core angiosperms. The emergence of the Ancestral group was mapped as the latest event subsequent to the aforementioned divergence of groups I-to-III, I-to-II, and III-A-to-III-B (Fig. 3). In this light, the term “group IV” (Miedes and Lorences 2007) may be preferable to “Ancestral” for future use.
Four *A. thaliana* XTHs were members of the Ancestral group. Of these, AtXTH3 had cellulose:cellulooligosaccharide and cellulose:xyloglucan-oligosaccharide transglycosylation activities alongside its XET activity (Shinohara et al. 2017). This broad substrate specificity resembles the activities of EfHTG (Simmons et al. 2015) and TmXET6.3 (Stratilová et al. 2019). However, AtXTH3 neither had the H102-Q116 combination nor belonged to group II (Fig. 3). Therefore, it is likely that multiple types of protein evolution have led to the evolutionary gain of broad substrate specificities in the XTH family.
Discussion

Co-occurrence of XTHs and aligned cellulose microfibrils

In the evolutionary history of the plant cell wall, the origins of the rosette-type cellulose-synthase complex and some of the core component proteins that facilitate the association between cellulose synthases and cortical microtubules are found in Zygnematophyceae (Lampugnani et al. 2019). This corresponds with the origin of the XTH family. Thus, there is a noticeable evolutionary shift in the plant cell wall between Coleochaetophyceae (and earlier charophycean algae) and the composite clade of Zygnematophyceae and land plants. This evolutionary change theoretically allows plant cells to grow in a microtubule-oriented manner with the aid of XTHs and is consistent with the previously reported notion that the function of XTHs is the reversible loosening of plant cell walls composed of aligned cellulose microfibrils enmeshed in matrix polysaccharides (Albersheim 1975; Nishitani 1997).

The mechanism underlying the expansion of the XTH multigene family

The genomes of two Zygnematophyceae species (S. muscicola and M. endlicherianum) contained multigene XTH families (Cheng et al. 2019). These two algae are unicellular (Cheng et al. 2019), and the expansion of the XTH family in Zygnematophyceae is therefore unlikely to be driven only by the acquisition of a multicellular body plan, which often involves multiple types of cell differentiation. The
question of how duplicated genes survive and undergo neofunctionalization over evolutionary time scales has been debated for decades and various theoretical models have been proposed (Soskine and Tawfik 2010). Currently favored models entail latent and promiscuous protein functions to explain both survival and neofunctionalization of duplicated genes (Soskine and Tawfik 2010). Some XTHs have additional enzymatic activities in addition to XET activity (Simmons et al. 2015; Shinohara et al. 2017; Stratilová et al. 2019; Stratilová et al. 2020), and this broader activity is likely to have emerged multiple times during evolution. Thus, the expansion and diversification of the XTH family may have been driven not only by the acquisition of new enzymatic activities (e.g., XEH in group III-A), but also by the relaxation of substrate specificity (e.g., the broader specificities of group II).

Vehicles for seemingly altruistic gene transfer

The bacterial origin of plant XTHs and other related enzymes, such as EG16s and EG16-2s, appeared to be Alphaproteobacteria ExoKs rather than the previously supposed Firmicutes licheninases. Alphaproteobacteria species are not uncommon donors of prokaryote-to-eukaryote HGT events, and the mechanisms responsible may include type IV secretion systems, gene transfer agents (GTAs), and Ti plasmids (Le et al. 2014). Of those mechanisms, GTAs are widely distributed in Rhodobacterales (Lang et al. 2012), whose ExoK-like enzymes are phylogenetically proximal to the root of charophycean EG16-2s. GTAs are phage-like particles loaded with random genomic fragments of donor bacteria (Lang et al. 2012) and are capable of transforming evolutionarily unrelated bacteria with high frequency (McDaniel et al. 2012).
2010). It is conceivable that the hypothetical HGT event of an Alphaproteobacteria *ExoK*-like gene to a *Klebsormidium*-related charophyte may have been caused by GTAs rather than by an accidental intake of naked DNA.

*Adaptation of plants to an ancient icy environment*

The HGT event from *ExoK*-like enzymes in bacteria to EG16-2s in plants was mapped to the time of divergence between *M. viride*, free-swimming freshwater algae covered with mineralized scaly cell walls (Domozych et al. 1991), and *Klebsormidium*, sessile but airborne algae that are found globally, including in polar regions (Ryšánek et al. 2016). The HGT event coincides with the evolutionary gain of the sessile vegetative phase in plants (Nishiyama et al. 2018) and the biosynthesis of xyloglucan (Del-Bem 2018; Herburger et al. 2018). Recent estimates (Donoghue and Paps 2020) mapped this HGT event to the Cryogenian (720–635 million years ago). The climate in this geological period was hypothesized to have led to global glaciation or Snowball Earth (Hoffman et al. 2017), and sessile plant traits on the modern Earth may have been co-opted as an adaptation to the ancient icy environment. In this light, acquired EG16-2s may have played a role in controlled cell-wall-loosening under sporadically occurring favorable conditions for growth on ice. Another possible role of EG16-2s may have been the modification of secreted polysaccharides, changing ice nucleation and allowing algae to effectively sunbathe on ice.

Recently, high xyloglucan contents were reported in soil of a glacial forefield (Galloway et al. 2018), and environmental xyloglucan secreted by plant roots and rhizoids was proposed to have a pivotal role in the
formation of the soil biosphere (Galloway et al. 2018). Such secretions may have originated from ancient charophycean snow algae.

The trinity of bacteria, plants, and fungi

The HGT event leading to the development of the fungal CRH family may have coincided with the emergence of terrestrial fungi. Fungal terrestrialization involved the loss of flagellar-mediated swimming ability and the gain of hyphal growth. Acquisition of these features overlapped with the Cryogenian period, which led to the notion that fungal terrestrialization took place in the ancient icy environment (Naranjo-Ortiz and Gabaldón 2019). In this light, as with the acquisition of EG16-2s by charophycean plants, the acquisition of CRHs by fungi may have assisted growth by controlling cell-wall-loosening under sporadically occurring favorable conditions on ice, allowing adaptation to the Snowball Earth. Overall, the ancient successive transfer of GH16 enzymes from bacteria to plants and fungi may have provided the basis for the sessile lifestyles of plants and fungi.

Conclusions

The aims of this study were to improve understanding of the evolutionary origins and development of XTHs. Several conclusions can be drawn. (1) The likely non-plant origin of the plant-specific XTH family is Alphaproteobacteria ExoKs, which cleave succinoglycans and loosen bacterial biofilms. This functional
framework is conserved among its putative descendants, including XTHs, which cleave xyloglucan in plant cell walls, and CRHs, which cleave chitin in fungal cell walls. (2) Free-swimming scaly algae related to *M. viride* may have acquired Alphaproteobacteria *ExoK*-like genes during the Cryogenian period. This HGT event may have assisted algae to develop sessile lifestyles and adapt to the ancient icy environment. Putative direct descendants of the acquired enzymes are *Klebsormidium* EG16-2s. (3) Flagellated fungi in the Cryogenian period may have experienced an HGT event leading to the fungal CRHs. This HGT event may have aided development of hyphal growth in fungi and survival in the ancient icy environment. (4) The first XTHs may have evolved in desmid-related freshwater algae through extension of the C-terminal region of EG16-2s. This event coincides with the evolutionary gain of microtubule-oriented directional growth, which is well conserved across the land plants. (5) Multiple XTH family diversification events occurred, primarily in the land plant lineage. These diversification events correlated with protein evolution (loop extension, coevolution of substrate-interacting residues, and intra-family gene fusion) and also correlated with the emergence of some plant taxonomic groups (land plants, euphyllophytes, and seed plants). At least some of the protein evolutionary events were related to the relaxation of substrate specificity, and this relaxation may have underpinned XTH family expansions. Overall, this study provides new insights into the evolutionary history of the XTH family and will facilitate future analysis of related enzymes.
Materials and Methods

Collection of protein sequences

Protein sequences of XTHs, EG16s, and EG16-2s were retrieved from public databases and published datasets. Identities and sequences are collated in Dataset S1. Retrieved sequences were examined using multiple sequence alignments and phylogenetic analysis to obtain a non-redundant set of sequences for each genome. Transcriptome datasets from a prior census (Behar et al. 2018) were used for analysis when genome datasets for species of interest were unavailable at the time of analysis. Bacterial and fungal sequences of the GH16_18, 19, 21, 22, and 23 subfamilies (Viborg et al. 2019) were retrieved from the CAZy database. Taxonomically representative and/or biochemically characterized microbial sequences were used for analysis and are listed in Dataset S1.

Indexing of protein sequence data

Sequences used in this study are detailed in Dataset S1. This comprises a “Summary” worksheet and four separate worksheets for enzyme groups “GH16_21”, “EG16s” (+ EG16-2s), “XTHs”, and “Fungal” GH16_18, _19, _22, and _23 enzymes. The “Summary” sheet details the members in each enzyme group and provides references within this study (e.g., figures and tables). The four enzyme group worksheets collate the protein sequences and their source organisms and are formatted as in the prior census (Behar et
al. 2018). Group information is also included where possible. Fragmentary sequences (often less than ~200 residues) and presumably mispredicted or pseudogenized protein sequences, some of which were omitted from phylogenetic analysis, are retained within Dataset S1 and are marked accordingly.

Species trees and time scale estimates

Classification of organisms was based on the NCBI taxonomy browser and the literature – for Streptophyta (Wickett et al. 2014), Zygnematophyceae (Cheng et al. 2019), Lycophyta (Field et al. 2016), Chlorophyta (Leliaert et al. 2012), fungi (Hibbett et al. 2007; Naranjo-Ortiz and Gabaldón 2019), and bacteria (Battistuzzi and Hedges 2009; Castelle and Banfield 2018). Estimates of divergence times and placement of evolutionary gain events were based on the literature – for divergence time in plant (Langdale and Harrison 2008; Donoghue and Paps 2020), fungal (Berbee et al. 2017; Tedersoo et al. 2018; Naranjo-Ortiz and Gabaldón 2019), and bacterial (Battistuzzi and Hedges 2009; Moore et al. 2017) lineages, and for the evolutionary gain events in plants (Del-Bem 2018; Nishiyama et al. 2018; Lampugnani et al. 2019).

User interfaces for sequence analysis

The SignalP 4.1 server (Petersen et al. 2011) was used for signal peptide prediction. SeaView software (Gouy et al. 2010) was used for MUSCULE sequence alignment (Edgar 2004), maximum-likelihood tree
construction (Guindon et al. 2010), and calculation of approximate likelihood-ratio test (aLRT) (Anisimova and Gascuel 2006) branch support values. The ggtree package (Yu et al. 2017) was used for tree rendering and graphical annotation.

**Computation of mutual information**

Mutual information values for amino acid variation (including gaps) and grouping of aligned sequences were calculated as previously reported (Weckwerth and Selbig 2003). Values were normalized according to Equation 4 (Kvålseth 1987) to allow comparison between datasets. An overview of the procedure is summarized (Fig. S7).

**Comparison of three-dimensional protein structures**

Three-dimensional coordinates of protein structures were retrieved from the RCSB Protein Data Bank (Burley et al. 2020) or were modeled on and retrieved from the SWISS-MODEL server (Waterhouse et al. 2018). The UCSF Chimera software (Pettersen et al. 2004) was used for comparison and rendering of the protein structures. The N-terminal loop extension (A45–R59 in 3o5s; S53–S67 in P33693) and donor-side (formerly C-terminal) extension (C207–I272 in 1umw) were annotated as described previously (Behar et al. 2018). The acceptor-side extension (K24–S41, A245–D266 in 6ibw) was annotated in this study.
Data Availability Statement

The data that supports the findings of this study are available in the Supplementary Data section of the online version of this article.

Supplementary Data

Supplementary Data are available at PCP online.

Funding

This work was in part supported by Grants-in-Aid from Japan Society for the Promotion of Science (JP24114005, JP17H03691, and JP18H05489 to K.N.).

Acknowledgments

We thank Tomoaki Nishiyama for valuable advice in computational analysis.

Disclosures

The authors declare that they have no conflict of interest.

Author Contributions

Naoki Shinohara: Conceptualization, Investigation, Writing - original draft, Writing - review & editing.
Kazuhiko Nishitani: Conceptualization, Writing - review & editing, Funding acquisition.
References

Albersheim, P. (1975) The walls of growing plant cells. *Sci. Am.* 232: 80-95.

Anisimova, M. and Gascuel, O. (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 55: 539-552.

Arroyo, J., Farkas, V., Sanz, A.B. and Cabib, E. (2016) Strengthening the fungal cell wall through chitin–glucan cross-links: effects on morphogenesis and cell integrity. *Cell. Microbiol.* 18: 1239-1250.

Barbeyron, T., Gerard, A., Potin, P., Henriissat, B. and Kloareg, B. (1998) The kappa-carrageenase of the marine bacterium *Cytophaga drobachiensis*. Structural and phylogenetic relationships within family-16 glycoside hydrolases. *Mol. Biol. Evol.* 15: 528-537.

Battistuzzi, F.U. and Hedges, S.B. (2009) A major clade of prokaryotes with ancient adaptations to life on land. *Mol. Biol. Evol.* 26: 335-343.

Baumann, M.J., Eklof, J.M., Michel, G., Kallas, A.M., Teeri, T.T., Czjzek, M., et al. (2007) Structural evidence for the evolution of xyloglucanase activity from xyloglucan endo-transglycosylases: biological implications for cell wall metabolism. *Plant Cell* 19: 1947-1963.

Behar, H., Graham, S.W. and Brumer, H. (2018) Comprehensive cross-genome survey and phylogeny of glycoside hydrolase family 16 members reveals the evolutionary origin of EG16 and XTH proteins in plant lineages. *Plant J.* 95: 1114-1128.

Berbee, M.L., James, T.Y. and Strullu-Derrien, C. (2017) Early diverging fungi: diversity and impact at the dawn of terrestrial life. *Annu. Rev. Microbiol.* 71: 41-60.

Bliven, S. and Prlić, A. (2012) Circular Permutation in Proteins. *PLoS Comput. Biol.* 8: e1002445.

Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., et al. (2017) Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* 171: 287-304. e215.

Burley, S.K., Bhukadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G.V., et al. (2020) RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* 49: D437-D451.

Campbell, P. and Braam, J. (1999a) *In vitro* activities of four xyloglucan endotransglycosylases from *Arabidopsis*. *Plant J.* 18: 371-382.

Campbell, P. and Braam, J. (1999b) Xyloglucan endotransglycosylases: diversity of genes, enzymes and potential wall-modifying functions. *Trends Plant Sci.* 4: 361-366.
Castelle, C.J. and Banfield, J.F. (2018) Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* 172: 1181-1197.

Cheng, S., Xian, W., Fu, Y., Marin, B., Keller, J., Wu, T., et al. (2019) Genomes of subaerial Zygmatophyceae provide insights into land plant evolution. *Cell* 179: 1057-1067. e1014.

Chodat, R. (1922) Matériaux pour l’histoire des algues de la Suisse. *Bull. Soc. Bot. Genève* série 2, 13: 66-114.

Corbin, D., Barran, L. and Ditta, G. (1983) Organization and expression of *Rhizobium meliloti* nitrogen fixation genes. *Proc. Natl. Acad. Sci., USA* 80: 3005-3009.

de Silva, J., Jarman, C.D., Arrowsmith, D.A., Stronach, M.S., Chengappa, S., Sidebottom, C., et al. (1993) Molecular characterization of a xyloglucan-specific endo-(1\(\rightarrow\)4)-\(\beta\)-D-glucanase (xyloglucan endo-transglycosylase) from nasturtium seeds. *Plant J.* 3: 701-711.

Del-Bem, L.-E. (2018) Xyloglucan evolution and the terrestrialization of green plants. *New Phytol.* 219: 1150-1153.

Del-Bem, L.-E. and Vincentz, M.G.A. (2010) Evolution of xyloglucan-related genes in green plants. *BMC Evol. Biol.* 10: 341.

Domozych, D.S., Wells, B. and Shaw, P.J. (1991) Basket scales of the green alga, *Mesostigma viride*: chemistry and ultrastructure. *J. Cell Sci.* 100: 397.

Donoghue, P. and Paps, J. (2020) Plant evolution: assembling land plants. *Curr. Biol.* 30: R81-R83.

Durrens, P., Nikolski, M. and Sherman, D. (2008) Fusion and fission of genes define a metric between fungal genomes. *PLoS Comput. Biol.* 4: e1000200.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32: 1792-1797.

Eklöf, J.M. and Brumer, H. (2010) The *XTH* gene family: an update on enzyme structure, function, and phylogeny in xyloglucan remodeling. *Plant Physiol.* 153: 456-466.

Eklöf, J.M., Shojania, S., Okon, M., McIntosh, L.P. and Brumer, H. (2013) Structure-function analysis of a broad specificity *Populus trichocarpa* endo-\(\beta\)-glucanase reveals an evolutionary link between bacterial licheninases and plant *XTH* gene products. *J. Biol. Chem.* 288: 15786-15799.

Fang, W., Sanz, A.B., Bartual, S.G., Wang, B., Ferenbach, A.T., Farkaš, V., et al. (2019) Mechanisms of redundancy and specificity of the *Aspergillus fumigatus* Crh transglycosylases. *Nat. Commun.* 10: 1669.

Farkas, V., Sulova, Z., Stratilova, E., Hanna, R. and Maclachlan, G. (1992) Cleavage of xyloglucan by nasturtium seed xyloglucanase and transglycosylation to xyloglucan subunit oligosaccharides. *Arch. Biochem. Biophys.* 298: 365-370.
Field, A.R., Testo, W., Bostock, P.D., Holtum, J.A. and Waycott, M. (2016) Molecular phylogenetics and the morphology of the Lycopodiaceae subfamily Huperzioideae supports three genera: *Huperzia*, *Phlegmariurus* and *Phylloglossum*. *Mol. Phylogenet. Evol.* 94: 635-657.

Fry, S.C., Smith, R.C., Renwick, K.F., Martin, D.J., Hodge, S.K. and Matthews, K.J. (1992) Xyloglucan endotransglycosylase, a new wall-loosening enzyme activity from plants. *Biochem. J.* 282: 821-828.

Furtado, G.P., Ribeiro, L.F., Santos, C.R., Tonoli, C.C., de Souza, A.R., Oliveira, R.R., et al. (2011) Biochemical and structural characterization of a β-1,3–1,4-glucanase from *Bacillus subtilis* 168. *Process Biochem.* 46: 1202-1206.

Galloway, A.F., Pedersen, M.J., Merry, B., Marcus, S.E., Blacker, J., Benning, L.G., et al. (2018) Xyloglucan is released by plants and promotes soil particle aggregation. *New Phytol.* 217: 1128-1136.

Gao, H., Xie, F., Zhang, W., Tian, J., Zou, C., Jia, C., et al. (2020) Characterization and improvement of curdlan produced by a high-yield mutant of *Agrobacterium* sp. ATCC 31749 based on whole-genome analysis. *Carbohydr. Polym.*: 116486.

Glucksmann, M.A., Reuber, T.L. and Walker, G.C. (1993) Genes needed for the modification, polymerization, export, and processing of succinoglycan by *Rhizobium meliloti*: a model for succinoglycan biosynthesis. *J. Bacteriol.* 175: 7045-7055.

Gouy, M., Guindon, S. and Gascuel, O. (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27: 221-224.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59: 307-321.

Hahn, M., Pons, J., Planas, A., Quevillon, E. and Heinemann, U. (1995) Crystal structure of *Bacillus licheniformis* 1,3-1,4-β-d-glucan 4-glucanohydrolase at 1.8 Å resolution. *FEBS Letters* 374: 221-224.

Halder, U., Banerjee, A. and Bandopadhyay, R. (2017) Structural and functional properties, biosynthesis, and patenting trends of bacterial succinoglycan: a review. *Indian J. Microbiol.* 57: 278-284.

Hara, Y., Yokoyama, R., Osakabe, K., Toki, S. and Nishitani, K. (2014) Function of xyloglucan endotransglucosylase/hydrolases in rice. *Ann. Bot.* 114: 1309-1318.

Herbarger, K., Ryan, L.M., Popper, Z.A. and Holzinger, A. (2018) Localisation and substrate specificities of transglycanases in charophyte algae relate to development and morphology. *J. Cell Sci.* 131: jcs203208.

Hibbett, D.S., Binder, M., Bischoff, J.F., Blackwell, M., Cannon, P.F., Eriksson, O.E., et al. (2007) A higher-level phylogenetic classification of the Fungi. *Mycol. Res.* 111: 509-547.

Hoffman, P.F., Abbot, D.S., Ashkenazy, Y., Benn, D.I., Brocks, J.J., Cohen, P.A., et al. (2017) Snowball Earth climate dynamics and Cryogenian geology-geobiology. *Sci. Adv.* 3: e1600983.
Holland, C., Simmons, T.J., Meulewaeter, F., Hudson, A. and Fry, S.C. (2020) Three highly acidic *Equisetum* XTHs differ from hetero-trans-β-glucanase in donor substrate specificity and are predominantly xyloglucan homo-transglucosylases. *J. Plant Physiol.* 153210.

Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., et al. (2014) *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* 5: 1-9.

Hrnova, M., Farkas, V., Harvey, A.J., Lahnstein, J., Wischmann, B., Kaewthai, N., et al. (2009) Substrate specificity and catalytic mechanism of a xyloglucan xyloglucosyl transferase HvXET6 from barley (*Hordeum vulgare* L.). *FEBS J.* 276: 437-456.

Hrnova, M., Farkas, V., Lahnstein, J. and Fincher, G.B. (2007) A barley xyloglucan xyloglucosyl transferase covalently links xyloglucan, cellulosic substrates, and (1,3;1,4)-β-D-glucans. *J. Biol. Chem.* 282: 12951-12962.

Husnik, F. and McCutcheon, J.P. (2018) Functional horizontal gene transfer from bacteria to eukaryotes. *Nat. Rev. Microbiol.* 16: 67.

Johansson, P., Brumer, H., Baumann, M.J., Kallas, Å.M., Henriksson, H., Denman, S.E., et al. (2004) Crystal structures of a poplar xyloglucan endotransglycosylase reveal details of transglycosylation acceptor binding. *Plant Cell* 16: 874-886.

Kaewthai, N., Gendre, D., Eklöf, J.M., Ibatullin, F.M., Ezcurra, L., Bhalerao, R.P., et al. (2013) Group III-A XTH genes of *Arabidopsis* encode predominant xyloglucan endohydrolases that are dispensable for normal growth. *Plant Physiol.* 161: 440-454.

Kvålseth, T.O. (1987) Entropy and correlation: some comments. *IEEE Trans. Syst. Man Cybern.* 17: 517-519.

Lampugnani, E.R., Flores-Sandoval, E., Tan, Q.W., Mutwil, M., Bowman, J.L. and Persson, S. (2019) Cellulose synthesis – central components and their evolutionary relationships. *Trends Plant Sci.* 24: 402-412.

Lang, A.S., Zhaxybayeva, O. and Beatty, J.T. (2012) Gene transfer agents: phage-like elements of genetic exchange. *Nat. Rev. Microbiol.* 10: 472-482.

Langdale, J.A. and Harrison, C.J. (2008) Developmental transitions during the evolution of plant form. In *Evolving Pathways: Key Themes in Evolutionary Developmental Biology*. Edited by Minelli, A. and Fusco, G. pp. 299-316. Cambridge University Press, Cambridge, UK.

Le, P.T., Pontarotti, P. and Raoult, D. (2014) Alphaproteobacteria species as a source and target of lateral sequence transfers. *Trends Microbiol.* 22: 147-156.

Leliaert, F., Smith, D.R., Moreau, H., Herron, M.D., Verbruggen, H., Delwiche, C.F., et al. (2012) Phylogeny and molecular evolution of the green algae. *Crit. Rev. Plant Sci.* 31: 1-46.
Mark, P., Baumann, M.J., Eklöf, J.M., Gullfot, F., Michel, G., Kallas, Å.M., et al. (2009) Analysis of nasturtium TmNXG1 complexes by crystallography and molecular dynamics provides detailed insight into substrate recognition by family GH16 xyloglucan endo-transglycosylases and endo-hydrolases. *Proteins* 75: 820-836.

McDaniel, L.D., Young, E., Delaney, J., Ruhnau, F., Ritchie, K.B. and Paul, J.H. (2010) High frequency of horizontal gene transfer in the oceans. *Science* 330: 50-50.

McGregor, N., Yin, V., Tung, C.C., Van Petegem, F. and Brumer, H. (2017) Crystallographic insight into the evolutionary origins of xyloglucan endotransglycosylases and endohydrolases. *Plant J.* 89: 651-670.

Mendis, H.C., Queiroux, C., Brewer, T.E., Davis, O.M., Washburn, B.K. and Jones, K.M. (2013) The succinoglycan endoglycanase encoded by exoK is required for efficient symbiosis of *Sinorhizobium meliloti* 1021 with the host plants *Medicago truncatula* and *Medicago sativa* (Alfalfa). *Mol. Plant Microbe Interact.* 26: 1089-1105.

Michel, G., Chantalat, L., Duee, E., Barbayron, T., Henrissat, B., Kloareg, B., et al. (2001) The κ-carrageenase of *P. carrageenovora* features a tunnel-shaped active site: a novel insight in the evolution of clan-B glycoside hydrolases. *Structure* 9: 513-525.

Miedes, E. and Lorences, E.P. (2007) The implication of xyloglucan endotransglucosylase/hydrolase (XTHs) in tomato fruit infection by *Penicillium expansum* link. *A. J. Agric. Food Chem.* 55: 9021-9026.

Mikhailyuk, T.I., Sluiman, H.J., Massalski, A., Mudimu, O., Demchenko, E.M., Kondratyuk, S.Y., et al. (2008) New streptophyte green algae from terrestrial habitats and an assessment of the genus *Interfilum* (Klebsormidiophyceae, Streptophyta). *J. Phycol.* 44: 1586-1603.

Moore, E.K., Jelen, B.I., Giovannelli, D., Raanan, H. and Falkowski, P.G. (2017) Metal availability and the expanding network of microbial metabolisms in the Archaean eon. *Nat. Geosci.* 10: 629-636.

Murphy, C.L., Youssef, N.H., Hanafy, R.A., Couger, M.B., Stajich, J.E., Wang, Y., et al. (2019) Horizontal gene transfer as an indispensable driver for evolution of Neocallimastigomycota into a distinct gut-dwelling fungal lineage. *Appl. Environ. Microbiol.* 85: e00988-00919.

Naranjo-Ortiz, M.A. and Gabaldón, T. (2019) Fungal evolution: major ecological adaptations and evolutionary transitions. *Biol. Rev.* 94: 1443-1476.

Nishitani, K. (1997) The role of endoxyloglucan transferase in the organization of plant cell walls. In *International Review of Cytology*. Edited by Jeon, K.W. pp. 157-206. Academic Press.

Nishitani, K. and Tominaga, R. (1991) In vitro molecular weight increase in xyloglucans by an apoplastic enzyme preparation from epicotyls of *Vigna angularis*. *Physiol. Plant.* 82: 490-497.
Nishitani, K. and Tominaga, R. (1992) Endo-xyloglucan transferase, a novel class of glycosyltransferase that catalyzes transfer of a segment of xyloglucan molecule to another xyloglucan molecule. *J. Biol. Chem.* 267: 21058-21064.

Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K.K., et al. (2018) The *Chara* genome: secondary complexity and implications for plant terrestrialization. *Cell* 174: 448-464. e424.

Okazawa, K., Sato, Y., Nakagawa, T., Asada, K., Kato, I., Tomita, E., et al. (1993) Molecular cloning and cDNA sequencing of endoxyloglucan transferase, a novel class of glycosyltransferase that mediates molecular grafting between matrix polysaccharides in plant cell walls. *J. Biol. Chem.* 268: 25364-25368.

Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8: 785-786.

Petersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, F.C., et al. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25: 1605-1612.

Planas, A. (2000) Bacterial 1,3-1,4-β-glucanases: structure, function and protein engineering. *Biochim. Biophys. Acta* 1543: 361-382.

Purugganan, M.M., Braam, J. and Fry, S.C. (1997) The Arabidopsis TCH4 xyloglucan endotransglycosylase (substrate specificity, pH optimum, and cold tolerance). *Plant Physiol.* 115: 181.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., et al. (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319: 64-69.

Rodríguez-Peña, J.M., Cid, V.J., Arroyo, J. and Nombela, C. (2000) A novel family of cell wall-related proteins regulated differently during the yeast life cycle. *Mol. Cell. Biol.* 20: 3245-3255.

Rose, J.K.C., Braam, J., Fry, S.C. and Nishitani, K. (2002) The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature. *Plant Cell Physiol.* 43: 1421-1435.

Ryšánek, D., Elster, J., Kováčik, L. and Škaloud, P. (2016) Diversity and dispersal capacities of a terrestrial algal genus *Klebsormidium* (Streptophyta) in polar regions. *FEMS Microbiol. Ecol.* 92.

Schoenheit, P., Buckel, W. and Martin, W.F. (2016) On the origin of heterotrophy. *Trends Microbiol.* 24: 12-25.

Shinohara, N., Sunagawa, N., Tamura, S., Yokoyama, R., Ueda, M., Igarashi, K., et al. (2017) The plant cell-wall enzyme AtXTH3 catalyses covalent cross-linking between cellulose and cello-oligosaccharide. *Sci. Rep.* 7: 46099.

Simmons, T.J., Mohler, K.E., Holland, C., Goubet, F., Franková, L., Houston, D.R., et al. (2015) Hetero-trans-β-glucanase, an enzyme unique to *Equisetum* plants,functionalizes cellulose. *Plant J.* 83: 753-769.
Smith, R.C. and Fry, S.C. (1991) Endotransglycosylation of xyloglucans in plant cell suspension cultures. *Biochem. J.* 279: 529-535.

Soskine, M. and Tawfik, D.S. (2010) Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* 11: 572-582.

Staehelin, C., Forsberg, L.S., D’Haeze, W., Gao, M.-Y., Carlson, R.W., Xie, Z.-P., et al. (2006) Exo-oligosaccharides of *Rhizobium* sp. strain NGR234 are required for symbiosis with various legumes. *J. Bacteriol.* 188: 6168-6178.

Stratilová, B., Firáková, Z., Klaudiny, J., Šesták, S., Kozmon, S., Strouhalová, D., et al. (2019) Engineering the acceptor substrate specificity in the xyloglucan endotransglycosylase TmXET6.3 from nasturtium seeds (*Tropaeolum majus* L.). *Plant Mol. Biol.* 100: 181-197.

Stratilová, B., Šesták, S., Mravec, J., Garajová, S., Pakanová, Z., Vadinová, K., et al. (2020) Another building block in the plant cell wall: Barley xyloglucan xyloglucosyl transferases link covalently xyloglucan and anionic oligosaccharides derived from pectin. *Plant J.* 104: 752-767.

Stülke, J. and Hillen, W. (2000) Regulation of carbon catabolism in *Bacillus* species. *Annu. Rev. Microbiol.* 54: 849-880.

Tabuchi, A., Mori, H., Kamisaka, S. and Hoson, T. (2001) A new type of endo-xyloglucan transferase devoted to xyloglucan hydrolysis in the cell wall of azuki bean epicotyls. *Plant Cell Physiol.* 42: 154-161.

Tedersoo, L., Sánchez-Ramírez, S., Kõljalg, U., Bahram, M., Döring, M., Schigel, D., et al. (2018) High-level classification of the Fungi and a tool for evolutionary ecological analyses. *Fungal Divers.* 90: 135-159.

Timme, R.E., Bachvaroff, T.R. and Delwiche, C.F. (2012) Broad phylogenomic sampling and the sister lineage of land plants. *PLoS One* 7: e29696.

Uchiyama, T., Uchihashi, T., Nakamura, A., Watanabe, H., Kaneko, S., Samejima, M., et al. (2020) Convergent evolution of processivity in bacterial and fungal cellulases. *Proc. Natl. Acad. Sci., USA* 117: 19896.

Viborg, A.H., Terrapon, N., Lombard, V., Michel, G., Cejzek, M., Henriassat, B., et al. (2019) A subfamily roadmap of the evolutionarily diverse glycoside hydrolase family 16 (GH16). *J. Biol. Chem.* 294: 15973-15986.

Wang, S., Li, Li, Li, H., Sahu, S.K., Wang, H., Xu, Y., et al. (2020) Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nat. Plants* 6: 95-106.

Wang, Y., Youssef, N.H., Couger, M.B., Hanafy, R.A., Elshahed, M.S. and Stajich, J.E. (2019) Molecular dating of the emergence of anaerobic rumen fungi and the impact of laterally acquired genes. *mSystems* 4: e00247-00219.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46: W296-W303.
Weckwerth, W. and Selbig, J. (2003) Scoring and identifying organism-specific functional patterns and putative phosphorylation sites in protein sequences using mutual information. *Biochem. Biophys. Res. Commun.* 307: 516-521.

Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., et al. (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci., USA* 111: E4859-4868.

Yokoyama, R. and Nishitani, K. (2001) A comprehensive expression analysis of all members of a gene family encoding cell-wall enzymes allowed us to predict cis-regulatory regions involved in cell-wall construction in specific organs of *Arabidopsis*. *Plant Cell Physiol.* 42: 1025-1033.

Yokoyama, R., Rose, J.K.C. and Nishitani, K. (2004) A surprising diversity and abundance of xyloglucan endotransglucosylase/hydrolases in rice. Classification and expression analysis. *Plant Physiol.* 134: 1088-1099.

Yokoyama, R., Uwagaki, Y., Sasaki, H., Harada, T., Hiwatashi, Y., Hasebe, M., et al. (2010) Biological implications of the occurrence of 32 members of the XTH (xyloglucan endotransglucosylase/hydrolase) family of proteins in the bryophyte *Physcomitrella patens*. *Plant J.* 64: 645-656.

Yu, G., Smith, D.K., Zhu, H., Guan, Y. and Lam, T.T.Y. (2017) GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8: 28-36.

Zhang, J., Fu, X.-X., Li, R.-Q., Zhao, X., Liu, Y., Li, M.-H., et al. (2020) The hornwort genome and early land plant evolution. *Nat. Plants* 6: 107-118.
Figure Legends:

**Fig. 1** Comparison of XTH and other related enzymes. Three-dimensional structures are derived from experimental (PDB code) or predicted (gene ID of modeled protein and PDB code of template protein) data. Substructures absent in the jellyroll β-sandwich of EG16 are colored in purple. A hypothetical evolutionary route is shown at the top, where gain (+) and loss (−) events ([1]-[5], explained at the left) were parsimoniously mapped, and the dotted line represents horizontal gene transfer. In the table, ticks denote presence, crosses denote absence, and ellipses denote unknown status.

**Fig. 2** Evolutionary history of the XTH family. Green and gray lines represent the evolutionary histories of Streptophyta and bacteria, respectively. Black lines represent the molecular evolution of XTH and related enzymes, and the black dotted line indicates horizontal gene transfer (HGT). A wide light blue bar represents the Cryogenian period (720–635 million years ago). Evolutionary gain events related to plant cell-wall evolution (details in Discussion) are shown in purple, where XyG indicates xyloglucan.

**Fig. 3** Phylogenetic inference of XTH diversification in the land plant lineage. A maximum-likelihood phylogram of XTH protein sequences in genome datasets of *Mesotaenium endlicherianum* (charophycean alga), *Spirogoea muscicola* (charophycean alga), *Marchantia polymorpha* (liverwort), *Selaginella moellendorffii* (lycophyte), *Oryza sativa* (monocot), and *Arabidopsis thaliana* (eudicot) and a transcriptome dataset of *Equisetum diffusum* (horsetail; fern ally) is shown. Individual XTHs of *EfHTG* (*E. fluviatile*; horsetail), *PttXET16-34* (*Populus tremula × P. tremuloides*; poplar), and *TmXET6.3* (*Tropaeolum majus*; nasturtium) are also included. Pairs of uppercase letters around the phylogram denote amino acid residues corresponding to Q102 and R116 of the mature *PttXET16-34* protein. Support values of major nodes are shown in parentheses. A list of source organisms and their mutual relationships is shown on the left, where the numbers of XTH sequences in genome [G] and transcriptome [T] datasets are given in brackets.

**Fig. 4** Curation of XTH-homologs in the *Klebsormidium nitens* genome. (A) A maximum-likelihood phylogram of XTH-homologs in a genome dataset of *K. nitens* (formerly *K. flaccidum*) is shown. A genome dataset of *Spirogoea muscicola* (Zygmatophyceae), previously reported EG16-2 s of
Charophyta, EG16s of Coleochaetales, moss, and eudicots, and XTHs of eudicots, are included as reference sequences. The branch position of Kfl00047_0180, which was determined by a separate phylogenetic analysis of its fragmentary sequence, is indicated with an arrow. Support values of major nodes are shown on branches. Triplets of letters on the right denote amino acid residues corresponding to the 44th–46th residues of the mature PttXET16-34 protein, and signature residues of EG16s are emphasized. (B) Schematic representation of *K. nitens* XTH-homolog proteins, with gaps, predicted signal peptides, active sites (EXDXE), and regions similar to the EG16 protein of VvEG16 (XP_002273975.1) shown. (C) Number of genes in each genome. Parentheses represent uncertain status. In (A) and (C), lists of source organisms and their mutual relationships are shown on the left.

**Fig. 5** Phylogenetic inference of the non-plant origin of XTH-related proteins. (A) A maximum-likelihood phylogram of representative members of GH16_18 (CRH1 of *Saccharomyces cerevisiae*), GH16_19 (CRH2 and CRR1 of *Saccharomyces cerevisiae*), GH16_20 (XTHs of *Mesotaenium endlicherianum*, Zyggnematophyceae), EG16-2 (GH16s of *Klebsormidium nitens*), and GH16_21 (QHA14546.1: an ExoK-like protein of *Rhodobacter sphaeroides*, a facultative photosynthetic bacterium; AAA16048.1: an ExoK of *Ensifer meliloti*, a nitrogen-fixing root-nodule bacterium) is shown. Support values are shown on branches. The output tree (A) supports the topology (B) and horizontal gene transfer (HGT) from bacteria via plants to fungi (D) and discounts the topology (C) and HGT from bacteria via fungi to plants (E). Topology (B) is also explained by two separate bacteria-to-fungi and bacteria-to-plant HGT events (F).

**Fig. 6** Phylogenetic inference of the bacterial origin of EG16-2s. A maximum-likelihood phylogram of protein sequences in GH16_21 and of Charophyta EG16-2s is shown. Arrows indicate biochemically characterized enzymes. Support values of major nodes are shown on branches.

**Fig. 7** Comparison between Firmicutes licheninases and Alphaproteobacteria ExoKs. (A) Mutual information profile estimating amino acid covariation in licheninase- and ExoK-clades of GH16_21. In the profile, peaks above 0.7 (> the 3-σ detection limit) are annotated, and the active site is shaded in light blue. (B) Frequency logos at the positions of the annotated peaks. (C) Three-dimensional structures of licheninase (3o5s) and ExoK (AAA16048 modeled using the 3o5s template).
Fig. 8 Comparison of XTH family groups. (A) Mutual information profiles estimating amino acid covariation in groups of the *Arabidopsis thaliana* XTH family. In the profiles, peaks above 0.8 (> the 3-σ detection limit) are annotated, and the active site is shaded in light blue. In the annotation, digits after + denote Nth gap sites from preceding residues in PttXET16-34. (B) Frequency logos at the positions of the annotated peaks. The empty symbol represents a gap in aligned amino acid sequences. Signature residues of group II are colored in purple, and the corresponding residues of the other groups are colored in green. (C) Q102 and R116 residues in the three-dimensional structure of the enzyme-substrate complex (1umz) of PttXET16-34 and xyloglucan nonasaccharide, where E89 is the catalytic nucleophile.
Figure 1

| Taxonomic distribution | Age estimates (million years ago) | Substrates | Heterotrophic digestion | N-terminal loop extension | Acceptor-side extension | Signal peptide | Donor-side extension | Transglycosylation | Substrate promiscuity |
|------------------------|----------------------------------|------------|-------------------------|--------------------------|------------------------|---------------|---------------------|-------------------|----------------------|
| Licheninase (3o5s)    | 3,000                            | β(1,3;1,4)-glucan succinoglycan ... | ✓                        | ✓                        | X                      | ✓             | X                   | ✓                 | ✓                    |
| ExoK (AAA18348-3o5s)   | 3,000                            | xyloglucan, β(1,3;1,4)-glucan   | ✓                        | ✓                        | ✓                      | ✓             | ✓                   | ✓                 | ✓                    |
| EG16-2 (KFL_00110255.56d) | 700                              | xyloglucan                  | ✓                        | ✓                        | ✓                      | ✓             | ✓                   | ✓                 | ✓                    |
| EG16 (5o8S)           | 600                              | xyloglucan                  | ✓                        | ✓                        | ✓                      | ✓             | ✓                   | ✓                 | ✓                    |
| XTH (Tumew)           | 600                              | chitin                     | ✓                        | ✓                        | ✓                      | ✓             | ✓                   | ✓                 | ✓                    |
| CRH (68w)             | 700                              | terrestrial fungi           | ✓                        | ✓                        | ✓                      | ✓             | ✓                   | ✓                 | ✓                    |
Figure 2
Figure 3

Reference proteins
- PixET16-34
- TrnXET6.3
- ERHTG

Land plants (Embryophyta)
- Arabidopsis thaliana [G:32]
- Glycine max [G:30]
- Equisetum arvense [T:12]
- Selaginella moellendorffii [G:27]
- Marchantia polymorpha [G:43]

Algae (Charophyta/Zygnematales)
- Mesotaenium endlichianum [G:3]
- Spirogyra muscicola [G:8]

Group I (0.71)
Group II (0.95)
Group III-B (0.81)
Group III-A (1.00)
Group IV (Ancestral) (0.94)

Algal (0.96)
Figure 4

A

Land plants (Embryophyta)
- Arabidopsis thaliana
- Vitis vinifera
- Physcomitrella patens

Algae (Charophyta)
- Mesotaenium endlicherianum
- Spiroglaoea muscicola
- Chaetosphaeridium globosum
- Klebsormidium subtile
- Klebsormidium nitens

B

Klebsormidium nitens genome annotation v1.1 (predicted GH16 proteins)

C

| Species                        | References      | EG16-2 | EG16 | XTH | Total |
|-------------------------------|-----------------|--------|------|-----|-------|
| Arabidopsis thaliana          | Rose et al., 2002 | 0      | 0    | 32  | 32    |
| Mesotaenium endlicherianum    | Cheng et al., 2019 | 3      | 0    | 3   | 6     |
| Spiroglaoea muscicola         | Cheng et al., 2019 | 5      | 0    | 8   | 13    |
| Coleochaete scutata a)        | Behar et al., 2018 | (5)    | (1)  | (0) | unknown |
| Chaetosphaeridium globosum a) | Behar et al., 2018 | (2)    | (1)  | (0) | unknown |
| Klebsormidium nitens          | Hori et al., 2014 | 6      | 0    | 0   | 6     |
| Chlorokybus amphitrichus       | Wang et al., 2020 | 0      | 0    | 0   | 0     |
| Mesostigma viride             | Wang et al., 2020 | 0      | 0    | 0   | 0     |

a) transcriptome datasets
Figure 5

A Maximum-likelihood phylogenetic tree of GH16 proteins

B Supported topology

C Unsupported topology

D Likely scenario

E Unlikely scenario

F Alternative scenario
Figure 7

A. Normalized Mutual Information

B. Frequency

C. Licheninase - ExoK

(Position in mature PtiXET16-34)

10 Å
