Phylogenomics reveals the evolutionary origins of lichenization in chlorophyte algae

Camille Puginier[1], Cyril Libourel[1], Juergen Otte[2], Pavel Skaloud[3], Mireille Haon[4,5], Sacha Grisel[4,5], Malte Petersen[6], Jean-Guy Berrin[4,5], Pierre-Marc Delaux[1]*, Francesco Dal Grande[2,7,8]*, Jean Keller[1,9*]

[1] Laboratoire de Recherche en Sciences Végétales (LRSV), Université de Toulouse, CNRS, UPS, INP Toulouse, 31320 Castanet-Tolosan, France.

[2] Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany

[3] Department of Botany, Faculty of Science, Charles University, Benátská 2, CZ-12800 Praha 2, Czech Republic

[4] INRAE, Aix Marseille Université, UMR1163 Biodiversité et Biotechnologie Fongiques (BBF), 13009 Marseille, France

[5] INRAE, Aix Marseille Université, 3PE Platform, 13009 Marseille, France

[6] High Performance Computing & Analytics Lab, University of Bonn, Friedrich-Hirzebruch-Allee 8, 53115 Bonn, Germany

[7] LOEWE Centre for Translational Biodiversity Genomics (TBG), Senckenberganlage 25, 60325 Frankfurt am Main, Germany

[8] Department of Biology, University of Padova, Padua, Italy

[9] Department of Insect Symbiosis, Max Planck Institute for Chemical Ecology, 07745 Jena, Germany

*To whom correspondence may be addressed: pierre-marc.delaux@cnrs.fr; francesco.dalgrande@unipd.it; jean.keller@cnrs.fr

Abstract

Mutualistic symbioses have contributed to major transitions in the evolution of life. Here, we investigate the evolutionary history and the molecular innovations at the origin of lichens, which are a symbiosis established between fungi and green algae or cyanobacteria. We *de novo* sequence the genomes or transcriptomes of 12 lichen algal symbiont (LAS) and closely related non-symbiotic algae (NSA) to improve the genomic coverage of Chlorophyte algae. We then perform ancestral state reconstruction and comparative phylogenomics. We identify at least three independent gains of the ability to engage in the lichen symbiosis, one in Trebouxiophyceae and two in Ulvophyceae, confirming the convergent evolution of the lichen symbioses. A carbohydrate-active enzyme from the glycoside hydrolase 8 (GH8) family was identified as a top candidate for the molecular-mechanism underlying lichen symbiosis in Trebouxiophyceae. This GH8 was acquired in lichenizing Trebouxiophyceae by horizontal gene transfer, concomitantly with the ability to associate with lichens fungal symbionts (LFS) and is able to degrade polysaccharides found in the cell wall of LFS. These findings indicate that a combination of gene family expansion and horizontal gene transfer provided the basis for lichenization to evolve in chlorophyte algae.

Introduction

Mutualistic interactions between plants and microorganisms are the foundation of plant diversification and adaptation to almost all terrestrial ecosystems[1,2]. An emblematic example of mutualism impact on Earth is the transition of plants from the aquatic environment to land, which occurred 450 million years ago and was partly enabled by the arbuscular mycorrhizal symbiosis formed with Glomeromycota fungi [2,3]. Another emblematic example of a plant-fungi symbiosis occurs in the mutualistic association between certain chlorophyte algae and fungi resulting in the formation of lichens [4,5].

Lichens are symbiotic structures composed of several types of organisms including a fungal partner, that most commonly belongs to the Ascomycetes and more rarely to the Basidiomycetes, and a photosynthetic partner, also called photobiont. Photobionts can be either cyanobacteria or algae belonging to the Chlorophytes. Certain lichens can contain both types of photobionts [6]. In this mutualistic symbiosis, both mycobionts and photobionts obtain benefits from their association. Carbohydrates from photosynthesis are supplied to the fungal partners, whereas the fungi create a favorable microenvironment shielding the photobionts from biotic and abiotic stresses [5,7]. Recently, metagenomic studies have demonstrated that other types of microorganisms, such as lichenicolous fungi and bacteria, are found within the lichen thallus and are likely important for its biology [8–10].

Because of their ecological and physiological importance, how these mutualistic interactions originated has been a central question for decades [11]. The comparison of genomes in a defined phylogenetic context (comparative phylogenomics) has successfully unraveled the evolutionary history of several mutualistic symbioses with complex evolutionary patterns, combining gains and losses across lineages [12–14]. In addition, such approaches have the potential to identify the molecular mechanisms associated with major innovations, including symbioses [15–17]. Even though lichens have been considered as a long-lasting mutualistic interaction between lichen fungal symbionts (LFS) and one or more photobionts, lichens have been asymmetrically investigated from the fungal perspective leading to the conclusion that the ability to form lichens has been originally acquired, lost, and regained multiple times during the evolution of the ascomycetes and basidiomycetes [18–20].

On the photobiont side, algal species that are known to establish the lichen symbioses (thereafter called lichen algal symbionts or LAS) are almost exclusively found in two of the eleven chlorophyte algae classes, the Ulvophyceae and the Trebouxiophyceae [5,21]. Such distribution of the ability to associate with LFS might be the result of either a single gain in the common ancestor of Ulvophyceae and Trebouxiophyceae followed by multiple losses, in a similar manner to other terrestrial endosymbioses [12,13,22,2], or multiple independent gains. Studies based on time-calibrated phylogenetic approaches provided strong support for the

convergent evolution of LFS and suggested a similar pattern for LAS [20,23]. However, the limited availability of LAS genomes has so far constrained molecular analyses to single algal species such as *Asterochloris glomerata* and *Trebouxia sp.* TZW2008 [24,25]. Thus, the evolutionary history of lichens on the green algal side and the underlying molecular mechanisms associated with lichenization (algae that are hosted and have a lifestyle inside of lichens symbioses) remain elusive. The initiation of contact between lichen symbionts hinges on mutual recognition, with emerging evidence suggesting the involvement of elicitors that interact with the cell wall (reviewed in [26]). On the mycobiont side, fungal stimuli may encompass the activities of carbohydrate-active enzymes (CAZymes), potentially enhancing the permeability of algal cell walls [27]. Sugars, sugar alcohols, along with other compound groups like secondary metabolites and antioxidants, are proposed as key elements in maintaining the intricately balanced symbiotic interplay between fungi and algae in lichens [26,28]. This process implies that LAS should manifest distinct genomic features compared to algae unable to establish symbiotic associations [29–32]. In this case as well, the limited availability of genomic information for LAS has thus far impeded the testing of this hypothesis.

In this study, we deployed unsupervised phylogenomic comparative approaches to decipher the evolutionary history and the genetic mechanisms conferring certain chlorophyte species the ability to engage in the lichen symbiosis. In this study, we *de novo* sequenced and annotated six LAS genomes, two LAS transcriptomes, three non-symbiotic algae (NSA) genomes, and one NSA transcriptome. We performed ancestral state reconstruction using this dataset along with 26 genomes and 103 transcriptomes of chlorophyte algae publicly available, demonstrating at least three convergent gains of lichenization in chlorophyte algae. We scrutinized one of these events through comparative phylogenomics cross-referenced with differential gene expression data and identified lichenization-related molecular mechanisms. We propose an evolutionary model for the evolution of lichens based on the projection of these molecular characteristics onto the phylogeny of chlorophyte algae. This scenario involves the expansion of gene families and horizontal gene transfers that likely facilitated the interaction between the symbiotic partners.

Results

Expanding the genomic coverage of the chlorophyte algae

To date, genomes and transcriptomes are available for only seven LAS. To investigate the evolution of lichens, we produced six new long-reads-based genome assemblies for LAS

species belonging to the Trebouxiales, Botryococcus and Apatococcus clades (Table 1, Fig. 1, Supplementary Data 1). We also sequenced three closely related NSA including species from the Apatoccocus and Myrmecia genera for which no genomes were previously available (Table 1, Fig. 1, Supplementary Data 1). Assemblies for eight of the nine species displayed an average scaffold length N50 of almost 2Mb, and an average of only 143 scaffolds (Table 1, Supplementary Data 2). The ninth assembly (*Apatococcus fuscideae* SAG2523) displayed a scaffold N50 of 50 kb and a much higher number of 2,319 scaffolds (Table 1, Supplementary Data 2). However, the genome completeness, estimated by BUSCO, indicated that most of the actual proteome was captured (75.3%, Table 1). To complete this dataset, the transcriptome of three additional species, two Trebouxiophyceae, and the symbiotic Ulvophyceae *Paulbroadya petersii*, were sequenced on an Illumina NovaSeq platform yielding an average of 40 million reads (details of sequencing and assembly statistics in Supplementary Data 3). The assembled transcriptomes reached 77.4% completeness when assessed by BUSCO.

The nine newly assembled genomes and three transcriptomes were combined with publicly available data mined from diverse databases, producing a final database of 141 species composed of 35 genomes (Supplementary Data 1) and 106 transcriptomes (Supplementary Data 1) of LAS and NSA species, covering all the chlorophyte classes.

Ancestral state reconstruction supports at least three independent gains of the ability to associate with lichen fungal symbionts in Chlorophytes

Textbooks and recent studies [20,23] all converge to a single hypothesis for the evolution of lichenization: it evolved in a convergent manner in Chlorophytes. Such evolutionary scenario had been proposed in the past for other type of symbioses, such as the nitrogen-fixing root nodule symbiosis [33], and later rejected [13,14,34]. The alternative hypothesis for the evolution of lichenization, a single gain followed by multiple losses, has so far not been explored. To determine the evolutionary history of lichenization in Chlorophytes, either rejecting the consensus convergent gains hypothesis or further supporting it, we conducted an Ancestral State Reconstruction (ASR) approach. The predicted proteomes from the 141 species were used as a matrix to reconstruct orthogroups using OrthoFinder, yielding a total of 2 157 361 genes (74.6% of the total) assigned to 197 669 hierarchical orthogroups (Supplementary Data 4). Based on the most informative orthogroups, a species tree of the 141 chlorophyte species was computed, rooted on *Prasinoderma coloniale* [35]. Either of two states for the lichenization capacity (LAS or NSA) were assigned to each chlorophyte algae present in the sampling. The status of algal species as symbionts was assigned following [21]. Furthermore, given that many algal species lack distinct morphological features and their lichenization status has often only

been reported in light microscopic studies, the determination of the lichenization status for each species in this study was based on a thorough review of published studies based on sequence data[36–41]. The ASR inferred three gains of lichenization within the Chlorophytes, one in Trebouxiophyceae and two in Ulvophyceae (Fig. 1). In Trebouxiophyceae, the single gain of the lichenization ability was followed by eleven putative losses in *Myrmecia bisecta* (SAG2043), *Apatococcus lobatus* (SAG2145), *Elliptochloris marina*, *Coccomyxa subellipsoidea* (C-169), *Botryococcus braunii*, *Botryococcus terriblis*, *Symbiochloris irregularis* (SAG2036), *Watanabea reniformis*, *Microthamnion kuetzingianum*, *Nannochloris atomus*, *Eremosphaera viridis* and *Geminella sp.* The more limited sampling in Ulvophyceae does not allow identifying loss events and, as well, may have masked additional gains. Based on the ASR, it can be proposed that the ability to engage in the symbiosis was acquired at least three times independently in Chlorophytes, aligning with the current consensus hypothesis for the evolution of this trait [23]. Although at the macroscopic level, the trait (*i.e.,* a chlorophyte alga hosted inside a fungal thallus) can be considered identical across the different clades, lichen symbiosis should be considered as a group of diverse symbioses rather than a single interaction.

Trebouxiophyceae symbionts share conserved hierarchical orthogroups (HOG)

Functional innovations are associated with the gains of genomic or genetic features which can be tracked by comparative genomics. Our gathered dataset encompasses 13 LAS and 23 NSA in the Trebouxiophyceae class allowing to conduct such a comparative analysis to identify genes and gene families associated with the ability to engage into lichens. When comparing general genomic features such as protein coding gene number, GC and transposable elements contents, and genome size no differences were observed between symbiotic and non-symbiotic Trebouxiophyceae (Supplementary Fig. 1, Supplementary Fig. 2, Supplementary Data 5) apart from the GC content which is lower in symbiotic Trebouxiophyceae. Hence, the gain of the ability to lichenize did not involve massive genomic modifications as it is the case for other symbioses [29–32].

To identify genes potentially associated with lichenization in Trebouxiophyceae, we focused on the computed orthogroups for the entire Chlorophytes using two complementary statistical approaches. First, we conducted a sparse Partial Least Square Discriminant Analysis (sPLS-DA) in which each orthogroup composition is analyzed to identify orthogroups whose composition clusters the 141 species into two groups: the symbiotic Trebouxiophyceae and the other Chlorophytes. The first two principal components are responsible for 2 and 1% of the discrimination of the species into the two groups respectively (Fig. 2a). Since the first component is the most discriminant one, we focused on the 100 orthogroups that contribute the most to it (Fig. 2b, Supplementary Fig. 3). In a complementary approach, we applied a

Mann-Whitney-Wilcoxon test to identify orthogroups that are significantly enriched in sequences from symbiotic Trebouxiophyceae. This approach identified 5 252 orthogroups (pvalue < 0.01, Fig. 2). When cross-referencing the sPLS-DA data with the Mann-Whitney-Wilcoxon test, we found a perfect overlap. Indeed, the 100 top orthogroups from the sPLS-DA were among the 5 252 orthogroups identified by the Mann-Whitney-Wilcoxon test (Fig. 2c, Supplementary Data 6). These 100 top orthogroups thus represent genes potentially associated with the evolution of lichenization in Trebouxiophyceae (Supplementary Data 7).

Gene family expansions are associated with the evolution of lichenization in Trebouxiophyceae

To narrow down the most promising candidates associated with the origin of lichenization in Trebouxiophyceae and test their symbiotic relevance, the 100 candidate orthogroups were further analyzed. First, because genes involved in symbiotic association often show differential regulation in the presence of the other symbiont [2,42], we determined whether the expression level of the candidates was affected during lichenization. For this, we collected RNAseq data previously obtained for *Trebouxia sp.* TZW2008 grown in the absence of symbiotic fungus, in co-culture with the LFS *Usnea hakonensis,* or in well-established lichens [25] and recomputed differentially expressed genes. This analysis revealed a total of 3 540 differentially regulated genes, either up- or down-regulated, when *Trebouxia sp.* TZW2008 associates with *Usnea hakonensis* (Supplementary Data 8)[25]. The differentially expressed genes were cross-referenced with the 100 orthogroups associated with lichenization in Trebouxiophyceae. Comparing the two datasets, we identified 42 orthogroups showing at least one *Trebouxia sp.* TZW2008 gene differentially regulated (14 HOG with up-regulated genes, seven with both up and down-regulated genes, and 21 with only down-regulated genes) in association with *Usnea hakonensis*.

Although OrthoFinder and other orthogroup-generating tools represent the only options to study genome-wide phylogenomic patterns, the resolution of the orthogroups is dependent on the sampled species and the gene-family complexity. In other words, orthogroups might either exclude actual orthologs or include non-orthologous genes. To reconstruct the evolutionary history of these candidate genes with higher confidence, we subjected them to phylogenetic analysis. Using targeted phylogenetic inference, five of the candidate genes were not found associated with lichenization anymore and five of the phylogenies were not resolved enough to conclude on the evolutionary history of the genes (Supplementary Data 9).

Thus, from the 42 candidate orthogroups, a total of 32 showed phylogenetic and differential gene expression (in *Trebouxia sp.* TZW2008) patterns associated with the symbiotic habit. Reverse genetic analyses will be required in the future to validate their functions when a

genetically tractable system and the *in* vitro resynthesis or lichen formation have been developed. Among the candidate genes associated with the symbiosis, eight contain genes that are annotated with IPR domains and can be associated with a putative function (Fig. 3, Supplementary Data 7). The 32 candidate orthogroups exhibit distinctive phylogenetic distributions, including indications of gene family expansions, exemplified by N0.HOG0002085, which encompasses genes annotated as glucose/ribitol dehydrogenase and short-chain dehydrogenase/reductase (SDR) (Fig. 3). Additionally, some candidates are LAS-specific, as seen in N0.HOG0012965, which contains a carbohydrate-active enzyme belonging to the glycoside hydrolase 8 family, or in N0.HOG0012501 that contains glutathione S-transferase enzymes (Supplementary Data 7). Furthermore, certain candidates such as N0.HOG0025580 and N0.HOG0025596 (both with an unknown function) display a specific distribution among Trebouxiales only (Fig. 3, Supplementary Data 9). Altogether, the phylogenomic comparison reveals that diverse genomic processes, including gene family expansions, contributed to the evolution of lichenization in the Trebouxiophyceae.

Horizontal gene transfers contributed to the evolution of Trebouxiophycean lichens

Besides genes family expansion, two genes seemed to be highly specific to the symbiotic Trebouxiophyceae and almost completely absent from non-symbiotic Trebouxiophyceae and other Chlorophytes (Supplementary Data 9, Supplementary Fig. 4). Such evolutionary pattern can be the result of *de novo* gene birth or horizontal gene transfer (HGT). HGTs have been found previously as drivers for the acquisition of functional innovations across living organisms, including plants[43–45]. To determine the origin of these two symbiosis-associated genes, further phylogenetic analysis of the symbiotic Trebouxiophyceae-specific orthogroups was conducted, using additional databases including the main eukaryotic and prokaryotic lineages, to search for putative homologs across the tree of life. An origin by HGT was clearly identified for the two candidates. The first one, the orthogroup N0.HOG0012965, corresponds to an enzyme from the GH8 family. Based on the CAZy classification, GH8 enzymes have only been found in bacteria.[46]. This orthogroup was ranked first in both the sPLSDA and the Mann-Whitney-Wilcoxon test (Fig. 2, Supplementary Data 6). Within the Trebouxiophyceae, GH8 are specifically present in LAS and in five NSA sister-species to well-characterized LAS (Supplementary Fig. 4, Supplementary Data 9). To ensure that the presence of the GH8 enzyme in LAS genomes was not due to potential contaminations, we scrutinized the scaffold they belong to. We found the GH8 well-anchored in their respective scaffolds, surrounded by algal genes. Re-mapping of the raw reads on these scaffolds excluded the possibility of chimeric scaffolds (Fig. 4b, Supplementary Fig. 5). The assignment of this orthogroup to the GH8 family was also confirmed by an unsupervised classification of carbohydrate-active enzymes using CUPP (Supplementary Data 10). The phylogenetic analysis identified GH8

members in bacteria, but also in 209 fungal species and strains (Fig. 4a, Supplementary Fig. 6, Supplementary Data 11), mostly from the non-symbiotic fungal phylum: the Mucoromycotina (Supplementary Data 11). Such a distribution could be explained by the presence of the GH8 enzyme clade in the eukaryotes most recent common ancestor, followed by losses and its specific retention in only two clades. However, such pattern would require losses in multiple eukaryotic lineages. The other hypothesis, the acquisition through an HGT, is more parsimonious, requiring only two events. The phylogenetic analysis thus supports at least two HGT events in the evolutionary history of the GH8 family. The GH8 originated in bacteria and was horizontally transferred to fungi and Trebouxiophyceae independently, or, alternatively, the GH8 was first transferred to Mucoromycotina fungi as an intermediate recipient between bacteria and the algae (Fig. 4a, Supplementary Fig. 6). Following this HGT, the GH8 enzyme was retained in most symbiotic Trebouxiophyceae species (11/13) but lost in most species that did not maintain the ability to engage in the lichen symbiosis (5/23 non-symbiotic Trebouxiophyceae). According to the CAZy database, members of the GH8 family catalyze the hydrolysis of fungal polymers such as chitosan or lichenans found in LFS [47,48]. In bacteria, the GH8 family has been divided into three subfamilies based on the position of the proton acceptor residue [47,48]. Alignment of reference proteins from the three bacterial GH8 subfamilies with the GH8 sequences identified in symbiotic Trebouxiophyceae revealed that they share the asparagine catalytic site with the GH8b subfamily (Supplementary Fig. 5c) known to encompass, among others, lichenase enzymes able to degrade lichenan [48]. Additionally, the 3D model of the GH8 from symbiotic Trebouxiophyceae generated using AlphaFold2 [49] positioned the catalytic asparagine in a pocket where the substrate could bind (Supplementary Fig. 7). To functionally test the enzymatic activity of the GH8 from Trebouxiophyceae, we cloned orthologs from three LAS and expressed them in a heterologous system. Among them, we successfully produced the recombinant GH8 from *Asterochloris glomerata*. Enzyme assays towards different polysaccharides confirmed a typical lichenase activity[50] for this enzyme with a significant cleavage of mixed linked β-1,3/β-1,4 glucans, while cleavage was neither observed on cellulose nor on chitosan substrates (Fig. 4c). The detailed chromatographic analysis of the major soluble products that accumulated over time upon enzyme action showed they do not correspond to β-1,4-linked cellooligosaccharides nor to β-1,3-linked laminari-oligosaccharides, thus suggesting that the degradation products belong to the mixed linked β-1,3/1,4 class. These results suggest that the GH8 enzyme was acquired through horizontal gene transfer (HGT) in the MRCA of Trebouxiophyceae, along with the ability to interact with LFS. The enzyme was later retained in species that engage in symbiosis.

The second candidate, the N0.HOG0012501 orthogroup, consists of genes belonging to the glutathione S-transferase enzyme family. Phylogenetic analysis revealed that the algae sequences are nested within bacterial clades, indicating that the likely donors of the HGT are bacteria. Moreover, the original orthogroup is dispersed across two distinct bacterial clades, suggesting the possibility of a double transfer of a similar gene: one present in nearly all symbiotic Trebouxiophyceae (9/13) and only a few non-symbiotic Trebouxiophyceae (4/23), and another one that seems Trebouxia specific (Supplementary Fig. 8). Here again, the scaffold anchoring and the read mapping did not show any sign of contamination (Supplementary Fig. 9). This family is well known for containing enzymes involved in a wide range of biological processes[51] but more especially in buffering oxidative stresses.

Discussion

Understanding the evolution of traits and the underlying genetic mechanisms has been studied in multiple contexts, from coat color in mice[52] to plant intracellular symbioses[22]. These genetic novelties may evolve through multiple mechanisms, from gene family expansion mediated by duplications to *de novo* gene birth by domain fusion or horizontal gene transfer.

One of the candidate genes originating from gene family expansion is annotated as a short-chain dehydrogenase/reductase (SDR). SDR belongs to a family encompassing enzymes involved in ribitol biosynthesis [53]. Ribitol is an acyclic pentose alcohol previously identified as the major sugar produced in lichens such as *Peltigera aphthosa* (*Coccomyxa* photobiont), *Xanthoria aureola* or *Gyalolechia bracteate* formed with *Trebouxia* spp. photobionts [54,55]. The addition of exogenous ribitol to a culture of LFS has been shown to stimulate fungal growth and developmental transitions and has been suggested as a signal for lichen initiation, in addition to being a source of carbon for the LFS [54,56]. We speculate that expansion of the SDR gene family in Trebouxiophyceae may have enhanced ribitol biosynthesis, stimulating lichen morphogenesis and carbon transfer to the LFS.

Although considered rare in eukaryotic genomes, horizontal gene transfer is becoming a common theme in the evolution of innovations. In our study, we found that two genes associated with lichenization were horizontally acquired. The first one is annotated as a glutathione S-transferase. These enzymes are known to play a role in oxidative stress response in a wide range of organisms, including in lichen fungal and algal symbionts [57]. This class of enzymes was previously identified in lichen symbionts to play a role in desiccation resistance [57].

The other horizontally-acquired orthogroup is the one that best discriminates symbiotic Trebouxiophyceae from other Chlorophytes at the phylogenetic level and belongs to the GH8 family. Contribution of HGT to the evolution of interactions between organisms has been

reported in multiple eukaryotic taxa. Ferns and *Cycas* have gained the ability to produce toxins improving resistance to herbivores[58,59] and Drosophila detoxifies cyanogenic compound found in plants [57]. In addition, such HGT events have been also identified between the LFS *Xanthoria parietina* and its associated algal symbiont, *Trebouxia decolorans*. The latter has likely horizontally acquired three genes that could have played a role in the evolution of lichenization ability [60]. The xenologous origin of cell wall-degrading enzymes has been also observed in parasites, including phytophagous insects[62] and nematodes[63]. This preponderance of plant cell wall degrading enzymes HGT might reflect a more general mechanism for the evolution of inter-organism interactions. Diversification of the enzymes occurs in microorganisms evolving on a given substrate, such as tree bark, and horizontally transferred to other eukaryotes from the same ecological niches. This transfer expands the cell wall degradation potential of the recipient species and facilitate interactions, either mutualistic, endophytic or parasitic. In the classical model for the evolution of novelties, potentiation – actualization – refinement, potentiation can be considered here as a community phenomenon associated to the diversification of enzymes.

Independently of their origin, the role of cell wall degrading enzymes, and in particular Glycoside Hydrolases (GHs), in mutualistic interactions has been well documented. The current lack of genetic model in LAS do not allow testing the biological role of the GH8 enzyme during lichenization, but a few hypotheses emerge based on the knowledge acquired on other symbiotic systems. First, the algal GH8 enzyme may play a crucial role in facilitating the establishment of a symbiotic interface between the fungus and the algae by breaking down the lichenan, a key component of the mycobiont's cell wall. Macrolichens consist of multiple layers, and lichenan has been identified predominantly in the medullary region of lichen species like *Cetraria islandica* [64]. In proximity to symbiotic Trebouxiophyceae, the fungal cell wall appears to be thinner[64], aligning with the ability of GH8 enzymes to break down lichenans present in this region. Other carbohydrate-active enzymes have previously been identified in the mycobiont *Usnea hakonensis*, and they are believed to be involved in creating a symbiotic interface by breaking down the algal cell walls (GH2, GH12), although these enzymes do not seem specific to LFS [25]. Overall, both the LFS and LAS appear to possess a genetic toolkit for degrading each other's cell walls [25,27]. Accommodation of microsymbionts via modification of the host cell wall is well described in plant symbioses such as the arbuscular mycorrhizal and ectomycorrhizal symbioses[65], and for endophytism[17], a similar function can thus be proposed for lichens. A second potential role for the GH8 enzyme could be to generate a carbon source that could be utilized by other partners within the lichen thallus. This mechanism was previously identified in other symbiotic relationships, such as that observed in *Streblomastix strix*, an oxymonad residing in the termite gut alongside a bacterial community. In this context,

bacteria employ their GHs to break down wood particles into monosaccharides, which they can subsequently assimilate [66]. Similarly, we can speculate that LAS engulfed in the macrolichens could generate simple sugars, in particular glucose monomers, from lichenans and use them as a carbon source as an alternative pathway to photosynthesis for carbon assimilation. The direct uptake of glucose by *Trebouxia* during symbiosis has been previously reported and proposed as an additional source of carbon to increase ribitol efflux [67]. Additionally, it was previously suggested that *Trebouxia* strains seem to adopt a heterotrophic lifestyle when cultured in the dark when a source of carbon such as glucose is provided in the culture medium. Hence, we can consider that LAS are mixotrophic and that the degradation of lichenans in simpler monomers could be used for their own nutrition [68]. As a working model, we propose that the combined gain of the GH8, increasing carbon availability to the LAS, and expansion of the SDR family, increasing ribitol biosynthesis, contributed to the evolution of lichenization in Trebouxiophyceae.

Our study expands the range of chlorophyte clades with sequenced genomes and revealed at least three independent origins of lichenization in green algae, confirming the convergent nature of this trait. We identified genes associated with lichenization originating from gene family expansion and horizontal gene transfers, including one enzyme able to degrade carbohydrate polymers formed in the thallus of lichen fungal symbionts. In the future, the development of genetic models in Trebouxiophyceae and lichen reconstitution in controlled conditions will allow proving the involvement of these genes in the symbiosis.

Methods

Algal cultures

Three *Trebouxia* photobionts with varied ecologies[69] were isolated from thalli of the lichen *Umbilicaria pustulata* using a micro-manipulator as described in [70]. Cultures were grown on solid 3N BBM + V medium (Bold's Basal Medium with vitamins and triple nitrate[71]) under a 30 µmol/m^2/s photosynthetic photon flux density with a 12 h photoperiod at 16°C. The identity of the isolated photobionts was validated by comparing the ITS sequence to those from [69]. Eleven additional algal cultures were obtained from the SAG Culture Collection (Göttingen). These represent a selection of both lichenized algae as well as closely-related, free-living lineages – i.e. species that were never reported to establish symbiotic associations with fungi - belonging to the classes Trebouxiophyceae and Ulvophyceae (Chlorophyta) (**Supplementary Data 1**). The cultures were maintained under the conditions described above and sub-cultured every two to three months onto fresh medium until sufficient biomass (~500 mg) for DNA isolation was obtained.

DNA isolation and sequencing

Prior to DNA isolation, we performed nuclei isolation to reduce the amount of organelle DNA, i.e. chloroplast and mitochondrial and non-target cytoplasmic components. This step has been shown to increase the read coverage of the targeted nuclear genomes and it is particularly recommended for long-read sequencing [72]. Green algae were transferred to fresh agar plates two days before nuclei isolation. For this, we used a modified protocol by Nishii et al. (2019, https://stories.rbge.org.uk/archives/30792) starting with 300-600 mg of algal material. Briefly, for each sample we prepared 20 ml of nuclei isolation buffer (NIB) consisting of 10 mM Tris-HCL pH 8.0, 30 mM EDTA pH 8.0, 100 mM KCl, 500 mM Sucrose, 5 mM Spermidine, 5 mM Spermine, 0.4% β-Mercaptoethanol, and 2% PVPP-30. The fine algal powder was transferred to 50 ml Falcon tubes with 10 ml ice-cold NIB and mixed gently. The homogenates were filtered into 50 ml centrifuge tubes through 20 µm cell strainers (pluriSelect, Leipzig, Germany), followed by a centrifugation at 2500 x g at 4°C for 10 minutes. The pellets were resuspended in 9 or 9.5 ml NIB by gently tapping the tubes. 1 or 0.5 ml of 10% Triton X-100 diluted NIB (NIBT). After a 15 min incubation on ice, the suspensions were centrifuged at 2 500 x g at 4°C for 15 minutes. The nuclei pellets were carefully resuspended in 20 ml Sorbitol buffer (100 mMTris-HCL pH 8.0, 5 mM EDTA pH 8.0, 0.35 M Sorbitol, 2% PVPP-30, 2% β-Mercaptoethanol). After a 15 min centrifugation at 5 000 x g and 4°C the supernatants were discarded, and the tubes were inverted on a paper towel to remove traces of buffer. After a RNAse A-/ Proteinase K digestion for several hours the gDNAs were isolated following the protocol by [73] with modifications described in [74] or with Qiagen Genomic-Tips.

Long-read DNA sequencing

SMRTcell libraries were constructed for samples passing quality control (**Supplementary Data 2**) according to the manufacturer's instructions of the SMRTcell Express Prep kit v2.0 following the Low DNA Input Protocol (Pacific Biosciences, Menlo Park, CA) as described in [74]. Genomic DNA was sheared to 20-kb fragments using Megaruptor 2 (Diagenode, Belgium) and then bead-size selected with AMPure PB beads (Pacific Biosciences) to remove <3-kb SMRTbell templates. SMRT sequencing was performed on the Sequel System II with Sequel II Sequencing kit 2.0 (Sequel Sequencing kit 2.1 for Sequel I system, see below) in 'circular consensus sequencing' (i.e. CCS) mode, 30 h movie time with pre-extension and Software SMRTLINK 8.0. Samples were barcoded using the Barcoded Overhang Adapters Kit-8A, multiplexed, and sequenced (3 samples/SMRT Cell) at the Genome Technology Center (RGTC) of the Radboud university medical center (Nijmegen, the Netherlands). Four samples were instead sequenced on the Sequel I system at BGI Genomics Co. Ltd. (Shenzhen, China) (**Supplementary Data 2**). In this case, one SMRT Cell was run for each sample.

RNA isolation and sequencing

For RNA isolations we used the Quick-RNA Fungal/ Bacterial Miniprep Kit (Zymo Research) starting with 30-50 mg of algal material. RNAs were further purified, when necessary, with the RNA Clean & Concentrator-5 Kit (Zymo Research). Total RNAs from the 12 algal cultures (**Supplementary Data 3**) were sent to Novogene (Hong Kong, China) for library preparation and sequencing. mRNA-seq was performed on the Illumina NovaSeq platform (paired-end 150 bp sequencing read length).

Genome assembly

Sequel II samples were demultiplexed using lima (v1.9.0, SMRTlink) and the options '--same --min-score 26 –peek-guess'. De novo assembly was carried out for each PacBio (Sequel/Sequel II) CLR subreads set using the genome assembler Flye (version 2.7-b1587)[75] in CLR mode and default parameters. Each assembly was polished once LAS part of the Flye workflow and a second time with the PacBio tool GCpp v2.0.0 with default parameters (v1.9.0, SMRTlink). The polished assemblies were scaffolded using SSPACE-LongRead v1.1[76] with default parameters.

The received scaffolds were taxonomically binned via BLASTx against the NCBI nr database (accessed in September 2020) using DIAMOND (--more-sensitive) in MEGAN v.6.7.7 [77], with an e-value threshold of 1E-10 and the MEGAN-LR algorithm [77]. Only scaffolds assigned to the Chlorophyta were retained for subsequent analysis.

Genome and transcriptome annotation

Genome assemblies were softmasked using Red [78] and annotated using BRAKER2 pipeline [79]. BRAKER2 was run with '--etpmode --softmasking --gff3 --cores 1' options. The pipeline in etpmode first train GeneMark-ETP with proteins of any evolutionary distance (i.e. OrthoDB) and RNA-Seq hints and subsequently trains AUGUSTUS based on GeneMark-ETP predictions. AUGUSTUS predictions are also performed with hints from both sources. The OrthoDB input proteins used by ProtHint is a combination of OrthoDB v10 (https://v100.orthodb.org/download/odb10_plants_fasta.tar.gz) and proteins from six species investigated in this study. To complement orthology-based annotation, available or generated RNA-Seq data for each species were used LAS hints in BRAKER2. Adapters and low-quality sequences were removed from the raw fastq files using cutadapt v2.1 [80] and TrimGalore v0.6.5, (https://github.com/FelixKrueger/TrimGalore) with the options -q 30 --length 20. The cleaned reads were mapped against the corresponding genomes using HISAT2 v2.1.0 [81] with the options --score-min L,-0.6,-0.6 --max-intronlen 10000 --dta. Duplicated reads were

removed using the markdup command from SAMtools v1.10 [82]. These final alignments data were used LAS hints in BRAKER2 [79].

We also annotated transcriptomes of four species (*Paulbroadya petersii, Paulbroadya prostrata, Myrmecia biatorellae, Stichococcus bacillaris*). First, we assembled the transcriptomes from the raw reads RNAseq using DRAP v1.92 pipeline [83]. runDrap was first used on the unique samples applying the Oases RNAseq assembly software[84]. Predictions of protein-coding genes were performed using TransDecoder v5.5.0 [85] (https://github.com/TransDecoder/TransDecoder) and hits from BLASTp search in the Swissprot database (downloaded on September 2021) as well as HMMER search in the Pfam v34 database [86,87]. Completeness of newly sequenced and annotated genomes and transcriptomes was assessed using BUSCO V5.4.4 [88] with default parameters and using the Chlorophyta "odb10" database (1,519 core genes) as reference. Transcriptomes from the 1KP project were also annotated following the same procedure with SwissProt downloaded on January 2019 and Pfam v32.
Finally, functional annotation was performed for all species investigated using the InterProScan suite v5.48-83.0 [84] with the following analysis enabled: PFAM v33.1, ProSite profiles and patterns (2019_11), Panther v15.0, TIGERFAM v15.0, Gene3D v4.3.0, CDD v3.18, HAMAP 2020_05, PIRSF v3.10, PRINTS v42.0, SFLD v4.0, SMART v7.1 and SUPERFAMILY v1.75.

Proteome database building

The corresponding proteomes of the 12 newly sequenced genomes and transcriptomes were added to a database that was built with proteomes extracted from public databases such as the NCBI and ORCAE (**Supplementary Data 1**). In total, the final database is composed of 141 species that cover all the chlorophyte clades and contains both LAS and NSA.

Orthogroups reconstruction

Orthogroups reconstruction was performed using OrthoFinder v2.5.4 [89] using the 141 species database with DIAMOND v0.9.19 [90] set in ultra-sensitive mode. The estimated species tree based on orthogroups was then manually controlled and re-rooted on the outgroup species *Prasinoderma coloniale*. OrthoFinder was then re-run with this correctly rooted tree and with the MSA option to improve the orthogroups reconstruction (**Supplementary Data 4**).

Ancestral state reconstruction

The ultrametric version of the 141 species tree obtained using OrthoFinder was generated using the 'phytools' package [91] (v1.9.16) and used to perform an Ancestral State

Reconstruction (ASR). All the species were coded as LAS or NSA. The ASR was then conducted using the 'ape' [92] package (v5.7-1) and plotted using the 'phytools' package (v1.9.16) in R (v4.2.2). Both equal rate (ER) and all rate different (ARD) models were tested, and the all-rate different model was retained based on the log-likelihood value.

Genome streamlining investigation

Multiple symbionts exhibit strong genome reduction and modifications [29–32]. We compared genome size, the number of protein-coding genes, the GC content, and the transposable elements repertoire of LAS and NSA species. Each comparison has been performed using a Wilcoxon test using R v4.2.2 [93].

Transposable element annotation

We used EDTA v2.0.1 [94] to annotate transposable elements. The EDTA pipeline combines an array of specific tools for different TE types, such as long terminal repeat transposons (LTRs): LTR_FINDER v1.07 [95], LTR_retriever v2.6 [96]; terminal inverted repeat transposons (TIRs): TIR-Learner v1.19 [97]; Helitrons: HelitronScanner v1.1 [98]; terminal direct repeats (TDRs), miniature inverted transposable elements (MITEs), TIRs, LTRs: Generic Repeat Finder v1.0 [99]. It also runs RepeatModeler v2.0.3 [100] to identify unknown TEs that were not found by the other tools. The pipeline then creates a filtered, combined repeat library of consensus sequences from the different sources, uses TEsorter [101] to de-duplicate and classify consensus sequences, and finally employs RepeatMasker v4.1.2-p1 [102] to annotate the TEs in the genome. A widespread strategy is to use RepeatModeler alone to generate a library of consensus sequences and annotate them in the genome with RepeatMasker. We opted to use EDTA instead for its multi-faceted approach and filtering procedure that produces a more informative repeat library, and thus a more detailed TE annotation in the genomes. EDTA was run with the options `--sensitive 1 --anno 1 --evaluate 1.

Identification of genes associated with lichenization in Trebouxiophyceae

To identify genes linked to lichenization in Trebouxiophyceae, a sparse Partial Least Square Discriminant Analysis (sPLS-DA) was conducted using the 'mixOmics' package (v6.22.0) [103]. To do that, the 141 species were divided into two classes: the symbiotic Trebouxiophyceae and the other algae (including all the non-symbiotic species and the symbiotic Ulvophyceae). The orthogroup count (**Supplementary Data 4**) was then used as the quantitative dataset to identify the 100 orthogroups that discriminate the two classes best (**Supplementary Fig. 3**). To have a better resolution, the species-specific orthogroups and the orthogroups with two

species were removed from the study. In parallel, a Wilcoxon test was conducted on the same orthogroup dataset to identify the ones that are enriched in symbiotic Trebouxiophyceae.

Phylogenetic analysis of candidate proteins

To place expansions, contractions, and gene losses in an evolutionary context, candidate proteins were subjected to phylogenetic analysis. First, homologs of sequences from orthogroups were searched against a database containing the 141 investigated species using the BLASTp v2.9.0 algorithm [104] and an e-value threshold of 1^{e-10}. Then, retained sequences were aligned using MUSCLE v5.1.0 [105] with default parameters and obtained alignments cleaned with trimAl v1.4.1 [106] to remove positions with more than 60% of gaps. Finally, alignments were used as a matrix for maximum likelihood analysis. First, phylogenetic reconstructions have been performed using IQ-Tree v2.1.3 [107] with default parameters to obtain a global topology of the tree. Before tree reconstruction, the best-fitting evolutionary model was tested using ModelFinder [108] implemented in IQ-TREE. Branch supports were tested using 10 000 replicates of both SH-aLRT [109] and ultrafast bootstrap [110]. Trees were visualized and annotated in the iTOL v6 platform [111]. All candidate trees are provided in Supplementary Data 12.

Identification of genes differentially expressed genes between the symbiotic state of Trebouxia sp **TZW2008**

The raw reads were downloaded and submitted to the nf-core/rnaseq v3.4 [112] workflow in nextflow v21.04 [113] using '-profile debug,genotoul --skip_qc --aligner star_salmon' options. Nextflow nf-core rnaseq workflow used bedtools v2.30.0 [114], cutadapt v3.4 [80] implemented in TrimGalore! v 0.6.7, picard v2.25.7 (https://broadinstitute.github.io/picard), salmon v1.5.2 [115], samtools v1.13 [116], star v2.6.1d and v2.7.6a [117], stringtie v2.1.7 [118] and UCSC tools v377 [119].

The counted data were analysed using *edgeR* package v2.1.7 [120] with R v4.1.1 [93]. Two samples of synthetic lichen showed distant clustering to other synthetic lichen samples (DRR200314 and DRR200315, named Tresp_LicSynt_R1 and Tresp_LicSynt_R2 respectively), so we decided to remove them. Then, we removed consistently low expressed genes with less than 10 reads across each class of samples (Algal culture, Synthetic lichen, and Field lichen). After, gene counts were normalized by library size and trimmed mean of M-values (i.e. TMM) normalization method [121]. We estimated differentially expressed genes (DEGs) by comparing synthetic lichen samples and field lichen samples to algal culture. DEGs were considered with adjusted p-value (FDR method) < 0.05 and $|logFC|>1.5$ (**Supplementary Data 8**).

Horizontal Gene Transfer demonstration

Three different approaches were used to validate the putative horizontal gene transfer of the GH8. First, the GH8-coding gene of algae was verified to be anchored in large scaffolds and surrounded by other algal genes. Visualization of GH8-like positions on scaffolds was performed using the R package chromoMap v0.3.1 [122]. Secondly, reads from sequencing were mapped on the region containing the algal GH8 enzyme to control for chimeric assembly using minimap2 v2.17-r941 [123] and default parameters. Finally, a phylogenetic analysis was conducted to place algal GH8 in an evolutionary context. Using the BLASTp v2.13.0.1+ algorithm [104] with an e-value threshold of 1^{e-30} homologs of algal GH8-like were searched for in three different databases: the JGI fungal resources (accessed in February 2020, contains more than 1600 fungal genomes) and the non-redundant protein database from NCBI (May 2022) and the algae transcriptomes from the one KP project. Obtained sequences were subjected to phylogenetic analysis as described above and using MUSCLE5 [105] (v5.1.0) Super5 option for the alignment step and FastTree v2.1.10 [124]. The presence of the GH8 functional domain was determined using hmmsearch from the HMMER v3.3.1 package [125] with default parameters and using the GH8 domain model (Pfam accession: PF01270). Protein structure was predicted using AlphaFold v2.1.0 [49].

Identification of carbohydrates active enzyme using CUPP

To identify all the carbohydrate active enzymes in Chlorophytes, CUPP v4.0.0 [126] was used with default parameters and with the 2023 CUPP library.

GH8 enzyme activity analysis

Recombinant production of GH8 enzyme

From all the LAS GH8, we selected the protein from *A. glomerata* to test its enzymatic activity. Its nucleotide sequence was first codon optimized for Pichia pastoris. The gene was synthesized by Genewiz (South Plainfield, New-Jersey, USA) and inserted in the expression vector pPICZαA (Invitrogen, Carlsbad, California, USA) in frame with the C-terminal poly-histidine tag. Transformation of competent *P. pastoris* X33 cells (Invitrogen, Carlsbad, California, USA) was performed by electroporation using the PmeI-linearized pPICZαA recombinant plasmid as described in [127] using the facilities of the 3PE Platform (Pichia Pastoris Protein Express; www.platform3pe.com/). Zeocin-resistant transformants were then screened for protein production. The best-producing transformant was grown in 4 L BMGY medium (10 g.L-1 glycerol, 10 g.L-1 yeast extract, 20 g.L-1 peptone, 3.4 g.L-1 YNB, 10 g.L-1 ammonium suASte, 100 mM phosphate buffer pH 6 and 0.2 g.L−1 of biotin) at 30°C and 200 rpm to an optical density at 600 nm of 2-6. Expression was induced by transferring cells into 800 mL of BMMY media at 20 °C in an orbital shaker (200 rpm) for another 3 days. Each day, the

medium was supplemented with 3% (v/v) methanol. The cells were harvested by centrifugation, and just before purification, the pH was adjusted to 7.8 and was filtrated on 0.45-µm membrane (Millipore, Burlington, Massachusetts, USA).

Purification by affinity chromatography

Filtered culture supernatant was loaded onto a 20 mL HisPrep FF 16/10 column (Cytiva, Vélizy-Villacoublay, France) equilibrated with buffer A (Tris-HCl 50 mM pH 7.8, NaCl 150 mM, imidazole 10 mM) that was connected to an Äkta pure (Cytiva). The (His)6-tagged recombinant protein was eluted with buffer B (Tris-HCl 50 mM pH 7.8, NaCl 150 mM, imidazole 500 mM). Fractions containing the recombinant protein were pooled, concentrated, and dialyzed against sodium acetate buffer 50 mM, pH 5.2. The concentration of the purified protein was determined by absorption at 280 nm using a Nanodrop ND-200 spectrophotometer (Thermo Fisher Scientific) with calculated molecular mass and molar extinction coefficients derived from the sequence.

Substrate cleavage assays

The enzymatic activity of *A. glomerata* GH8 was tested on different types of substrates, i.e. cellulose (Avicel), lichenan (β-1,3/1,4-glucan), and chitosan. All substrates except Avicel cellulose (PH-101, Sigma-Aldrich), were purchased from Megazyme (Bray, Ireland). Amorphous cellulose (Phosphoris acid swollen cellulose or PASC) was prepared from Avicel cellulose as described in [128]. Enzyme assays were performed in a total volume of 200 µL containing 1% (w/v) polysaccharides or 0.5 mM of oligosaccharides in 50 mM pH 7.0 sodium phosphate buffer with 4 µM of *A. glomerata* GH8. The samples were incubated in a thermomixer (Eppendorf) at 30°C and 1000 rpm, for 24-48 h. The samples were then boiled for 10 min to stop the enzymatic reaction and centrifuged at 15,000 *g* for 5 min. The enzyme reactions were analyzed by high-performance anion-exchange chromatography coupled with pulsed amperometric detection (HPAEC-PAD) (Dionex ICS6000 system, Thermo Fisher Scientific, Waltham, MA, USA). The system is equipped with a CarboPac-PA1 guard column (2 x 50 mm) and a CarboPac-PA1 column (2 x 250 mm) kept at 30°C. Elution was carried out at a flow rate of 0.25 mL.min-1 and 25 µL was injected. The solvents used were NaOH 100 mM (eluent A) and NaOAc (1 M) in 100 mM NaOH (eluent B). The initial conditions were set to 100% eluent A, following gradient was applied: 0-10 min, 0-10% B; 10-35 min, 10-30% B; 35-40 min, 30-100% B (curve 6); 40-41 min, 100-0% B; 41-50 min, 100% A.

Data availability

Genome and transcriptome data from this study were deposited in NCBI under the BioProject PRJNA790449.

The following database were used in this study: NCBI NR (accessed September 2020 for genome annotation and May 2022 for HGT investigation), Pfam v32 (1KP transcriptome annotation) and v34 (this study transcriptome annotation), OrthoDB v10, Mycocosm (last accessed in February 2020) and SwissProt (last accessed September 2021 for transcriptome annotation from this study and January 2019 for transcriptome annotation from the 1KP project).

Publicly available genomes used in this study can be found in the NCBI under the following accession codes: *Auxenochlorella protothecoides* 710 [GCF_000733215.1], *Auxenochlorella protothecoides* UTEX25 [GCA_003709365.1], *Chlorella sorokiniana* [GCA_002245835.2], *Chlorella variabilis* NC64A [GCF_000147415.1], *Helicosporidium sp.* ATCC 50920 [GCA_000690575.1], *Micractinium conductrix* SAG241.80 [GCA_002245815.2], *Parachlorella kessleri* iCABeR21 [GCA_015712045.1], *Prototheca wickerhamii* HMC1 [GCA_003255715.1], *Ulva prolifera* [GCA_004138255.1].

Picochlorum sp. RCC4223 and *Ulva mutabilis* genomes were retrieved from ORCAE database respectively at https://bioinformatics.psb.ugent.be/gdb/RCC4223/ and https://bioinformatics.psb.ugent.be/gdb/ulva/.

Transcriptomes of *Pseudochlorella pringsheimii*, *Watanabea reniformis, Elliptochloris marina* were assembled from SRA available in the NCBI under the respective series of SRA accession codes: [SRR11611235, SRR11611236, SRR11611237, SRR11611238], [SRR16849198] and [SRR3952294, SRR5133332] respectively.

Annotations of the 1KP transcriptomes have been deposited in FigShare under the DOI: 10.6084/m9.figshare.25611138.

Code availability

No custom code has been produced in this study.

References

1. Field, K. J., Pressel, S., Duckett, J. G., Rimington, W. R. & Bidartondo, M. I. Symbiotic options for the conquest of land. *Trends in Ecology & Evolution* **30**, 477–486 (2015).
2. Rich, M. K. *et al.* Lipid exchanges drove the evolution of mutualism during plant terrestrialization. *Science* **372**, 864–868 (2021).
3. Rensing, S. A. Great moments in evolution: the conquest of land by plants. *Current Opinion in Plant Biology* **42**, 49–54 (2018).
4. Puginier, C., Keller, J. & Delaux, P.-M. Plant–microbe interactions that have impacted plant terrestrializations. *Plant Physiology* **190**, 72–84 (2022).
5. Nash, T. H. *Lichen Biology*. (Cambridge University Press, Cambridge, 2008).
6. Grimm, M. *et al.* The Lichens' Microbiota, Still a Mystery? *Front. Microbiol.* **12**, 623839 (2021).
7. Spribille, T., Resl, P., Stanton, D. E. & Tagirdzhanova, G. Evolutionary biology of lichen symbioses. *New Phytologist* **234**, 1566–1582 (2022).
8. Spribille, T. *et al.* Basidiomycete yeasts in the cortex of ascomycete macrolichens. *Science* **353**, 488–492 (2016).
9. Hawksworth, D. L. & Grube, M. Lichens redefined as complex ecosystems. *New Phytologist* **227**, 1281–1283 (2020).
10. Tagirdzhanova, G. *et al. Evidence for a Core Set of Microbial Lichen Symbionts from a Global Survey of Metagenomes*. http://biorxiv.org/lookup/doi/10.1101/2023.02.02.524463 (2023) doi:10.1101/2023.02.02.524463.
11. Pirozynski, K. A. & Malloch, D. W. The origin of land plants: A matter of mycotrophism. *Biosystems* **6**, 153–164 (1975).
12. Delaux, P.-M. *et al.* Comparative Phylogenomics Uncovers the Impact of Symbiotic Associations on Host Genome Evolution. *PLoS Genet* **10**, e1004487 (2014).
13. Griesmann, M. *et al.* Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* **361**, eaat1743 (2018).

681 14. van Velzen, R. *et al.* Comparative genomics of the nonlegume *Parasponia* reveals insights into
682 evolution of nitrogen-fixing rhizobium symbioses. *Proc. Natl. Acad. Sci. U.S.A.* **115**, (2018).
683 15. Mycorrhizal Genomics Initiative Consortium *et al.* Convergent losses of decay mechanisms and
684 rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat Genet* **47**, 410–415 (2015).
685 16. Kiss, E. *et al.* Comparative genomics reveals the origin of fungal hyphae and multicellularity.
686 *Nat Commun* **10**, 4080 (2019).
687 17. Mesny, F. *et al.* Genetic determinants of endophytism in the Arabidopsis root mycobiome. *Nat*
688 *Commun* **12**, 7227 (2021).
689 18. Gargas, A., DePriest, P. T., Grube, M. & Tehler, A. Multiple Origins of Lichen Symbioses in Fungi
690 Suggested by SSU rDNA Phylogeny. *Science* **268**, 1492–1495 (1995).
691 19. Lutzoni, F., Pagel, M. & Reeb, V. Major fungal lineages are derived from lichen symbiotic
692 ancestors. *Nature* **411**, 937–940 (2001).
693 20. Nelsen, M. P., Lücking, R., Boyce, C. K., Lumbsch, H. T. & Ree, R. H. The macroevolutionary
694 dynamics of symbiotic and phenotypic diversification in lichens. *Proc. Natl. Acad. Sci. U.S.A.*
695 **117**, 21495–21503 (2020).
696 21. Sanders, W. B. & Masumoto, H. Lichen algae: the photosynthetic partners in lichen symbioses.
697 *The Lichenologist* **53**, 347–393 (2021).
698 22. Radhakrishnan, G. V. *et al.* An ancestral signalling pathway is conserved in intracellular
699 symbioses-forming plant lineages. *Nat. Plants* **6**, 280–289 (2020).
700 23. Lutzoni, F. *et al.* Contemporaneous radiations of fungi and plants linked to symbiosis. *Nat*
701 *Commun* **9**, 5451 (2018).
702 24. Armaleo, D. *et al.* The lichen symbiosis re-viewed through the genomes of Cladonia grayi and
703 its algal partner Asterochloris glomerata. *BMC Genomics* **20**, 605 (2019).
704 25. Kono, M., Kon, Y., Ohmura, Y., Satta, Y. & Terai, Y. In vitro resynthesis of lichenization reveals
705 the genetic background of symbiosis-specific fungal-algal interaction in Usnea hakonensis. *BMC*
706 *Genomics* **21**, 671 (2020).

707 26. Pichler, G., Muggia, L., Carniel, F. C., Grube, M. & Kranner, I. How to build a lichen: from
708 metabolite release to symbiotic interplay. *New Phytologist* **238**, 1362–1378 (2023).
709 27. Resl, P. *et al.* Large differences in carbohydrate degradation and transport potential among
710 lichen fungal symbionts. *Nat Commun* **13**, 2634 (2022).
711 28. Wang, Y. *et al.* Regulation of symbiotic interactions and primitive lichen differentiation by
712 UMP1 MAP kinase in Umbilicaria muhlenbergii. *Nat Commun* **14**, 6972 (2023).
713 29. Ran, L. *et al.* Genome Erosion in a Nitrogen-Fixing Vertically Transmitted Endosymbiotic
714 Multicellular Cyanobacterium. *PLoS ONE* **5**, e11486 (2010).
715 30. McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nat Rev*
716 *Microbiol* **10**, 13–26 (2012).
717 31. Nechitaylo, T. Y. *et al.* Incipient genome erosion and metabolic streamlining for antibiotic
718 production in a defensive symbiont. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023047118 (2021).
719 32. Noh, S. Linear paths to genome reduction in a defensive symbiont. *Proc. Natl. Acad. Sci. U.S.A.*
720 **118**, e2106280118 (2021).
721 33. Soltis, D. E. *et al.* Chloroplast gene sequence data suggest a single origin of the predisposition
722 for symbiotic nitrogen fixation in angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 2647–2651
723 (1995).
724 34. Van Velzen, R., Doyle, J. J. & Geurts, R. A Resurrected Scenario: Single Gain and Massive Loss of
725 Nitrogen-Fixing Nodulation. *Trends in Plant Science* **24**, 49–57 (2019).
726 35. Li, L. *et al.* The genome of Prasinoderma coloniale unveils the existence of a third phylum
727 within green plants. *Nat Ecol Evol* **4**, 1220–1231 (2020).
728 36. Thüs, H. *et al.* Revisiting photobiont diversity in the lichen family Verrucariaceae (Ascomycota).
729 *European Journal of Phycology* **46**, 399–415 (2011).
730 37. Nyati, S., Beck, A. & Honegger, R. Fine Structure and Phylogeny of Green Algal Photobionts in
731 the Microfilamentous Genus *Psoroglaena* (Verrucariaceae, Lichen-Forming Ascomycetes).
732 *Plant Biology* **9**, 390–399 (2007).

733 38. Zahradníková, M., Andersen, H. L., Tønsberg, T. & Beck, A. Molecular Evidence of Apatococcus ,
734 including A. fuscideae sp. nov., as Photobiont in the Genus Fuscidea. *Protist* **168**, 425–438
735 (2017).
736 39. Malavasi, V. *et al.* DNA-Based Taxonomy in Ecologically Versatile Microalgae: A Re-Evaluation
737 of the Species Concept within the Coccoid Green Algal Genus Coccomyxa (Trebouxiophyceae,
738 Chlorophyta). *PLoS ONE* **11**, e0151137 (2016).
739 40. Darienko, T., Gustavs, L. & Pröschold, T. Species concept and nomenclatural changes within the
740 genera *Elliptochloris* and *Pseudochlorella* (Trebouxiophyceae) based on an integrative
741 approach. *Journal of Phycology* **52**, 1125–1145 (2016).
742 41. Darienko, T. & Pröschold, T. Toward a monograph of non-marine Ulvophyceae using an
743 integrative approach (Molecular phylogeny and systematics of terrestrial Ulvophyceae II.).
744 *Phytotaxa* **324**, 1 (2017).
745 42. Libourel, C. *et al.* Comparative phylotranscriptomics reveals ancestral and derived root nodule
746 symbiosis programmes. *Nat. Plants* **9**, 1067–1080 (2023).
747 43. Wickell, D. A. & Li, F. On the evolutionary significance of horizontal gene transfers in plants.
748 *New Phytol* **225**, 113–117 (2020).
749 44. Cheng, S. *et al.* Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant
750 Evolution. *Cell* **179**, 1057-1067.e14 (2019).
751 45. Ma, J. *et al.* Major episodes of horizontal gene transfer drove the evolution of land plants.
752 *Molecular Plant* **15**, 857–871 (2022).
753 46. Drula, E. *et al.* The carbohydrate-active enzyme database: functions and literature. *Nucleic*
754 *Acids Research* **50**, D571–D577 (2022).
755 47. Perlin, A. S. & Suzuki, S. The structure of lichenin: selective enzymolysis studies. *Can. J. Chem.*
756 **40**, 50–56 (1962).
757 48. Adachi, W. *et al.* Crystal Structure of Family GH-8 Chitosanase with Subclass II Specificity from
758 Bacillus sp. K17. *Journal of Molecular Biology* **343**, 785–795 (2004).

759 49. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–
760 589 (2021).
761 50. Lafond, M. *et al.* The Quaternary Structure of a Glycoside Hydrolase Dictates Specificity toward
762 β-Glucans. *Journal of Biological Chemistry* **291**, 7183–7194 (2016).
763 51. Hernández Estévez, I. & Rodríguez Hernández, M. Plant Glutathione S-transferases: An
764 overview. *Plant Gene* **23**, 100233 (2020).
765 52. Barrett, R. D. H. *et al.* Linking a mutation to survival in wild mice. *Science* **363**, 499–504 (2019).
766 53. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27–
767 30 (2000).
768 54. Meeßen, J., Eppenstein, S. & Ott, S. Recognition mechanisms during the pre-contact state of
769 lichens: II. Influence of algal exudates and ribitol on the response of the mycobiont of Fulgensia
770 bracteata. *Symbiosis* **59**, 131–143 (2013).
771 55. Palmqvist, K. Carbon economy in lichens. *New Phytologist* **148**, 11–36 (2000).
772 56. Ahmadjian, V. The Lichen Symbiosis. *Nordic Journal of Botany* **14**, 588–588 (1994).
773 57. Kranner, I. *et al.* Antioxidants and photoprotection in a lichen as compared with its isolated
774 symbiotic partners. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 3141–3146 (2005).
775 58. Li, F.-W. *et al.* Fern genomes elucidate land plant evolution and cyanobacterial symbioses.
776 *Nature Plants* **4**, 460–472 (2018).
777 59. Liu, Y. *et al.* The Cycas genome and the early evolution of seed plants. *Nat. Plants* **8**, 389–401
778 (2022).
779 60. Beck, A., Divakar, P. K., Zhang, N., Molina, M. C. & Struwe, L. Evidence of ancient horizontal
780 gene transfer between fungi and the terrestrial alga Trebouxia. *Org Divers Evol* **15**, 235–248
781 (2015).
782 61. Wang, B. *et al.* Co-opted genes of algal origin protect C. elegans against cyanogenic toxins.
783 *Current Biology* **32**, 4941-4948.e3 (2022).

784 62. Kirsch, R. *et al.* Metabolic novelty originating from horizontal gene transfer is essential for leaf
785 beetle survival. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2205857119 (2022).
786 63. Haegeman, A., Jones, J. T. & Danchin, E. G. J. Horizontal Gene Transfer in Nematodes: A
787 Catalyst for Plant Parasitism? *MPMI* **24**, 879–887 (2011).
788 64. Honegger, R. & Haisch, A. Immunocytochemical location of the (1→3) (1→4)-β-glucan lichenin
789 in the lichen-forming ascomycete *Cetraria islandica* (Icelandic moss) [1]. *New Phytologist* **150**,
790 739–746 (2001).
791 65. Gong, Y., Lebreton, A., Zhang, F. & Martin, F. Role of carbohydrate-active enzymes in
792 mycorrhizal symbioses. *Essays in Biochemistry* **67**, 471–478 (2023).
793 66. Treitli, S. C., Kolisko, M., Husník, F., Keeling, P. J. & Hampl, V. Revealing the metabolic capacity
794 of *Streblomastix strix* and its bacterial symbionts using single-cell metagenomics. *Proc. Natl.*
795 *Acad. Sci. U.S.A.* **116**, 19675–19684 (2019).
796 67. Tapper, R. Glucose uptake by *Trebouxia* and associated fungal symbiont in the lichen
797 symbiosis. *FEMS Microbiology Letters* **10**, 103–106 (1981).
798 68. Ahmadjian, V. Trebouxia: Reflections on a Perplexing and Controversial Lichen Photobiont. in
799 *Symbiosis: mechanisms and model systems* (ed. Seckbach, J.) 375–383 (Kluwer Academic
800 Publishers, Dordrecht ; Boston, 2002).
801 69. Dal Grande, F. *et al.* Environment and host identity structure communities of green algal
802 symbionts in lichens. *New Phytol* **217**, 277–289 (2018).
803 70. Beck, A. & Hans-Ulrich, K. Analysis of the Photobiont Population in Lichens Using a Single-Cell
804 Manipulator. *Symbiosis* 57–67 (2001).
805 71. Ahmadjian, V. Lichen synthesis. *Österreichische Botanische Zeitschrift 116* 306–311 (1969).
806 72. Bethune, K. *et al.* Long-fragment targeted capture for long-read sequencing of plastomes. *Appl*
807 *Plant Sci* **7**, e1243 (2019).
808 73. Mayjonade, B. *et al.* Extraction of high-molecular-weight genomic DNA for long-read
809 sequencing of single molecules. *BioTechniques* **61**, 203–205 (2016).

810 74. Merges, D., Dal Grande, F., Greve, C., Otte, J. & Schmitt, I. Virus diversity in metagenomes of a
811 lichen symbiosis (*Umbilicaria phaea*): complete viral genomes, putative hosts and elevational
812 distributions. *Environmental Microbiology* **23**, 6637–6650 (2021).
813 75. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using
814 repeat graphs. *Nat Biotechnol* **37**, 540–546 (2019).
815 76. Boetzer, M. & Pirovano, W. SSPACE-LongRead: scaffolding bacterial draft genomes using long
816 read sequence information. *BMC Bioinformatics* **15**, 211 (2014).
817 77. Huson, D. H. *et al.* MEGAN-LR: new algorithms allow accurate binning and easy interactive
818 exploration of metagenomic long reads and contigs. *Biol Direct* **13**, 6 (2018).
819 78. Girgis, H. Z. Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the
820 genomic scale. *BMC Bioinformatics* **16**, 227 (2015).
821 79. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic
822 eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein
823 database. *NAR Genomics and Bioinformatics* **3**, lqaa108 (2021).
824 80. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
825 *EMBnet j.* **17**, 10 (2011).
826 81. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and
827 genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).
828 82. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
829 83. Cabau, C. *et al.* Compacting and correcting Trinity and Oases RNA-Seq *de novo* assemblies.
830 *PeerJ* **5**, e2988 (2017).
831 84. Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. *Oases:* robust *de novo* RNA-seq assembly
832 across the dynamic range of expression levels. *Bioinformatics* **28**, 1086–1092 (2012).
833 85. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity
834 platform for reference generation and analysis. *Nat Protoc* **8**, 1494–1512 (2013).
835 86. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).

836 87. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Research* **49**, D412–
837 D419 (2021).
838 88. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO:
839 assessing genome assembly and annotation completeness with single-copy orthologs.
840 *Bioinformatics* **31**, 3210–3212 (2015).
841 89. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative
842 genomics. *Genome Biol* **20**, 238 (2019).
843 90. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using
844 DIAMOND. *Nat Methods* **18**, 366–368 (2021).
845 91. Revell, L. J. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and
846 other things). *PeerJ* **12**, e16505 (2024).
847 92. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary
848 analyses in R. *Bioinformatics* **35**, 526–528 (2019).
849 93. R Core Team. R: A language and environment for statistical computing. (2013).
850 94. Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a
851 streamlined, comprehensive pipeline. *Genome Biol* **20**, 275 (2019).
852 95. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR
853 retrotransposons. *Nucleic Acids Research* **35**, W265–W268 (2007).
854 96. Ou, S. & Jiang, N. LTR_retriever: A Highly Accurate and Sensitive Program for Identification of
855 Long Terminal Repeat Retrotransposons. *Plant Physiol.* **176**, 1410–1422 (2018).
856 97. Su, W., Gu, X. & Peterson, T. TIR-Learner, a New Ensemble Method for TIR Transposable
857 Element Annotation, Provides Evidence for Abundant New Transposable Elements in the Maize
858 Genome. *Molecular Plant* **12**, 447–460 (2019).
859 98. Xiong, W., He, L., Lai, J., Dooner, H. K. & Du, C. HelitronScanner uncovers a large overlooked
860 cache of *Helitron* transposons in many plant genomes. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 10263–
861 10268 (2014).

862 99. Shi, J. & Liang, C. Generic Repeat Finder: A High-Sensitivity Tool for Genome-Wide De Novo
863 Repeat Detection. *Plant Physiol.* **180**, 1803–1815 (2019).
864 100. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element
865 families. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9451–9457 (2020).
866 101. Zhang, R.-G., Wang, Z.-X., Ou, S. & Li, G.-Y. *TEsorter: Lineage-Level Classification of*
867 *Transposable Elements Using Conserved Protein Domains*.
868 http://biorxiv.org/lookup/doi/10.1101/800177 (2019) doi:10.1101/800177.
869 102. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-4.0. (2013).
870 103. Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixOmics: An R package for 'omics feature
871 selection and multiple data integration. *PLoS Comput Biol* **13**, e1005752 (2017).
872 104. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
873 105. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
874 *Nucleic Acids Research* **32**, 1792–1797 (2004).
875 106. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated
876 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
877 107. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in
878 the Genomic Era. *Molecular Biology and Evolution* **37**, 1530–1534 (2020).
879 108. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder:
880 fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**, 587–589 (2017).
881 109. Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies:
882 Assessing the Performance of PhyML 3.0. *Systematic Biology* **59**, 307–321 (2010).
883 110. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the
884 Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **35**, 518–522 (2018).
885 111. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
886 display and annotation. *Nucleic Acids Research* **49**, W293–W296 (2021).

887 112. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat*
888 *Biotechnol* **38**, 276–278 (2020).
889 113. Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat Biotechnol*
890 **35**, 316–319 (2017).
891 114. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features.
892 *Bioinformatics* **26**, 841–842 (2010).
893 115. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-
894 aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).
895 116. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079
896 (2009).
897 117. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
898 118. Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2.
899 *Genome Biol* **20**, 278 (2019).
900 119. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling
901 browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).
902 120. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR : a Bioconductor package for differential
903 expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
904 121. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression
905 analysis of RNA-seq data. *Genome Biol* **11**, R25 (2010).
906 122. Anand, L. & Rodriguez Lopez, C. M. ChromoMap: an R package for interactive visualization of
907 multi-omics data and annotation of chromosomes. *BMC Bioinformatics* **23**, 33 (2022).
908 123. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100
909 (2018).
910 124. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing Large Minimum Evolution Trees
911 with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution* **26**, 1641–1650
912 (2009).

125. Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, 431 (2010).

126. Barrett, K. & Lange, L. Peptide-based functional annotation of carbohydrate-active enzymes by conserved unique peptide patterns (CUPP). *Biotechnol Biofuels* **12**, 102 (2019).

127. Haon, M. *et al.* Recombinant protein production facility for fungal biomass-degrading enzymes using the yeast Pichia pastoris. *Front. Microbiol.* **6**, (2015).

128. Bennati-Granier, C. *et al.* Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by Podospora anserina. *Biotechnol Biofuels* **8**, 90 (2015).

Acknowledgements

The authors thank the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, https://doi.org/10.15454/1.5572369328961167E12) for providing computing resources. J.K., C.L. and P-M.D are supported by the project Engineering Nitrogen Symbiosis for Africa (ENSA) currently funded through a grant to the University of Cambridge by the Bill & Melinda Gates Foundation (OPP1172165) and the UK Foreign, Commonwealth and Development Office as Engineering Nitrogen Symbiosis for Africa (OPP1172165). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 101001675 - ORIGINS) to P-M.D. This work was supported by the "Laboratoires d'Excellence (LABEX)" TULIP (ANR-10-LABX-41)" and by the "École Universitaire de Recherche (EUR)" TULIP-GS (ANR-18-EURE-0019). F.D.G. was supported by the LOEWE-Centre for Translational Biodiversity Genomics (TBG) funded by the Hessen State Ministry of Higher Education, Research and the Arts (HMWK). We thank Carola Greve for assistance with PacBio library preparation, Anjuli Calchera for bioinformatic support, Andreas Beck for assistance with the micromanipulator algal cell isolation, Nicolas Piganeau and Bas Tolhuis for their helpful advice on PacBio sequencing and technical support.

Author Contributions

FDG, PMD and JK designed the project. JK, FDG, CP, CL, PS, MP, MH, SG and JO conducted experiments. JK, FDG, PS, CP, CL, JGB and PMD analyzed data. CP, JK and P-MD wrote the manuscript with inputs from all authors.

Competing Interests

The authors declare no competing interests.

Figures

Table 1 | List of species sequenced, the class they belong to, their symbiotic status (LAS: lichen algal symbionts, NSA: non-symbiotic algae), the resource type (G: genomic, T: transcriptomic), the genome sizes, the N50, the number of protein and the BUSCO completeness.

Species	Lichens	Class	Resource type	Genome size (Mb)	N50	Number of CDS/proteins	Busco score (%)
Apatococcus fuscideae (ApafusSAG2523)	LAS	Trebouxiophyceae	G	102.683555	10579	12399	75.3
Apatococcus lobatus (ApalobSAG2145)	NSA	Trebouxiophyceae	G	106.452463	15974	11112	95.6
Coccomyxa pringsheimii (CocpriSAG2167)	LAS	Trebouxiophyceae	G	50.915843	15647	10022	95.4
Elliptochloris bilobata (EllbilSAG24580)	LAS	Trebouxiophyceae	G	52.254682	9434	8676	93.7
Myrmecia biatorellae (MyrbiaSAG882)	LAS	Trebouxiophyceae	T	NA	NA	15547	73.6
Myrmecia bisecta (MyrbisSAG2043)	NSA	Trebouxiophyceae	G	83.484054	15027	12551	96.9
Symbiochloris irregularis (SymirrSAG2036)	NSA	Trebouxiophyceae	G	65.271117	10287	10921	91
Trebouxia sp (TrespOTU5)	LAS	Trebouxiophyceae	G	68.275875	70602	11710	94.4
Trebouxia sp (TrespOTU1)	LAS	Trebouxiophyceae	G	70.863246	9004	12712	96.3
Trebouxia sp (TrespOTU3)	LAS	Trebouxiophyceae	G	62.215734	9923	11096	93
Paulbroadya petersii (PaupetSAG2240)	LAS	Ulvophyceae	T	NA	NA	18999	76.8
Paulbroadya prostrata (PauproSAG2392)	NSA	Ulvophyceae	T	NA	NA	15610	81.9

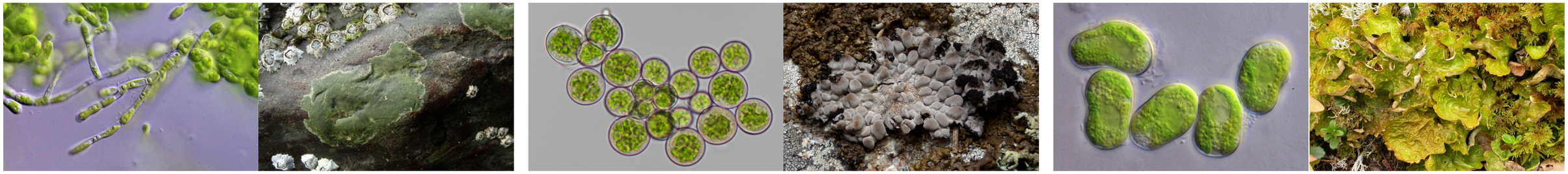
Fig. 1 | Algal species genomes and transcriptomes sampling and ancestral state reconstruction. (**a**) Pictures of three lichens and their algal partners: *Paulbroadya petersii and Verrucaria mucosa (right), Trebouxia sp* OTU1 and *Umbilicaria pustulata* (middle), *Myrmecia biatorellae* and *Lobaria linita* (right). (**b**) Algal genomes (oranges) and transcriptomes (blue) sampling, their sources, and their symbiotic habit (LAS: lichen algal symbionts, NAS: non-symbiotic algae). (**c**) Phylogenetic tree of Chlorophytes and ancestral state reconstruction of lichenization (ability to be involved in lichens) using the All Rates Different (ARD) model. LAS are indicated in black and NSA in grey. Black stars indicate the species that were sequenced for this study.

Fig. 2 | sparse PLS-DA and Mann-Whitney-Wilcoxon results. (**a**) Individual sPLS-DA plot of the first two components discriminating chlorophyte species into two groups (grey: symbiotic Trebouxiophyceae, green: other algae) according to the 100 best orthogroups. (**b**) The 100 best orthogroups identified with the sPLS-DA and their contribution on the first component. (**c**) Barplots of the pvalues of the significant orthogroups using a two-sided Mann-Whitney-Wilxocon test (pvalue < 0.01 or –log10(pvalue)>2). The bars are colored according to the Wilcoxon pvalues.

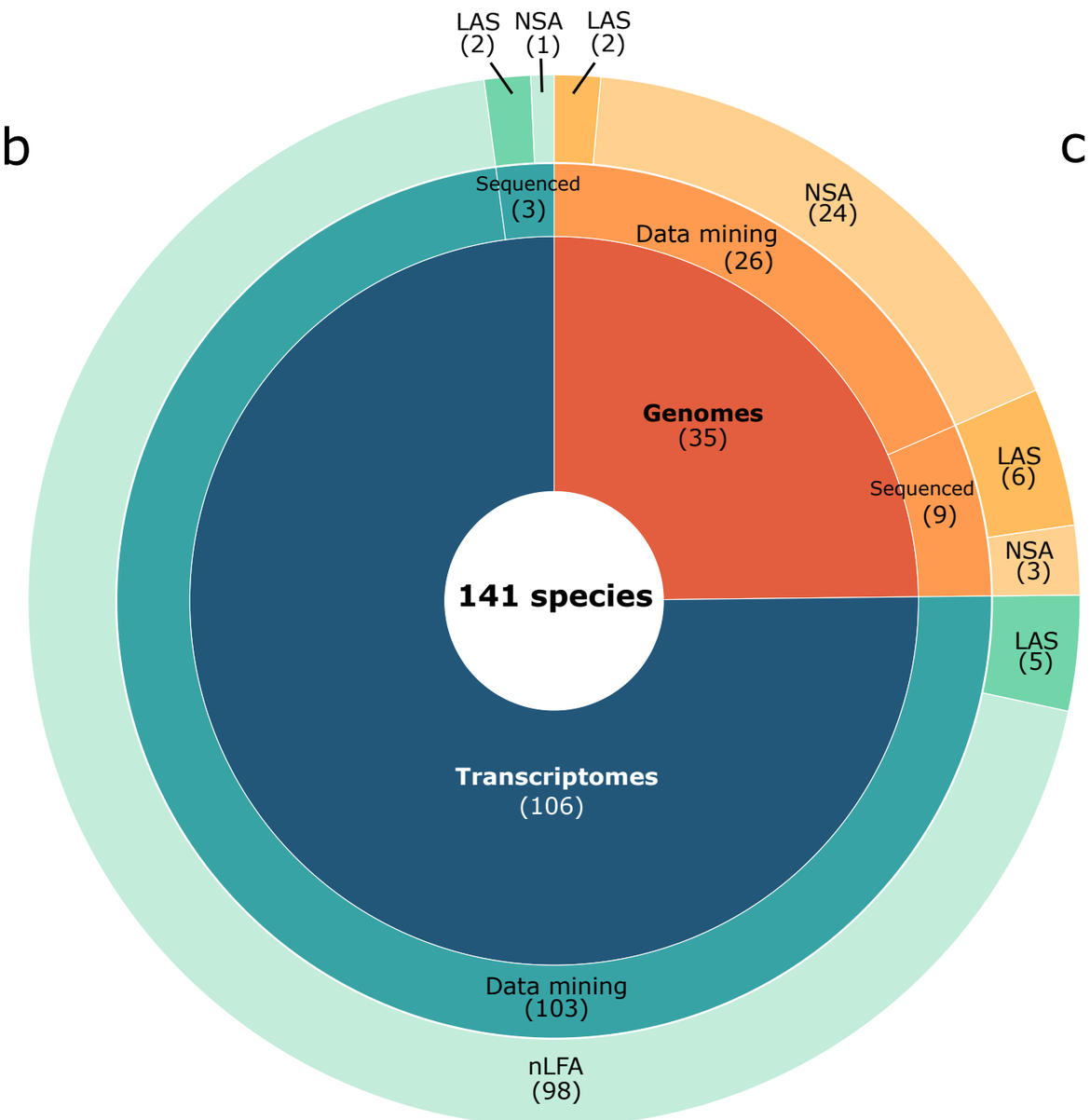
Fig. 3 | Distribution of symbiotic-associated candidate genes in Chlorophytes. Chlorophytes phylogeny, heatmap of the number of genes per species and per orthogroup for the ones that contain at least one differentially expressed gene (up and/or down-regulated) in symbiosis for *Trebouxia sp* (TZW2008) according to the data from Kono *et al.*, 2020, the contribution of each orthogroup according to Fig. 2, the transcriptomic state of the differentially expressed genes in symbiosis and the main functional annotation of each orthogroup (all the IPR domains and GO terms found in the orthogroups are available in Supplementary Table 7). Symbiotic Trebouxiophyceae are indicated with black dots. The orthogroups without a functional annotation are listed as “unknown function”.

Fig. 4 | Evolution of the GH8 enzyme in algae.
(**a**) Unrooted maximum likelihood tree of the sequences corresponding to the GH8 family. Branches are colored as follow: bacteria in pink, fungi in brown and chlorophyte algae in green. (**b**) GH8-like anchoring in scaffold of the different chlorophyte algae. The GH8 position is indicated by the orange lines. (**c**) Enzymatic activity of the GH8 enzyme from *A. glomerata* on, from top to bottom: lichenan, cellulose and chitosan (data presented as mean values +/- standard deviation with n=3 independant replicates).

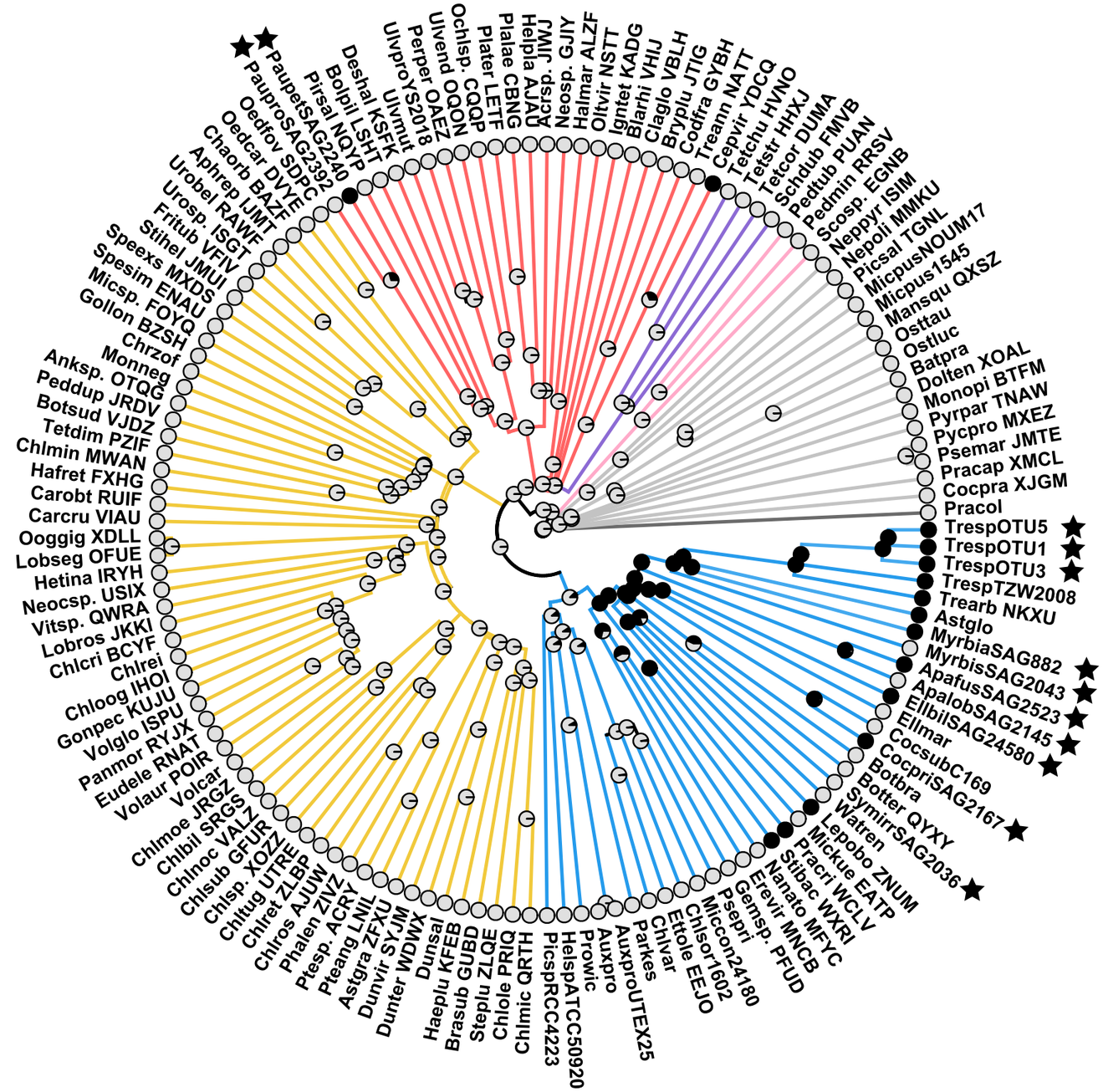
a

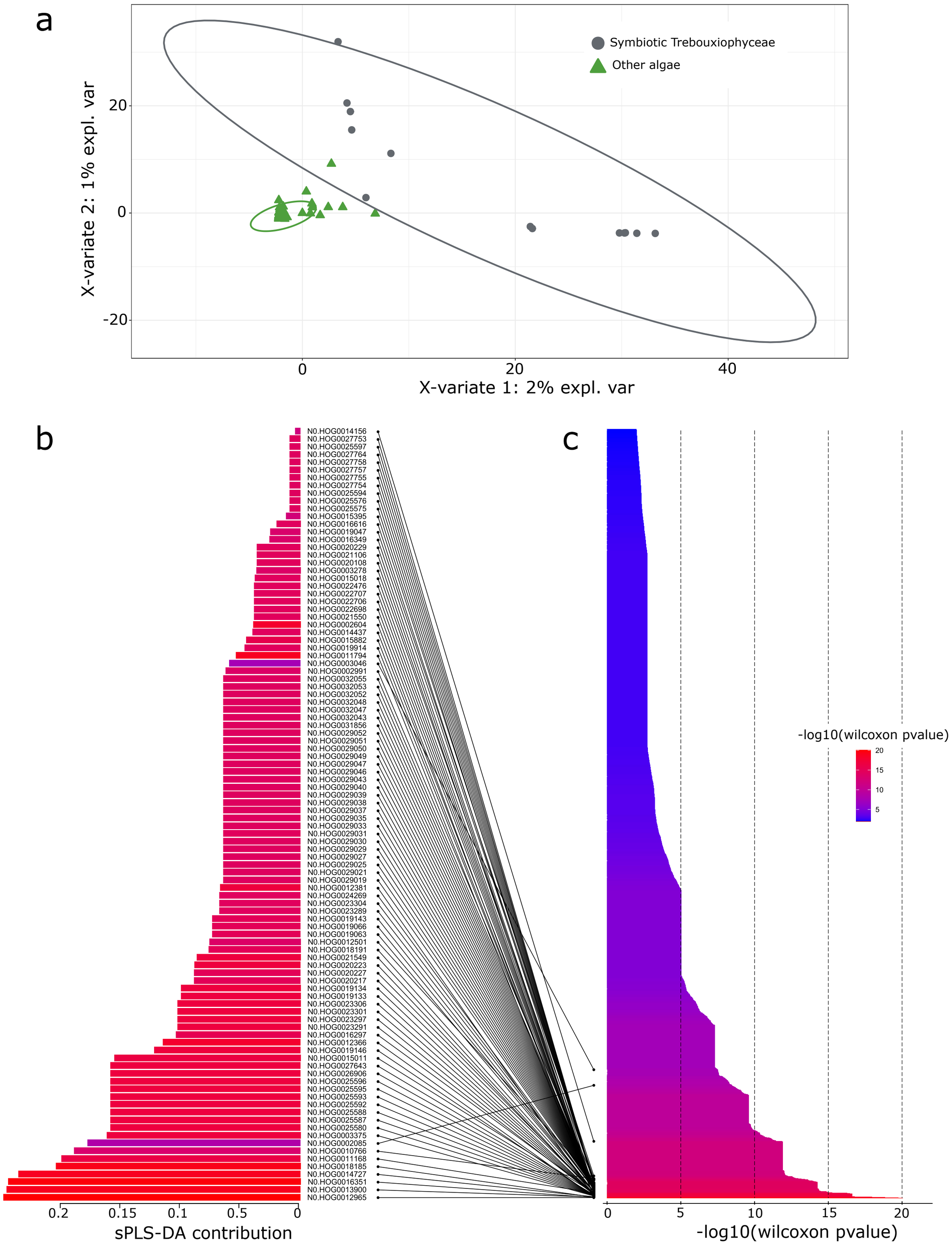


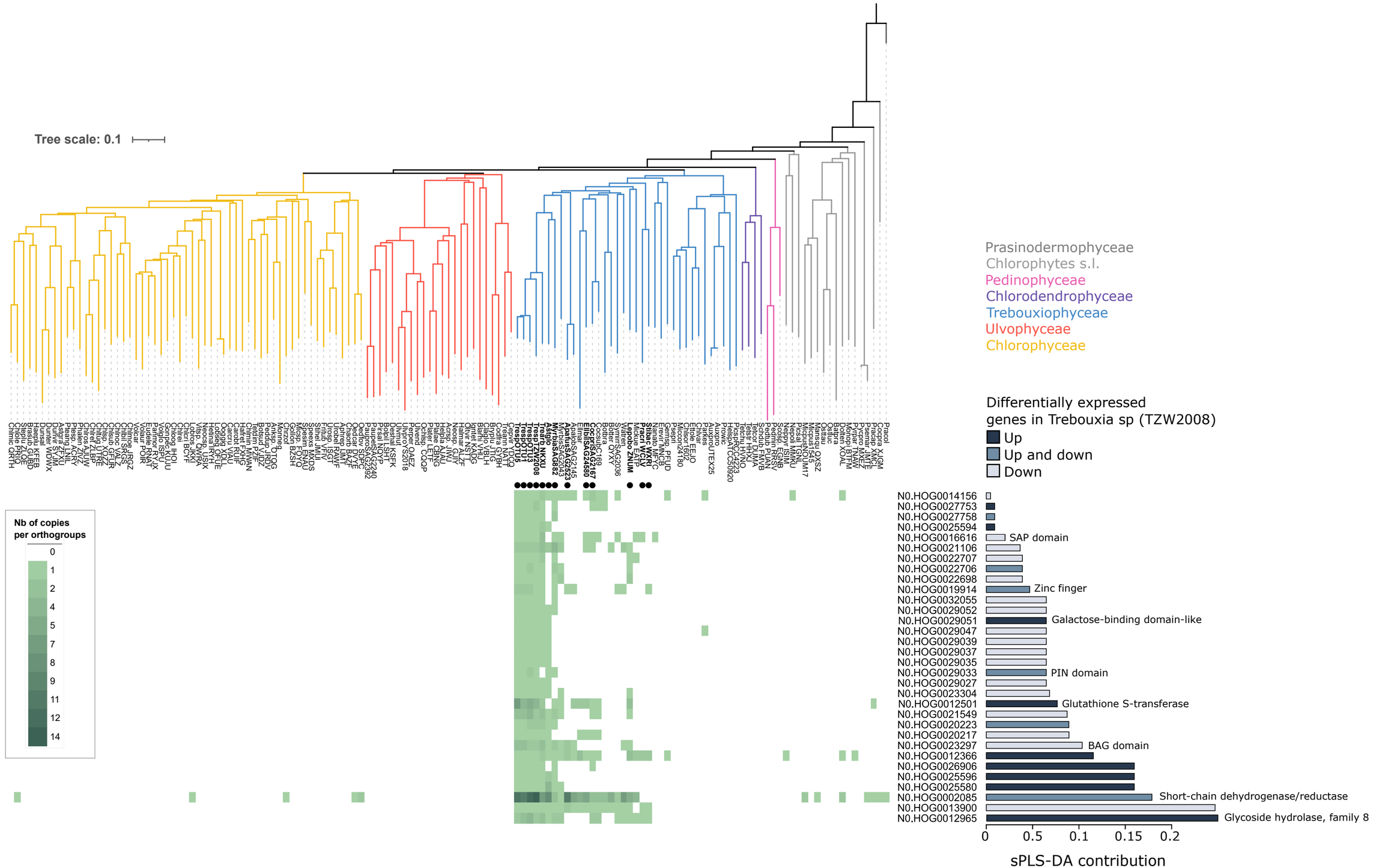
b

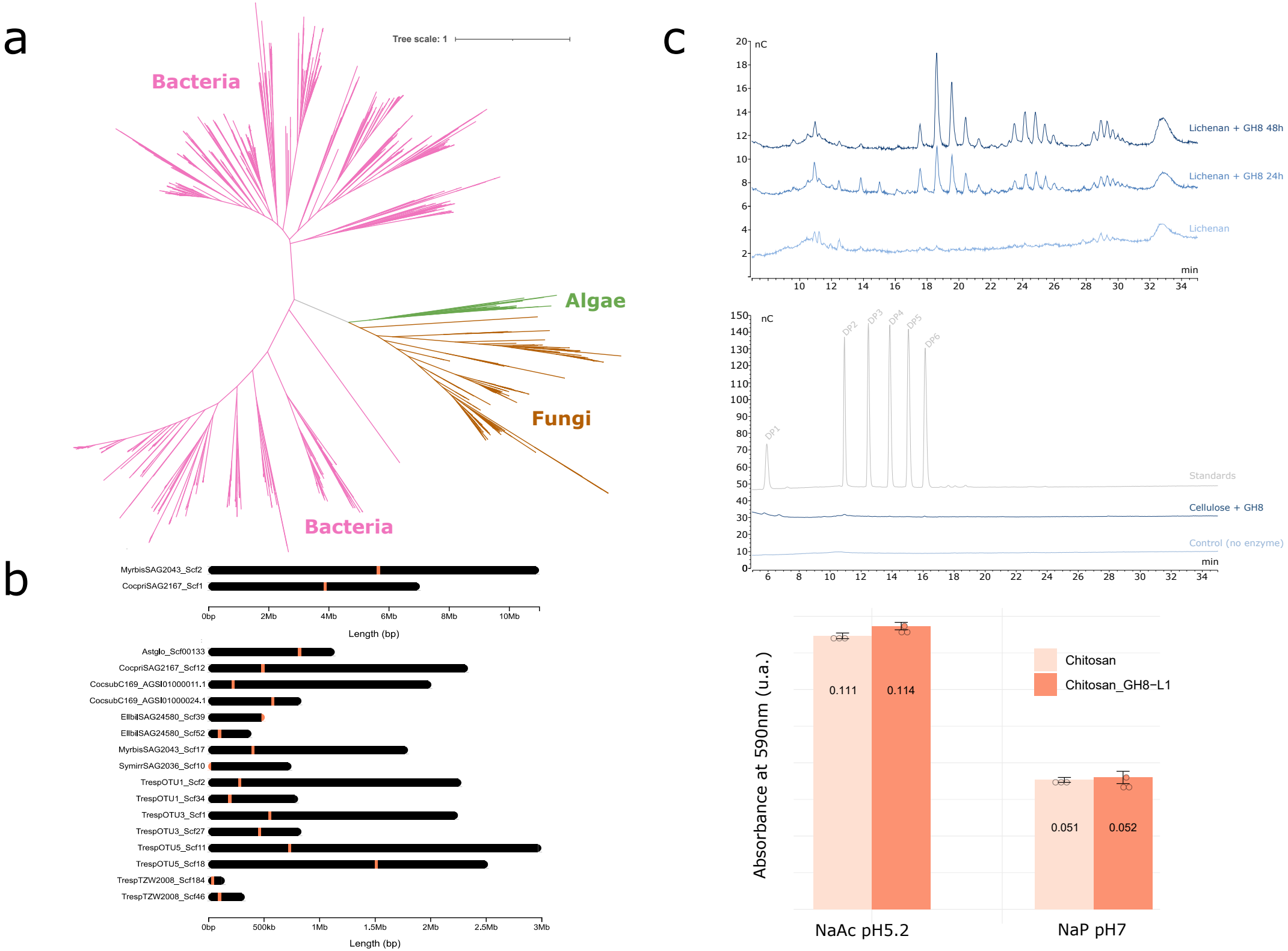

LAS (2)
NSA (1)
LAS (2)
Sequenced (3)
NSA (24)
Data mining (26)
Genomes (35)
Sequenced (9)
LAS (6)
NSA (3)
141 species
LAS (5)
Transcriptomes (106)
Data mining (103)
nLFA (98)

c


PauproSAG2392
PauppetSAG2240
Oedfov SDPC
Oedcar DVYE
Chaorb BAZH
Aphrep IJMT
Urobel RAWF
Urosp. ISGT
Fritub VFIV
Stihel JMUI
Speexs MXDS
Spesim ENAU
Micsp. FOYQ
Gollon BZSH
Chrzof
Monneg
Anksp. OTQG
Peddup JRDV
Botsud VJDZ
Tetdim PZIF
Chlmin MWAN
Hafret FXHG
Carobt RUIF
Carcru VIAU
Ooggig XDLL
Lobseg OFUE
Hetina IRYH
Neocsp. USIX
Vitsp. QWRA
Lobros JKKI
Chlcri BCYF
Chlrei
Chloog IHOI
Gonpec KUJU
Volglo ISPU
Panmor RYJX
Eudele RNAT
Volaur POIR
Volcar
Chlmoe JRGZ
Chlbil SRGS
Chlnoc VALZ
Chlsub GFUR
Chlsp. XOZZ
Chltug UTRE
Chlret ZLBP
Chlros AJUW
Phalen ZIVZ
Ptesp. ACRY
Pteang LNIL
Astgra ZFXU
Dunvir SYJM
Dunter WDWX
Dunsal
Haeplu KFEB
Brasub GUBD
Steplu ZLQE
Chlole PRIQ
Chlmic QRTH
PicspRCC4223
HelspATCC50920
Prowic
Auxpro
AuxproUTEX25
Parkes
Chlvar
Ettole EEJO
Chlsor1602
Miccon24180
Psepri
Gemsp. PFUD
Erevir MNCB
Nanato MFYC
Stlbac WXRI
Pracri WCLV
Mickue EATP
Lepobo ZNUM
Watren
SymirrSAG2036
Botter QYXY
Botbra
CocpriSAG2167
CocsubC169
Ellmar
EllbilSAG24580
ApalobSAG2145
ApafusSAG2523
MyrbisSAG2043
MyrbiaSAG882
Astglo
Trearb NKXU
TrespTZW2008
TrespOTU3
TrespOTU1
TrespOTU5
Pracol
Cocpra XJGM
Pracap XMCL
Psemar JMTE
Pycpro MXEZ
Pyrpar TNAW
Monopi BTFM
Dolten XOAL
Batpra
Ostluc
Osttau
Mansqu QXSZ
MicpusNOUM17
Micpus1545
Picsal TGNL
Nepoli MMKU
Neppyr ISIM
Scosp. EGNB
Pedmin RRSV
Pedtub PUAN
Schdub FMVB
Tetcor DUMA
Tetstr HHXJ
Tetchu HVNO
Cepvir YDCQ
Treann NATT
Codfra GYBH
Bryplu JTIG
Claglo VBLH
Blarhi VHJ
Igntet KADG
Oltvir NSTT
Halmar ALZF
Neosp. GJIY
Acrsp. AJAU
Helpla AJAU
Plalae CBNG
Plater LETF
Ochlsp. CQQP
Ulvend OQON
Perper OAEZ
UlvproYS2018
Ulvmut
Deshal KSFK
Bolpil LSHT
Pirsal NQYP


a
Symbiotic Trebouxiophyceae
Other algae
X-variate 2: 1% expl. var
X-variate 1: 2% expl. var
20
0
-20
0
20
40
b
N0.HOG0014156
N0.HOG0027753
N0.HOG0025597
N0.HOG0027764
N0.HOG0027758
N0.HOG0027757
N0.HOG0027755
N0.HOG0027754
N0.HOG0025594
N0.HOG0025576
N0.HOG0025575
N0.HOG0015395
N0.HOG0016616
N0.HOG0019047
N0.HOG0016349
N0.HOG0020229
N0.HOG0021106
N0.HOG0020108
N0.HOG0003278
N0.HOG0015018
N0.HOG0022476
N0.HOG0022707
N0.HOG0022706
N0.HOG0022698
N0.HOG0021550
N0.HOG0002604
N0.HOG0014437
N0.HOG0015882
N0.HOG0019914
N0.HOG0011794
N0.HOG0003046
N0.HOG0002991
N0.HOG0032055
N0.HOG0032053
N0.HOG0032052
N0.HOG0032048
N0.HOG0032047
N0.HOG0032043
N0.HOG0031856
N0.HOG0029052
N0.HOG0029051
N0.HOG0029050
N0.HOG0029049
N0.HOG0029047
N0.HOG0029046
N0.HOG0029043
N0.HOG0029040
N0.HOG0029039
N0.HOG0029038
N0.HOG0029037
N0.HOG0029035
N0.HOG0029033
N0.HOG0029031
N0.HOG0029030
N0.HOG0029029
N0.HOG0029027
N0.HOG0029025
N0.HOG0029021
N0.HOG0029019
N0.HOG0012381
N0.HOG0024269
N0.HOG0023304
N0.HOG0023289
N0.HOG0019143
N0.HOG0019066
N0.HOG0019063
N0.HOG0012501
N0.HOG0018191
N0.HOG0021549
N0.HOG0020223
N0.HOG0020227
N0.HOG0020217
N0.HOG0019134
N0.HOG0019133
N0.HOG0023306
N0.HOG0023301
N0.HOG0023297
N0.HOG0023291
N0.HOG0016297
N0.HOG0012366
N0.HOG0019146
N0.HOG0015011
N0.HOG0027643
N0.HOG0026906
N0.HOG0025596
N0.HOG0025595
N0.HOG0025593
N0.HOG0025592
N0.HOG0025588
N0.HOG0025587
N0.HOG0025580
N0.HOG0003375
N0.HOG0002085
N0.HOG0010766
N0.HOG0011168
N0.HOG0018185
N0.HOG0014727
N0.HOG0016351
N0.HOG0013900
N0.HOG0012965
0.2
0.15
0.1
0.5
0
sPLS-DA contribution
c
-log10(wilcoxon pvalue)
20
15
10
5
0
5
10
15
20
-log10(wilcoxon pvalue)


Tree scale: 0.1
Prasinodermophyceae
Chlorophytes s.l.
Pedinophyceae
Chlorodendrophyceae
Trebouxiophyceae
Ulvophyceae
Chlorophyceae
Differentially expressed genes in Trebouxia sp (TZW2008)
Up
Up and down
Down
Nb of copies per orthogroups
0
1
2
4
5
7
8
9
11
12
14
N0.HOG0014156
N0.HOG0027753
N0.HOG0027758
N0.HOG0025594
N0.HOG0016616
SAP domain
N0.HOG0021106
N0.HOG0022707
N0.HOG0022706
N0.HOG0022698
N0.HOG0019914
Zinc finger
N0.HOG0032055
N0.HOG0029052
N0.HOG0029051
Galactose-binding domain-like
N0.HOG0029047
N0.HOG0029039
N0.HOG0029037
N0.HOG0029035
N0.HOG0029033
PIN domain
N0.HOG0029027
N0.HOG0023304
N0.HOG0012501
Glutathione S-transferase
N0.HOG0021549
N0.HOG0020223
N0.HOG0020217
N0.HOG0023297
BAG domain
N0.HOG0012366
N0.HOG0026906
N0.HOG0025596
N0.HOG0025580
N0.HOG0002085
Short-chain dehydrogenase/reductase
N0.HOG0013900
N0.HOG0012965
Glycoside hydrolase, family 8
0
0.5
0.1
0.15
0.2
sPLS-DA contribution


a
Tree scale: 1
Bacteria
Algae
Fungi
Bacteria
b
MyrbisSAG2043_Scf2
CocpriSAG2167_Scf1
0bp
2Mb
4Mb
6Mb
8Mb
10Mb
Length (bp)
Astglo_Scf00133
CocpriSAG2167_Scf12
CocsubC169_AGSI01000011.1
CocsubC169_AGSI01000024.1
EllbilSAG24580_Scf39
EllbilSAG24580_Scf52
MyrbisSAG2043_Scf17
SymirrSAG2036_Scf10
TrespOTU1_Scf2
TrespOTU1_Scf34
TrespOTU3_Scf1
TrespOTU3_Scf27
TrespOTU5_Scf11
TrespOTU5_Scf18
TrespTZW2008_Scf184
TrespTZW2008_Scf46
0bp
500kb
1Mb
1.5Mb
2Mb
2.5Mb
3Mb
Length (bp)
c
nC
Lichenan + GH8 48h
Lichenan + GH8 24h
Lichenan
min
nC
DP1
DP2
DP3
DP4
DP5
DP6
Standards
Cellulose + GH8
Control (no enzyme)
min
Absorbance at 590nm (u.a.)
Chitosan
Chitosan_GH8-L1
0.111
0.114
0.051
0.052
NaAc pH5.2
NaP pH7