Phylogenomics reveals the evolutionary origin of lichenization in chlorophyte algae

Camille Puginier[1], Cyril Libourel[1], Jürgen 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*]

[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

*To whom correspondence may be addressed: pierre-marc.delaux@cnrs.fr; francesco.dalgrande@unipd.it; jean.keller@univ-tlse3.fr

Author Contributions:

FDG, P-M.D, PS and JK designed the project

JK, FDG, CP, CL, PS, MP, MH, SG and JO conducted experiments

JK, FGD, PJ, CP, CL, JGB and PMD analyzed data.

P-M.D, CP and JK wrote the manuscript with input from all authors.

Competing Interest Statement: The authors declare no competing interests.

Classification: Plant Biology, Evolution

Keywords: Lichens, symbiosis, evolution, phylogenomics, green algae

Highlights

- We sequenced 12 lichen-forming and non-lichen-forming chlorophytes
- Lichens evolved independently in Ulvophyceae and Trebouxiophyceae
- Gene family expansion correlate with the ability to form lichens
- Horizontal Gene Transfer of a glycoside hydrolase is linked with lichenization

Summary

Mutualistic symbioses, such as lichens formed between fungi and green algae or cyanobacteria, have contributed to major transitions in the evolution of life and are at the center of extant ecosystems. However, our understanding of their evolution and function remains elusive in most cases. Here, we investigated the evolutionary history and the molecular innovations at the origin of lichens in green algae. We *de novo* sequenced the genomes or transcriptomes of 12 lichen-forming and closely related non-lichen-forming algae and performed comparative phylogenomics with 22 genomes previously generated. We identified 39 functional categories and 11 orthogroups significantly enriched in lichen-forming chlorophyte green algae. Among them, functions such as carbohydrates metabolism and stress tolerance could be related to the pioneer characteristics of lichens. We detected lichen-specific expansions of glycosyde hydrolase gene families known to remodel cell walls, including the glycosyde hydrolase 8 (GH8) which encodes a glucanase and was acquired in lichenizing Trebouxiophyceae by horizontal gene transfer from bacteria, concomitantly with the ability to form lichens. Mining genome-wide orthogroups, we found additional evidence supporting at least two independent origins of lichen-forming ability in chlorophyte green 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 was enabled by the arbuscular mycorrhizal symbiosis formed with Glomeromycota fungi [1,3], leading to massive geochemical changes [4]. Although the association with Glomeromycota fungi is at the base of most of the terrestrial plant diversity which is observed in embryophytes (*i.e.* land plants), association between green algae and fungi to form lichens resulted in a second event of terrestrialization. Lichens are symbiotic organisms composed of a fungal partner (usually from the Ascomycota phylum, more rarely from Basidiomycota) and a photosynthetic partner, the photobiont, which is a chlorophyte in 90% of the known lichens [5]. Lichens exhibit diverse shape and are adapted to a wide range of ecosystems, from temperate to extreme cold and arid environments [6]. Lichens are also key players in ecosystems dynamics as they are pioneer species and provide food and shelter to a wide range of animal and microbial species [7]. Despite their major ecological importance, little is known about lichen evolution. The fossil record is rich in lichen-like structures ([8] for review), however consortia of filamentous and photosynthetic organisms may have been wrongly interpreted as lichens [9,10]. The most ancient convincing fossils of lichens date from the Early Devonian, an origin supported by dated phylogenies of the fungal and green algal partners [9,11]. The comparison of genomes in a defined phylogenetic context (comparative phylogenomics) has successfully unraveled the evolution of plant mutualistic symbioses with complex evolutionary histories [12–14]. In addition, such approaches have the potential to shed light into the molecular mechanisms associated with major innovations, including symbioses [15–18]. In the context of lichens, these approaches have been exclusively conducted with a 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 [9]. Lichen-Forming Algae (LFA) are almost exclusively found in two of the 13 classes of chlorophyte algae, the Ulvophyceae and Trebouxiophyceae [6]. Such distribution of lichen-forming ability might be the result of either a single gain in the common ancestor of Ulvophyceae and Trebouxiophyceae followed by multiple losses, similar to other terrestrial endosymbioses [1,12,13,16], or multiple independent (convergent) gains. Recently, studies based on ancestral trait reconstruction and time-calibrated phylogenetic approaches provided strong support to the convergent evolution of lichen-forming fungi and suggested a similar pattern for LFA [9,11,19]. However, the limited availability of LFA genomes constrained so far molecular analyses to single algal species such as *Asterochloris glomerata* and *Trebouxia sp.* TZW2008[20,21]. Thus, the evolutionary history of lichens and the molecular mechanisms linked to the ability to form lichens in green algae remain elusive. In this study, we sequenced the genomes of eight LFA and four non-Lichen-Forming Algae (nLFA) and used comparative phylogenomics to discover molecular processes in algae associated with the ability to lichenize. Projecting these molecular features on the chlorophyte algae phylogeny, we propose an evolutionary scenario for the evolution of lichens, involving the expansion of gene families and horizontal gene transfers that facilitated the physical interactions between the symbiotic partners.

Results

General genome characteristics do not differ between lichen-forming and non-lichen-forming chlorophyte algae.

Using the PacBio long-read platform we sequenced the genomes of six LFA belonging to the Trebouxiales, Botryococcus- and Apatococcus-clades (**Figure 1, Table S1**). In parallel, we sequenced three closely related nLFA including species from the *Symbiochloris* and *Myrmecia* genera for which no genomes were available previously (**Figure 1, Table S1**). Assemblies for eight of the nine species displayed high quality statistics with an average scaffold N50 of almost 2Mb, and an average of only 143 scaffolds (**Figure 1, Table S2**). The ninth assembly (*Apatococcus fuscidae* SAG2523) displayed a scaffold N50 of 50 kb and a higher number of 2,319 scaffolds (**Figure 1, Table S2)**. To complete this dataset, the transcriptome of six additional species, including the LFA Ulvophyceae *Dilabifilum* (now *Paulbroadya) petersii*, were sequenced on an Illumina NovaSeq platform yielding an average of 40 million reads (details of sequencing and assembly statistics in **Table S3**). Finally, we downloaded publicly available genome assemblies for 22 species covering other classes of chlorophyte green algae, and re-annotated six of them (**Table S1**). Completion of the predicted proteomes estimated for all species using BUSCO ranged from 18.3% to 99.6% (**Figure 1, Figure S1**). To ensure the reliability of the comparative analyses, a threshold of 65% completeness and less than 30% of missing BUSCO orthologs was set up, leading to a total of 38 green algal species covering the four main classes of the chlorophytes and including 10 LFA species (**Figure 1**, nine Trebouxiophyceae and one Ulvophyceae).

We searched for genomic signatures associated with the ability to form lichens, starting with general genomic parameters which have been associated with transition to symbiotic states [22–25.] We compared

LFA and nLFA genome size, GC content and coding sequence number(**Figure 1**). A significant difference was observed for the GC content of coding sequences which was lower in LFA compared to nLFA (**Figure S2**). Careful examination showed that the lower GC content in LFA is due to the *Trebouxia* species (53.19% GC in average) and does not correlate with the lichenization lifestyle **(Figure S2).** We did not detect significant differences between LFA and nLFA in the other parameters **(Figure S2)**. Instead, these comparisons highlighted a lineage-specific genome expansion in the Chlorophyceae class, with significantly larger genomes and higher number of genes compared to other classes **(Figure S2)**. Transposable elements (TEs) are known drivers of trait evolution [26]. We identified the TE repertoires of each algal species and compared them between LFA and nLFA species. We observed a very significant variation of overall TEs percentages among species which does not correlate with the ability to form lichens. At the level of TEs subfamilies, we identified two (CACTA and PIF/HARBINGER) of the 13 investigated significantly different between LFA and nLFA. However, it seems that it could be due to a confounding phylogenetic effect and further investigation are needed (**Figure S2 and S3, Table S4)**. Altogether this indicates that general genomic features are similar between LFA and nLFA.

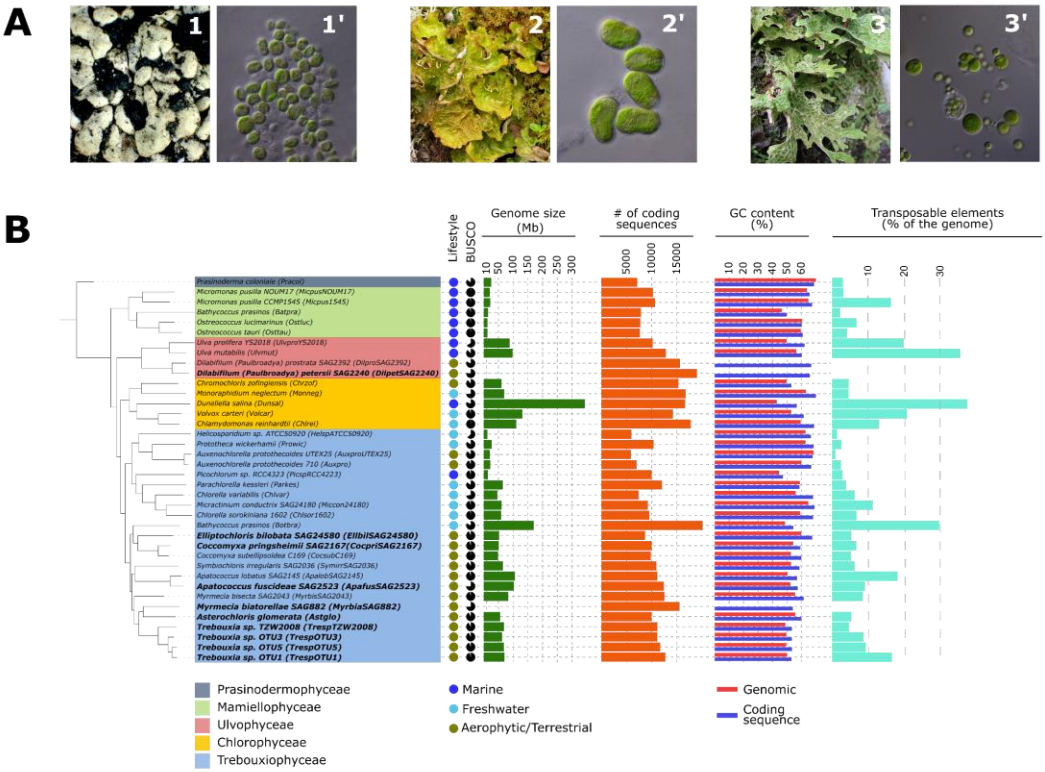


Figure 1. Lichens sampling, distribution of habitats and genomic parameters.
(A) Pictures of three lichens and their algal partner: *Catolechia wahlenbergii* (1) and *Elliptochloris bilobata* SAG245.80 (1'), *Lobaria linita* (2) and *Myrmecia biatorellae* SAG8.82 (2'), *Lobaria pulmonaria* (3) and *Symbiochloris irregularis* SAG2036 (3').
(B) From left to right: phylogenetic tree reconstructed based on orthogroups (tree is rooted on *Prasinoderma coloniale*), lifestyle of the different algae species investigated, BUSCO completion of proteomes (full circle

indicates 100% completeness level), size of the genome in Mb (empty rows for species with only transcriptomic data), number of protein-coding genes, GC content for protein-coding genes (blue bars) and genomes (red bars), the percentage of genomes covered by transposable elements (LRT, non LTR, TIR and repeat regions). Lichen-forming species are indicated in bold

Phylogenomics support at least two independent gains of the lichen-forming ability in chlorophyte algae.

To trace the evolution of lichen-forming capacity, we conducted two complementary statistical approaches. On one hand, Wilcoxon and Fisher tests were used to identify functional terms (IPR) and orthogroups significantly more present in LFA compared two nLFA. On the other hand, we applied a sparse Partial Least Square Discriminant Analysis (sPLSDA) to identify the IPR and orthogroups that most discriminate LFA and nLFA (**Figure S4**). Then, we compared the results from the two complementary approaches and retained exclusively the IPR and orthogroups found in both. We then cross-referenced these two lists with the genes of *Trebouxia sp.* TZW2008 which are differentially regulated during the interaction with lichen-forming fungi compared to axenic cultures. Sparse PLSDA on orthogroups and IPR functional terms placed LFA and nLFA species in separated clusters. Within the LFA cluster, the Ulvophyceae species *Dilabifilum (*now *Paulbroadya*) *petersii* is divergent from all Trebouxiophyceae LFA species (**Figure S5**). In addition, we looked for orthogroups only composed of LFA species, which would reflect a single origin of the ability to form lichens followed by multiple losses and found only one (corresponding to the rhodopsin functional category). Taken together, these results support the prevailing hypothesis of the convergent evolution in chlorophytes for the ability to engage in lichens, with at least two independent gains.

Gene family expansion in LFA are associated with stress tolerance and carbohydrate metabolism

Besides informing the evolutionary history of the ability to form lichen, the phylogenomic data offer the opportunity to identify the molecular mechanisms associated with these gains. Through the combination of sPLSDA, Wilcoxon/Fisher tests and cross-referencing with up-regulated genes of *Trebouxia sp* TZW2008 during lichen formation (**Table S6**), we identify 39 IPR (20 expansions/contractions and 19 specific to LFA) and 11 orthogroups (one expansion, 10 mainly specific to LFA) representing candidates associated with the lichen formation ability (**Table S5**). Among them, 9 IPR are either specific to or expanded in LFA and are associated with the adaptation to stress **(Figure 2)**. Indeed, four catalase-related domains (highlighted in purple in **Table S5**) that are antioxidant enzymes involved in the protection of cells against oxidative stress [27] are expanded in LFA (between 1.7 to 2 times more present in LFA). Furthermore, two TspO/MBR domains (highlighted in pink in **Table S5**) which were identified with a two-fold increase in abundance in LFA compared to nLFA, have been found induced in drought stress condition as well as essential for adaptation to salt stress in the model moss *Physcomitrium patens* [28]. Finally, three aquaporins-related domains (**Figure 3C**; highlighted in purple in **Table S5**) are between 1.8 and 2.4 times more present in LFA compared to nLFA (**Figure 2**). Aquaporins were previously identified as important actors of the lichen symbiosis since they are upregulated during dehydration state in *Trebouxia gelatinosa*, a lichen photobiont

[29]. Such expansions in functions related to stress tolerance could be linked to the adaptation of lichens to extreme environments.

We also identified candidates linked with the carbohydrate metabolism **(Figure 2)**. An orthogroup that contains genes annotated as a short-chain dehydrogenase/reductase (SDR) (in average 8.5 times more present in LFA) stands out among the candidates and is responsible for 0.32% of the total 6.1% explained variance in sPLSDA (**Figure 2; Figure 3A; Figure S4;** highlighted in orange in **Table S5**). SDR belongs to a family containing enzymes involved in ribitol biosynthesis. Ribitol is a linear pentose alcohol previously identified as the major sugar produced in lichens such as *Peltigera aphtosa* (*Coccomyxa* photobiont), *Xanthoria aureola* or *Fulgensia bracteata* (*Trebouxia* spp. photobionts) [30,31]. The addition of exogenous ribitol to culture of lichen-forming fungi is known to stimulate fungal growth and developmental transitions and has been suggested as a signal for lichen initiation [31,32]. The enrichment in SDR across LFA suggests that the expanded ability to produce ribitol is not a species-specific feature but is rather linked with the ability to lichenize itself. Along with the SDR, we identified one glycoside hydrolase orthogroup (**Figure 2; Figure 3B**; highlighted in brown in **Table S5**) that is specific to LFA, and one functional domain that corresponds to a glycoside hydrolase 8 (GH8) mainly present among the LFA from the class Trebouxiophyceae (**Figure 2**). Glycoside hydrolases are known to hydrolase a large diversity of polysaccharides including those forming the plant and fungal cell walls [33]. Cell wall remodeling occurs in all symbioses between plants and microorganisms, facilitating the progression of the microbial partner in the plant tissues in the case of endosymbiosis [34,35], and is observed at the contact sites between symbionts in the case of lichens [36]. The occurrence of an expanded and specific repertoire of glycoside hydrolases in LFA echoes the peculiar patterns of retention and losses of cell wall degrading enzymes linked to symbiotic abilities in fungi [15,18,37,38]. Considering the absence of specificities in the glycoside hydrolases repertoire from lichen-forming fungi [39], this indicates that LFA may directly contribute to the remodeling of the fungal partners' cell wall during lichenization. Finally, four IPR and one orthogroup corresponding to various nitrile hydratase (NH) functional categories were identified as highly specific of LFA (**Figure 2, Figure 3D**). However, the putative function of these proteins in the lichenization process remains elusive.

The specific presence or expansion of given genes in LFA might result from various evolutionary processes, including gene duplication or horizontal gene transfer [26]. We inferred phylogenies of each candidate to investigate their evolutionary history. These phylogenies revealed that 56% of the IPR functional categories and 55% of the orthogroups were expanded or associated with the advent of lichen formation ability within the Trebouxiophyceae class (**Table S7**). Further phylogenetic analysis of the new candidates revealed that most of them are present in all chlorophyte algae but the sister clade of the Trebouxiophycean (**Figure S6**). Sequencing LFA species in that clade would allow confirming whether these candidates are linked to the lichen forming capacity in Trebouxiophycean.

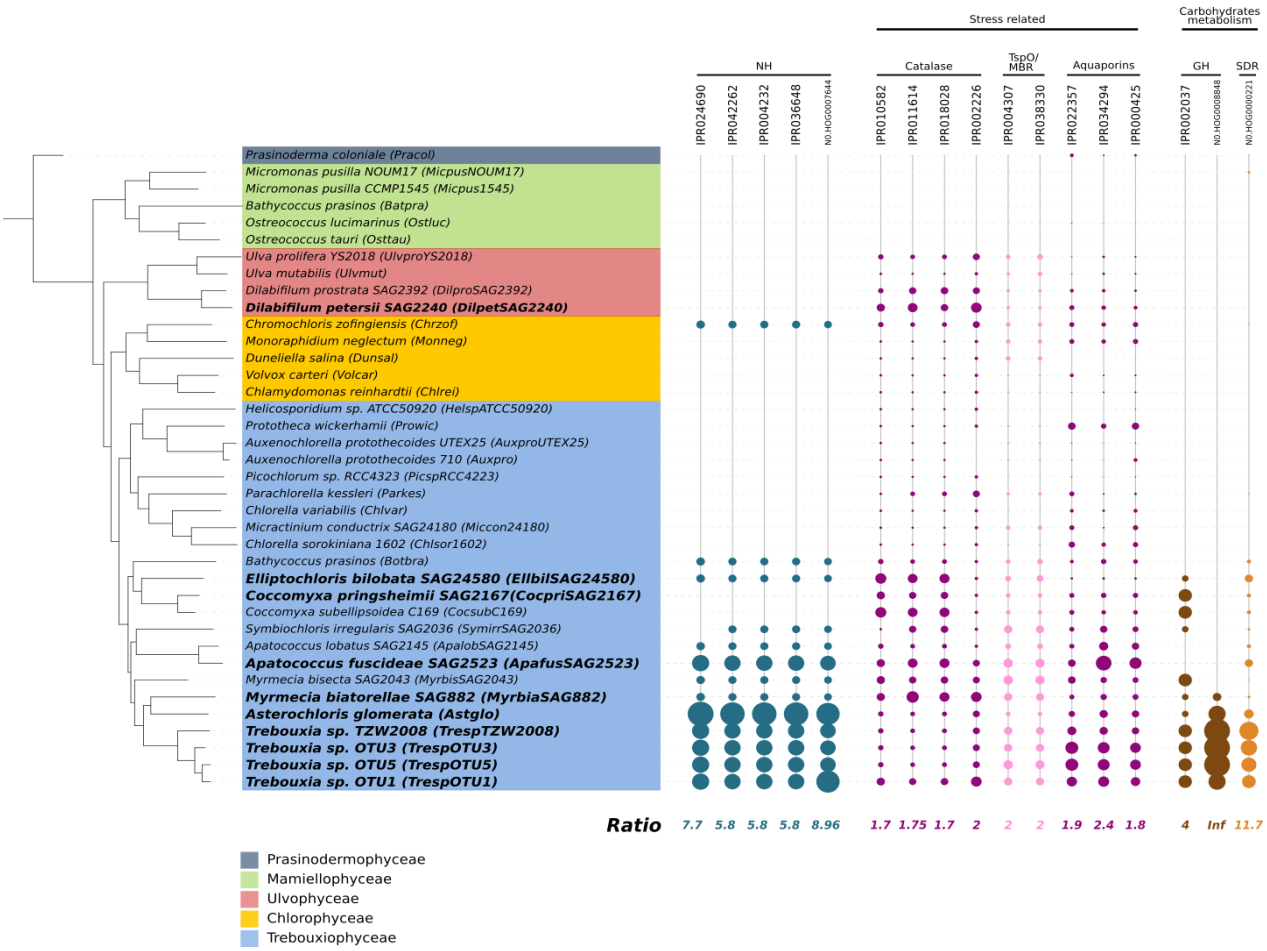


Figure 2. Mapping of functional family and orthogroups significantly more present in lichen-forming algae compared to non-lichen-forming algae.

Functional families (IPR) and orthogroups related to stress tolerance (Catalases, TspO/MBR, Aquaporins), carbohydrates metabolism (glycoside hydrolase, short-chain dehydrogenase) and nitrile hydratases are indicated in pink/purple, brown/orange and blue respectively. The circles correspond to the number of IPR or to number of genes in orthogroups for each species. Size of the circles is normalized for each column and are not relative between columns. The numbers below each column correspond to the ratio of the numbers of IPR or orthologs between lichen-forming algae and non-lichen-forming algae. Lichen-forming species are indicated in bold.

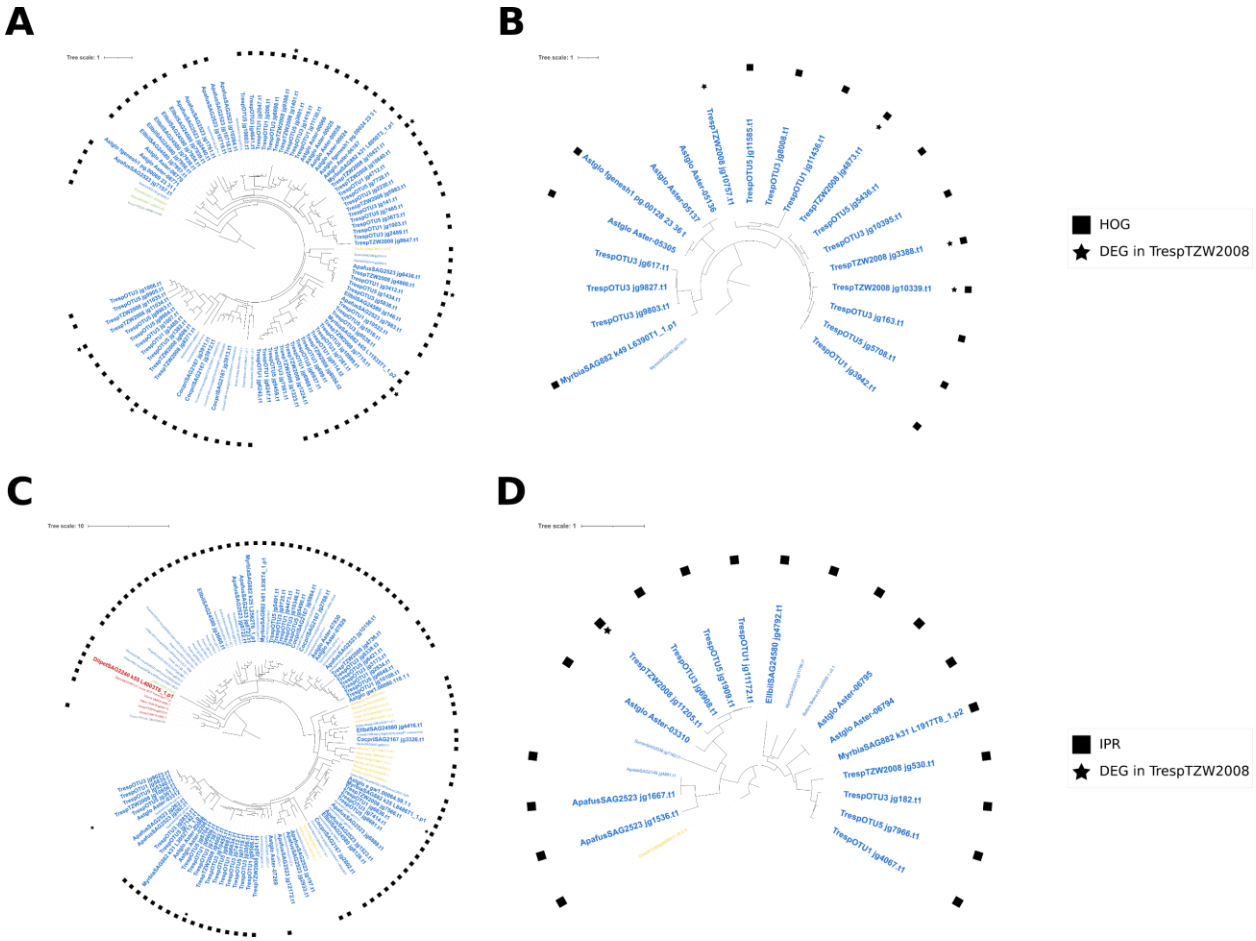


Figure 3: Phylogenetic trees of carbohydrates metabolism and stress related functional families and orthogroups.

(A-B) Phylogenetic trees of carbohydrates metabolism related orthogroups. (A) Maximum likelihood phylogenetic tree (WAG+R6) of orthogroup N0.HOG0000221 with genes belonging to the Short-chain Dehydrogenase/Reductase (SDR) family, HAD family or NADP-dependent oxidoreductase. The tree is rooted on *Prasinoderma coloniale* (Prasinodermophcyeae). (B) Maximum likelihood phylogenetic tree of orthogroup N0.HOG0008848 with genes belonging either to the glycosyde hydrolase superfamily or to the Alpha/Beta hydrolase family. The tree is rooted on *Myrmecia bisecta* SAG2043. (C) Maximum likelihood phylogenetic tree of stress related domain IPR022357 (Major intrinsic protein, aquaporin) rooted on *Prasinoderma coloniale* (Prasinodermophcyeae). (D) Maximum likelihood phylogenetic tree of IPR024690 (Nitrile hydratase beta subunit domain) rooted on *Chromochloris zofingiensis* (Chlorophyceae).
On each tree, the lichen-forming species are indicated in bold. Black squares mark the proteins from the orthogroup as reconstructed by OrthoFinder for tree concerning orthogroups while they indicate the presence of the functional domain (IPR) on trees on functional categories. Black stars mean that the gene is up-regulated during the lichen formation of Trebouxia sp. TZW2008. Trebouxiophyceae, Chlorophyceae, Ulvophyceae, Mamiellophyceae and Prasinodermophcyeae are showed in blue, yellow, red, green and grey respectively. All trees were rooted on the most distant species according to the species trees obtained in OrthoFinder. Bootstrap values are available in the newick files in Supplementary Data S1.

Horizontal gene transfer shaped the evolution of lichens

Horizontal gene transfer is a key driver for the acquisition of functional innovations across living organisms [26]. Among the candidate genes identified through the phylogenomic comparisons and the gene family-

specific phylogenetic analysis, we identified six genes that could derive from HGT events. Using the non-redundant database from NCBI we reconstructed the evolution of all candidates and found that 5 of them had an uncertain evolutionary pattern while one, the GH8, originates from HGT. According to the CAZy classification, all GH8 members have been found in bacteria [40]. Phylogenetic analysis suggests a bacterial origin of the GH8 that have been transferred mostly in Trebouxiophyceae and in one nLFA Ulvophyceae species (**Figure 4A**). Within the Trebouxiophyceae, GH8 are only present in LFA and close relative nLFA algae (**Figure S7**). Additionally, we identified a putative HGT to fungi, mostly from the non-lichen forming mucoromycotina phylum, but the direction of the transfer remains elusive (**Table S8**). HGT was also supported by the gene position which is well anchored in the scaffolds and by mapping the raw reads which excluded the possibility of chimeric scaffolds (**Figure 4B; Figure S7C**). According to the CAZy database, members of the GH8 family are able to cleave either fungal polymers such as chitosan or β-1,3/1,4-glucans, known as lichenans and found in lichens [41,42]. The GH8 family has been divided into three subfamilies based on the position of the proton acceptor residue [42,43]. Alignment of reference proteins from the three GH8 subfamilies with the LFA GH8 sequences revealed that they shared the asparagine catalytic site with the GH8b subfamily (**Figure S7D**) known to encompass, among others, lichenase enzymes able to degrade lichenan. Additionally, the 3D model of the LFA GH8 generated using AlphaFold2 [44] positioned the catalytic asparagine in a pouch where the substrate could bind (**Figure S8**). To functionally test whether the LFA GH8 can degrade lichenan, we cloned orthologs from three species and successfully produced the recombinant GH8 from *Asterochloris glomerata*. Enzyme assays towards different polysaccharides confirmed a typical lichenase activity [45] for this enzyme with a significant cleavage of mixed linked β-1,3/β-1,4 glucans, while no cleavage was observed on cellulose nor chitosan substrates **(Figure 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. Altogether, these results indicate that the ability to form lichens in Trebouxiophyceae is linked with the HGT of GH8 from bacteria, which could be involved in cell-wall remodeling during the interaction.

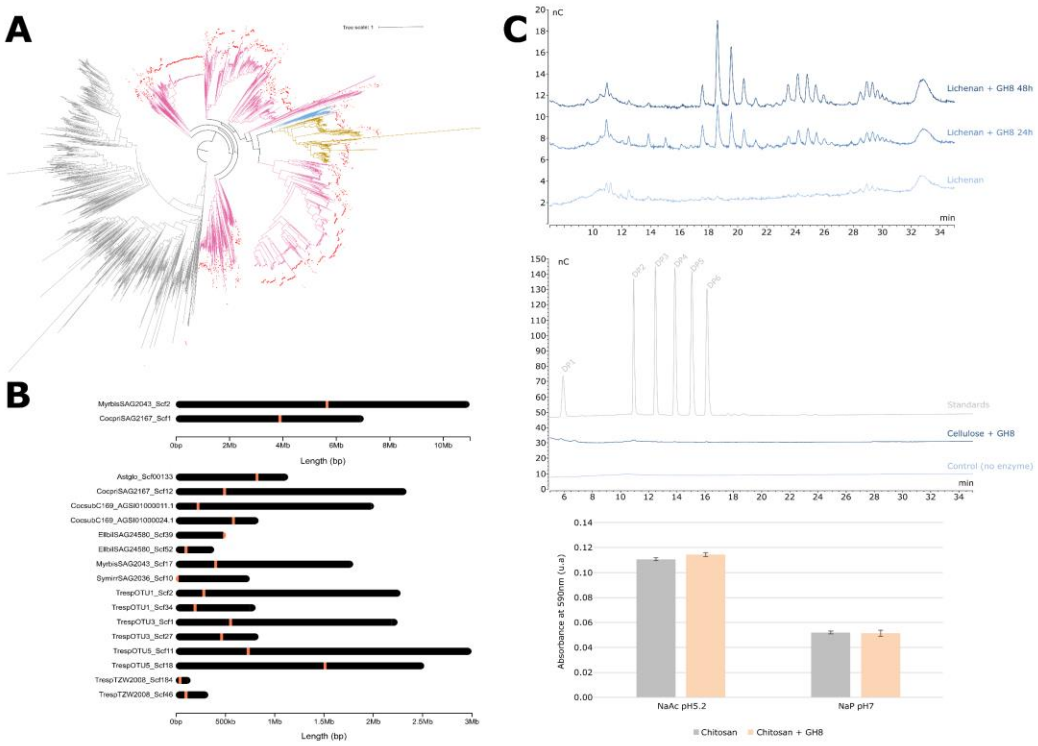


Figure 4: Evolution of GH8 lichenases in algae.
(A) Maximum Likelihood tree of the GH8 proteins. Branches are colored as follow: outgroup in grey, bacteria in pink, fungi in brown and chlorophytes algae in blue. Red dots indicate the presence of the GH8 functional domain. Tree is rooted on distant GH8 homologs. (B) GH8-like anchoring in scaffold of the different chlorophytes algae. The GH8 position is indicated by the green line. (C) Enzymatic activity of the GH8 from *A. glomerata* on, from top to bottom: lichenan, cellulose and chitosan.

Discussion

Lichens evolution has been unsymmetrically studied from the fungal partner perspective. These studies have revealed the convergent evolution of lichens ability resulting in over 20,000 fungal species able to form this symbiosis [9,46]. Recently, a detailed phylogenomic comparison of Lecanoromycetes revealed that lichenization capacity is the result of complex loss and re-acquisition processes driven not only by host–symbiont specificity but also by multiple environmental and biological causes [9]. The dynamic evolutionary pattern of lichenization on the fungal side is associated with interactions with diverse photobionts, including cyanobacteria and chlorophyte green algae and possibly other eukaryotic symbionts [47,48]. The chlorophyte green algae encompass more than 8,650 species divided in 13 classes [49]. LFA have been described in two of these classes, the Trebouxiophyceae and the Ulvophyceae (AlgaeBase, accessed 08/12/2022). The scattered distribution of the lichenization habit observed in these classes suggests either that this ability evolved independently multiple times in both classes or originated in their most recent common ancestor and was subsequently lost many times, the former hypothesis being currently favored. Here, using the phylogenetic pattern of genome-wide orthogroups and genes associated with lichenization

as proxy, we did not find evidence for an ancestral gain of the lichenizing ability in chlorophyte green algae. By contrast, these analyses supported at least two gains, one in the Ulvophyceae, and one either at the base of the Trebouxiophyceae or at the base of the core trebouxiophycean lineage which is consistent with previous ancestral trait reconstruction and time-calibrated approaches [11,48,50]. The exact nodes at which the independent gains occurred need the sequencing and the analysis of more LFA genomes, especially within the Ulvophyceae family and the Trebouxiophycean sister clade.

Here, comparing a large diversity of LFA and close relative nLFA species we can propose molecular mechanisms associated with the evolution of lichen formation (Fig. 4). We discovered shared expansions of gene families in LFA, and LFA-specific genes. Among them, we identified a glycosyde hydrolase which is possibly involved in cell wall remodeling process, a critical step of plant–microbe interactions [34]. In its forced lichen-like association with the non-lichen-forming fungus *Aspergillus nidulans,* the nLFA *Chlamydomonas reinhardtii* exhibits thinner cell wall at the points of contact with the fungus, suggesting that an inherent capacity to generate an interface may exist in chlorophyte green algae [51]. The evolution of lichenization would thus build by enhancing this ability, and the gain of GH8 via horizontal gene transfer represents such improvement. Further biochemical characterization of the GH8 from *A glomerata* revealed its ability to degrade lichenan which supports its important role in lichen formation. The gain of Cell-Wall degrading activity seems to be a common theme in the evolution of symbiotic abilities, as exemplified by a recent study aiming at identifying the genetic basis of endophytism in the *Arabidopsis thaliana* microbiome which detected a GH10 [18], or the gain of GH2 and GH5 linked with Ectomycorrhizae in the Basidiomycete *Laccaria bicolor* [52,53]. Along with the GH8, we also identified the expansions of SDR gene family putatively involved in the biosynthesis of ribitol. Ribitol has been demonstrated to impact fungal growth in various lichens and is thought to act as a signal to initiate lichen formation [31], and its biosynthesis has been found regulated in the lichens *Cladonia grayi* and *Usnea hakonensis* [20,21]. Additional expanded functional categories likely associated with the lichen formation ability encompass different functions related to stress tolerance such as desiccation. Such expansion is in line with the ability of lichen to develop in extreme condition as well as with the terrestrial lifestyle of the LFA species.

Most of the expansion or specific functional categorizes and orthogroups associated with LFA were identified in the Trebouxiophyceae class. This could come from a lack of available genomic resources in Ulvophyceae and be investigated through the sequencing and analysis of more LFA genomes from this understudied class. However, this also supports the hypothesis of a convergent evolution in these two classes. In Trebouxiophyceae, almost all the candidate orthogroups and functional categories were still present, although contracted, in nLFA species whose are sister species of a LFA (for example, *C. subellipsoidea C-169, nLFA* and sister species of *C. pringsheimii*). This result strongly suggests either an unknown remnant capacity for this nLFA species to form lichen or a recent loss with a partial conservation of the genetic toolkit enabling the lichenization.

Besides providing a general framework for the evolution of lichenization in chlorophytes, the comparative analysis presented here identified biological processes and actual genes that may be considered as the *lichenization toolkit* similar to the *symbiotic toolkit* previously proposed for embryophytes [54]. To ascertain the role of the identified genes it will be now essential to develop genetically tractable models in the Trebouxiophyceae for loss of function approaches (for recent advances in Trebouxiophyceae transformation see: [55]) or high-throughput gain of function in well-established chlorophyte models, such as *Chlamydomonas reinhardtii.*

Methods

Algal cultures

Three *Trebouxia* photobionts with varied ecologies [56] were isolated from thalli of the lichen *Umbilicaria pustulata* using a micro-manipulator as described in [57]. Cultures were grown on solid 3N BBM + V medium (Bold's Basal Medium with vitamins and triple nitrate [58] 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 [56]. 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 form symbiotic associations with fungi - belonging to the classes Trebouxiophyceae and Ulvophyceae (Chlorophyta) (*SI Appendix,* Table S1). 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 [59]. 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 2500 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 5000 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 [60] with modifications described in [61] or with Qiagen Genomic-Tips.

Long-read DNA sequencing

SMRTbell libraries were constructed for samples passing quality control (*SI Appendix,* Table S2) according to the manufacturer's instructions of the SMRTbell Express Prep kit v2.0 following the Low DNA Input Protocol (Pacific Biosciences, Menlo Park, CA) as described in [61]. 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) (*SI Appendix,* Table S2). 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 (*SI Appendix,* Table S3) 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) [62] in CLR mode and default parameters. Each assembly was polished once as 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 [63] 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 [64], with an e-value threshold of 1E-10 and the MEGAN-LR algorithm [64]. Only scaffolds assigned to the Chlorophyta were retained for subsequent analysis.

Genome and transcriptome annotations

To investigate evolution of the lichen symbiosis at a large evolutionary scale, we additionally annotated six published green algal genomes without structural annotations (*SI Appendix,* Table S1).

Genome assemblies were softmasked using Red [65] and annotated using BRAKER2 pipeline [66–74]. 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 (last column of *SI Appendix,* Table S1).

For 14 species, available or generated RNA-Seq data were used as hints in BRAKER2 (*SI Appendix,* Table S1). Adapters and low-quality sequences were removed from the raw fastq files using cutadapt v2.1, [75] 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 [76] 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 [77]. These final alignments data were used as input in BRAKER2.

We also annotated transcriptomes of four species. First, we assembled the transcriptomes from the raw reads RNAseq using DRAP v1.92 pipeline [78]. runDrap was first used on the unique samples applying the Oases RNAseq assembly software [79]. For *Trebouxia gelatinosa*, runMeta was used to merge, without redundancy, the 3 transcripts assemblies predicted based on a fpkm of 1. Predictions of protein-coding genes were performed using TransDecoder v5.5.0 (https://github.com/TransDecoder/TransDecoder, [80])

and hits from BLASTp search in the Swissprot database (downloaded on September 2021) as well as HMMER search in the Pfam v34 database [81,82]. Completeness of newly sequenced and annotated genomes and transcriptomes was assessed using BUSCO V4.1.4 [83] with default parameters and using the Chlorophyta "odb10" database (1519 core genes) as reference.

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.

Orthogroups reconstruction

Orthogroups reconstruction was performed using OrthoFinder v2.5.2 [85] with DIAMOND set in ultra-sensitive mode. The estimated 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 the MSA option to improve the orthogroups reconstruction (used orthogroups are available in Table S9).

Genome streamlining investigation

Multiple symbionts exhibit strong genome reduction and modifications [22–25]. We compared genome size, number of protein coding genes, the GC content and the transposable elements repertoire of LFA and nLFA species. Each comparison has been performed using a Wilcoxon test using R v4.1.1 [86].

Transposable element annotation

We used EDTA v2.0.1 [87] 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 [88], LTR_retriever v2.6 [89]; terminal inverted repeat transposons (TIRs): TIR-Learner v1.19 [90]; Helitrons: HelitronScanner v1.1 [91]; terminal direct repeats (TDRs), miniature inverted transposable elements (MITEs), TIRs, LTRs: Generic Repeat Finder v1.0 [92]. It also runs RepeatModeler v2.0.3 [93] 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 [94] to de-duplicate and classify consensus sequences and finally employs RepeatMasker v4.1.2-p1 [95] 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 differentially expressed genes between the symbiotic state of Trebouxia sp TZW2008

In order to increase our statistical power by crossing deregulated genes in symbiotic context we used the transcriptomic data from *Trebouxia sp* TZW2008 [21] with results of our contraction/expansion analysis. To do so, the raw reads were downloaded (*SI Appendix,* Table S1) and submitted to the nf-core/rnaseq v3.4 [96] workflow in nextflow v21.04 [97] using *-profile debug,genotoul --skip_qc --aligner star_salmon* options. Nextflow nf-core rnaseq workflow used *bedtools* v2.30.0 [98], *cutadapt* v3.4 [75] implemented in *TrimGalore!* v 0.6.7, *picard* v2.25.7 (https://broadinstitute.github.io/picard), *salmon* v1.5.2 [99], *samtools* v1.13 [69], *star* v2.6.1d and v2.7.6a [100], *stringtie* v2.1.7 [101] and *UCSC tools* v377 [102].

The counted data were analysed using *edgeR* package v3.36.0 [103] with R v4.1.1 [86]. Two samples of synthetic lichen shown distant clustering to other synthetic lichens 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 [104]. 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.

Identification of functions and genes linked to lichenization

Two statistical analyses were combined to identify contraction/expansion or gain/loss of functions linked to lichenization. The first analysis consists on performing a Wilcoxon test (detect IPR/orthogroups with a significant difference between the LFA and the nLFA) coupled with a Fisher exact test to detect (1) IPR/orthogroups for which the difference between LFA and NLFA is linked to lichenization with a fisher pval < 0.01 that correspond to gene gain/loss (2) IPR/orthogroups for which the difference between LFA and NLFA is linked to the number of genes copies with a fisher pval > 0.01 that correspond to gene family expansion/contraction.The second analysis consists in performing a sparse PLSDA (sparse Partial Least Square Discriminant Analysis) using the 'mixOmics' R package [105] to identify IPR/orthogroups that best discriminate LFA and nLFA. For the LFA/nLFA comparison, the most explicative 100 IPR and the 50 orthogroups were kept (Figure S4). Significative IPR/orthogroups from both analyses were crossed and those appearing in both were further investigated. To empower the statistical approaches, transcriptomic data from *Trebouxia sp.* (TZW2008 isolate, [21]) was used to analyze significative IPR/orthogroups that also corresponds to a DEG in *Trebouxia sp.* TZW2008.

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 38 investigated species using the BLASTp v2.10.1 algorithm [106] and an e-value threshold of 1e-30. Then, retained sequences were aligned using MUSCLE v3.8.1551 [107] or Clustal Omega [108] with default parameters and obtained alignments cleaned with trimAl v1.4.1 [109] to remove positions with more than 80% of gaps. Finally, alignments were used as matrix for Maximum Likelihood analysis. First, phylogenetic reconstructions have been performed using FastTree v2.1.11 [110] with default parameters to obtain a global topology of the tree. Detailed analyses of subclades of interest were conducted using with IQ-TREE v2.0.3 [111]. Prior to tree reconstruction, the best fitting evolutionary model was tested using ModelFinder [112] as implemented in IQ-TREE. Branch supports were tested using 10,000 replicates of both SH-aLRT [113] and ultrafast bootstrap [114]. Trees were visualized and annotated in the iTOL v6 platform [115].

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 unsing the R package chromoMap v0.3.1. Secondly, reads from sequencing were mapped on the region containing the algal GH8 to control for chimeric assembly using minimap2 v2.17-r941 [116] and default parameters. Finally, a phylogenetic analysis was conducted to place algal GH8 in an evolutionary context. Using the BLASTp v2.10.1+ algorithm [106] with an e-value threshold of 1e-30 homologs of algal GH8-like were searched for in three different database: the JGI fungal resources (accessed in February 2020, contains more than 1600 fungal genomes) and the non-redundant protein database from NCBI (May 2022). Obtained sequences were subjected to phylogenetic analysis as described above and using MAFFT v7.453 [117] for the alignment step and FastTree v2.1.11 [110]. Presence of the GH8 functional domain was determined using hmmsearch from the HMMER v3.3 package [118] with default parameters and using the GH8 domain model (Pfam accession: PF01270). Protein structure was predicted using AlphaFold v2.1.0 [44].

GH8 enzyme activity analysis

<u>*Recombinant production of GH8 enzyme*</u>

From all the LFA GH8, we selected the protein *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 [119] 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 sulfate, 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 (Sigma-Aldrich), were purchased from Megazyme (Bray, Ireland). Amorphous cellulose (Phosphoris acid swollen cellulose or PASC) was prepared from Avicel cellulose as described in [120]. 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.

Acknowledgments

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 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. We are grateful to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, https://doi.org/10.15454/1.5572369328961167E12) for providing computing resources. Part of the work described was performed using services provided by the 3PE platform, a member of IBISBA-FR

(https://doi.org/10.15454/08BX-VJ91; www.ibisba.fr), the French node of the European research infrastructure, EU-IBISBA (www.ibisba.eu).

All data were deposited in NCBI under the BioProject PRJNA790449.

References

1. Rich, M.K., Vigneron, N., Libourel, C., Keller, J., Xue, L., Hajheidari, M., Radhakrishnan, G.V., Le Ru, A., Diop, S.I., Potente, G., et al. (2021). Lipid exchanges drove the evolution of mutualism during plant terrestrialization. Science *372*, 864–868. 10.1126/science.abg0929.

2. Field, K.J., Pressel, S., Duckett, J.G., Rimington, W.R., and Bidartondo, M.I. (2015). Symbiotic options for the conquest of land. Trends in Ecology & Evolution *30*, 477–486. 10.1016/j.tree.2015.05.007.

3. Rensing, S.A. (2018). Great moments in evolution: the conquest of land by plants. Current Opinion in Plant Biology *42*, 49–54. 10.1016/j.pbi.2018.02.006.

4. Beerling, D. (2007). The Emerald Planet: How plants changed Earth's history (Oxford University Press) 10.1093/oso/9780192806024.001.0001.

5. Sanders, W.B., and Masumoto, H. (2021). Lichen algae: the photosynthetic partners in lichen symbioses. The Lichenologist *53*, 347–393. 10.1017/S0024282921000335.

6. Nash, T.H. ed. (2008). Lichen biology 2nd ed. (Cambridge University Press).

7. Asplund, J., and Wardle, D.A. (2017). How lichens impact on terrestrial community and ecosystem properties: How lichens impact on communities and ecosystems. Biol Rev *92*, 1720–1738. 10.1111/brv.12305.

8. Berbee, M.L., Strullu-Derrien, C., Delaux, P.-M., Strother, P.K., Kenrick, P., Selosse, M.-A., and Taylor, J.W. (2020). Genomic and fossil windows into the secret lives of the most ancient fungi. Nat Rev Microbiol *18*, 717–730. 10.1038/s41579-020-0426-8.

9. Nelsen, M.P., Lücking, R., Boyce, C.K., Lumbsch, H.T., and Ree, R.H. (2020). No support for the emergence of lichens prior to the evolution of vascular plants. Geobiology *18*, 3–13. 10.1111/gbi.12369.

10. Honegger, R., Edwards, D., and Axe, L. (2013). The earliest records of internally stratified cyanobacterial and algal lichens from the Lower Devonian of the Welsh Borderland. New Phytol *197*, 264–275. 10.1111/nph.12009.

11. Lutzoni, F., Nowak, M.D., Alfaro, M.E., Reeb, V., Miadlikowska, J., Krug, M., Arnold, A.E., Lewis, L.A., Swofford, D.L., Hibbett, D., et al. (2018). Contemporaneous radiations of fungi and plants linked to symbiosis. Nat Commun *9*, 5451. 10.1038/s41467-018-07849-9.

12. Delaux, P.-M., Varala, K., Edger, P.P., Coruzzi, G.M., Pires, J.C., and Ané, J.-M. (2014). Comparative Phylogenomics Uncovers the Impact of Symbiotic Associations on Host Genome Evolution. PLoS Genet *10*, e1004487. 10.1371/journal.pgen.1004487.

13. Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M.B., Billault-Penneteau, B., Lauressergues, D., Keller, J., Imanishi, L., et al. (2018). Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. Science *361*, eaat1743. 10.1126/science.aat1743.

14. van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., Santuari, L., Cao, Q., Sharma, T., Shen, D., et al. (2018). Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. Proc Natl Acad Sci USA *115*, E4700–E4709. 10.1073/pnas.1721395115.

15. Kohler, A., Kuo, A., Nagy, L.G., Morin, E., Barry, K.W., Buscot, F., Canbäck, B., Choi, C., Cichocki, N., Clum, A., et al. (2015). Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. Nat Genet *47*, 410–415. 10.1038/ng.3223.

16. Radhakrishnan, G.V., Keller, J., Rich, M.K., Vernié, T., Mbadinga Mbadinga, D.L., Vigneron, N., Cottret, L., Clemente, H.S., Libourel, C., Cheema, J., et al. (2020). An ancestral signalling pathway is conserved in intracellular symbioses-forming plant lineages. Nat. Plants *6*, 280–289. 10.1038/s41477-020-0613-7.

17. Kiss, E., Hegedüs, B., Virágh, M., Varga, T., Merényi, Z., Kószó, T., Bálint, B., Prasanna, A.N., Krizsán, K., Kocsubé, S., et al. (2019). Comparative genomics reveals the origin of fungal hyphae and multicellularity. Nat Commun *10*, 4080. 10.1038/s41467-019-12085-w.

18. Mesny, F., Miyauchi, S., Thiergart, T., Pickel, B., Atanasova, L., Karlsson, M., Hüttel, B., Barry, K.W., Haridas, S., Chen, C., et al. (2021). Genetic determinants of endophytism in the Arabidopsis root mycobiome. Nat Commun *12*, 7227. 10.1038/s41467-021-27479-y.

19. Friedl, T., and Bhattacharya, D. (2004). Origin and Evolution of Green Lichen Algae. In Symbiosis Cellular Origin, Life in Extreme Habitats and Astrobiology., J. Seckbach, ed. (Kluwer Academic Publishers), pp. 341–357. 10.1007/0-306-48173-1_21.

20. Armaleo, D., Müller, O., Lutzoni, F., Andrésson, Ó.S., Blanc, G., Bode, H.B., Collart, F.R., Dal Grande, F., Dietrich, F., Grigoriev, I.V., et al. (2019). The lichen symbiosis re-viewed through the genomes of Cladonia grayi and its algal partner Asterochloris glomerata. BMC Genomics *20*, 605. 10.1186/s12864-019-5629-x.

21. Kono, M., Kon, Y., Ohmura, Y., Satta, Y., and Terai, Y. (2020). In vitro resynthesis of lichenization reveals the genetic background of symbiosis-specific fungal-algal interaction in Usnea hakonensis. BMC Genomics *21*, 671. 10.1186/s12864-020-07086-9.

22. Ran, L., Larsson, J., Vigil-Stenman, T., Nylander, J.A.A., Ininbergs, K., Zheng, W.-W., Lapidus, A., Lowry, S., Haselkorn, R., and Bergman, B. (2010). Genome Erosion in a Nitrogen-Fixing Vertically Transmitted Endosymbiotic Multicellular Cyanobacterium. PLoS ONE *5*, e11486. 10.1371/journal.pone.0011486.

23. McCutcheon, J.P., and Moran, N.A. (2012). Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol *10*, 13–26. 10.1038/nrmicro2670.

24. Nechitaylo, T.Y., Sandoval-Calderón, M., Engl, T., Wielsch, N., Dunn, D.M., Goesmann, A., Strohm, E., Svatoš, A., Dale, C., Weiss, R.B., et al. (2021). Incipient genome erosion and metabolic streamlining

634 for antibiotic production in a defensive symbiont. Proc. Natl. Acad. Sci. U.S.A. *118*, e2023047118.
635 10.1073/pnas.2023047118.

636 25. Noh, S. (2021). Linear paths to genome reduction in a defensive symbiont. Proc Natl Acad Sci USA
637 *118*, e2106280118. 10.1073/pnas.2106280118.

638 26. Wickell, D.A., and Li, F. (2020). On the evolutionary significance of horizontal gene transfers in plants.
639 New Phytol *225*, 113–117. 10.1111/nph.16022.

640 27. Mhamdi, A., Queval, G., Chaouch, S., Vanderauwera, S., Van Breusegem, F., and Noctor, G. (2010).
641 Catalase function in plants: a focus on Arabidopsis mutants as stress-mimic models. Journal of
642 Experimental Botany *61*, 4197–4220. 10.1093/jxb/erq282.

643 28. Frank, W., Baar, K.-M., Qudeimat, E., Woriedh, M., Alawady, A., Ratnadewi, D., Gremillon, L., Grimm,
644 B., and Reski, R. (2007). A mitochondrial protein homologous to the mammalian peripheral-type
645 benzodiazepine receptor is essential for stress adaptation in plants: Knock-out of a TSPO homologue
646 in moss. The Plant Journal *51*, 1004–1018. 10.1111/j.1365-313X.2007.03198.x.

647 29. Carniel, F.C., Gerdol, M., Montagner, A., Banchi, E., De Moro, G., Manfrin, C., Muggia, L., Pallavicini,
648 A., and Tretiach, M. (2016). New features of desiccation tolerance in the lichen photobiont Trebouxia
649 gelatinosa are revealed by a transcriptomic approach. Plant Mol Biol *91*, 319–339. 10.1007/s11103-
650 016-0468-5.

651 30. Palmqvist, K. (2000). Carbon economy in lichens. New Phytologist *148*, 11–36. 10.1046/j.1469-
652 8137.2000.00732.x.

653 31. Meeßen, J., Eppenstein, S., and Ott, S. (2013). Recognition mechanisms during the pre-contact state
654 of lichens: II. Influence of algal exudates and ribitol on the response of the mycobiont of Fulgensia
655 bracteata. Symbiosis *59*, 131–143. 10.1007/s13199-012-0219-6.

656 32. Ahmadjian, V. (1993). The lichen symbiosis (John Wiley).

657 33. Minic, Z., and Jouanin, L. (2006). Plant glycoside hydrolases involved in cell wall polysaccharide
658 degradation. Plant Physiology and Biochemistry *44*, 435–449. 10.1016/j.plaphy.2006.08.001.

659 34. Brewin, N.J. (2004). Plant Cell Wall Remodelling in the Rhizobium–Legume Symbiosis. Critical Reviews
660 in Plant Sciences *23*, 293–316. 10.1080/07352680490480734.

661 35. Balestrini, R., and Bonfante, P. (2014). Cell wall remodeling in mycorrhizal symbiosis: a way towards
662 biotrophism. Front. Plant Sci. *5*. 10.3389/fpls.2014.00237.

663 36. Chapman, R.L. (1976). Ultrastructural investigation on the foliicolous pyrenocarpous lichen Strigula
664 elegans (Fée) Müll. Arg. Phycologia *15*, 191–196. 10.2216/i0031-8884-15-2-191.1.

665 37. Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., Charron, P., Duensing, N.,
666 Frei dit Frey, N., Gianinazzi-Pearson, V., et al. (2013). Genome of an arbuscular mycorrhizal fungus
667 provides insight into the oldest plant symbiosis. Proceedings of the National Academy of Sciences
668 *110*, 20117–20122. 10.1073/pnas.1313452110.

669 38. Martino, E., Morin, E., Grelet, G., Kuo, A., Kohler, A., Daghino, S., Barry, K.W., Cichocki, N., Clum, A.,
670 Dockter, R.B., et al. (2018). Comparative genomics and transcriptomics depict ericoid mycorrhizal
671 fungi as versatile saprotrophs and plant mutualists. New Phytol *217*, 1213–1229. 10.1111/nph.14974.

672 39. Resl, P., Bujold, A.R., Tagirdzhanova, G., Meidl, P., Rallo, S.F., Kono, M., Fernández-Brime, S.,
673 Guðmundsson, H., Andrésson, Ó.S., Muggia, L., et al. (2021). Large differences in carbohydrate
674 degradation and transport potential in the genomes of lichen fungal symbionts (Genomics)
675 10.1101/2021.08.01.454614.

676 40. Drula, E., Garron, M.-L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2022). The
677 carbohydrate-active enzyme database: functions and literature. Nucleic Acids Research *50*, D571–
678 D577. 10.1093/nar/gkab1045.

679 41. Perlin, A.S., and Suzuki, S. (1962). The Structure of Lichenin: Selective Enzymolysis studies. Can. J.
680 Chem. *40*, 50–56. 10.1139/v62-009.

681 42. Adachi, W., Sakihama, Y., Shimizu, S., Sunami, T., Fukazawa, T., Suzuki, M., Yatsunami, R., Nakamura,
682 S., and Takénaka, A. (2004). Crystal Structure of Family GH-8 Chitosanase with Subclass II Specificity
683 from Bacillus sp. K17. Journal of Molecular Biology *343*, 785–795. 10.1016/j.jmb.2004.08.028.

684 43. The CAZypedia Consortium (2018). Ten years of CAZypedia: a living encyclopedia of carbohydrate-
685 active enzymes. Glycobiology *28*, 3–8. 10.1093/glycob/cwx089.

686 44. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates,
687 R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold.
688 Nature *596*, 583–589. 10.1038/s41586-021-03819-2.

689 45. Lafond, M., Sulzenbacher, G., Freyd, T., Henrissat, B., Berrin, J.-G., and Garron, M.-L. (2016). The
690 Quaternary Structure of a Glycoside Hydrolase Dictates Specificity toward β-Glucans. Journal of
691 Biological Chemistry *291*, 7183–7194. 10.1074/jbc.M115.695999.

692 46. Gargas, A., DePriest, P.T., Grube, M., and Tehler, A. (1995). Multiple Origins of Lichen Symbioses in
693 Fungi Suggested by SSU rDNA Phylogeny. Science *268*, 1492–1495. 10.1126/science.7770775.

694 47. Divakar, P.K., Crespo, A., Wedin, M., Leavitt, S.D., Hawksworth, D.L., Myllys, L., McCune, B., Randlane,
695 T., Bjerke, J.W., Ohmura, Y., et al. (2015). Evolution of complex symbiotic relationships in a
696 morphologically derived family of lichen-forming fungi. New Phytol *208*, 1217–1226.
697 10.1111/nph.13553.

698 48. Nelsen, M.P., Lücking, R., Boyce, C.K., Lumbsch, H.T., and Ree, R.H. (2020). The macroevolutionary
699 dynamics of symbiotic and phenotypic diversification in lichens. Proc. Natl. Acad. Sci. U.S.A. *117*,
700 21495–21503. 10.1073/pnas.2001913117.

701 49. Hanschen, E.R., and Starkenburg, S.R. (2020). The state of algal genome quality and diversity. Algal
702 Research *50*, 101968. 10.1016/j.algal.2020.101968.

703 50. Lemieux, C., Otis, C., and Turmel, M. (2014). Chloroplast phylogenomic analysis resolves deep-level
704 relationships within the green algal class Trebouxiophyceae. BMC Evol Biol *14*, 211. 10.1186/s12862-
705 014-0211-2.

706 51. Hom, E.F.Y., and Murray, A.W. (2014). Niche engineering demonstrates a latent capacity for fungal-
707 algal mutualism. Science *345*, 94–98. 10.1126/science.1253320.

708 52. Zhang, F., Labourel, A., Haon, M., Kemppainen, M., Da Silva Machado, E., Brouilly, N., Veneault-
709 Fourrey, C., Kohler, A., Rosso, M., Pardo, A., et al. (2022). The ectomycorrhizal basidiomycete *Laccaria*
710 *bicolor* releases a GH28 polygalacturonase that plays a key role in symbiosis establishment. New
711 Phytologist *233*, 2534–2547. 10.1111/nph.17940.

712 53. Zhang, F., Anasontzis, G.E., Labourel, A., Champion, C., Haon, M., Kemppainen, M., Commun, C.,
713 Deveau, A., Pardo, A., Veneault-Fourrey, C., et al. (2018). The ectomycorrhizal basidiomycete *Laccaria*
714 *bicolor* releases a secreted β-1,4 endoglucanase that plays a key role in symbiosis development. New
715 Phytol *220*, 1309–1321. 10.1111/nph.15113.

716 54. Delaux, P.-M., Séjalon-Delmas, N., Bécard, G., and Ané, J.-M. (2013). Evolution of the plant–microbe
717 symbiotic 'toolkit.' Trends in Plant Science *18*, 298–304. 10.1016/j.tplants.2013.01.008.

718 55. Tatara, H., Sato, Y., Kasai, Y., Hagiwara, D., Makino, A., Imamura, S., Abe, J., Yoshimitsu, Y., and
719 Harayama, S. (2020). A method for the preparation of electrocompetent cells to transform unicellular
720 green algae, Coccomyxa (Trebouxiophyceae, Chlorophyta) strains Obi and KJ. Algal Research *48*,
721 101904. 10.1016/j.algal.2020.101904.

722 56. Dal Grande, F., Rolshausen, G., Divakar, P.K., Crespo, A., Otte, J., Schleuning, M., and Schmitt, I. (2018).
723 Environment and host identity structure communities of green algal symbionts in lichens. New Phytol
724 *217*, 277–289. 10.1111/nph.14770.

725 57. Beck, A., and Koop, H.-U. (2001). Analysis of the Photobiont Population in Lichens Using a Single-Cell
726 Manipulator. Symbiosis *31*, 57–67.

727 58. Ahmadjian, V. (1969). Lichen Synthesis. Österreichische Botanische Zeitschrift *116*, 306–311.

728 59. Bethune, K., Mariac, C., Couderc, M., Scarcelli, N., Santoni, S., Ardisson, M., Martin, J.-F., Montúfar,
729 R., Klein, V., Sabot, F., et al. (2019). Long-fragment targeted capture for long-read sequencing of
730 plastomes. Appl Plant Sci *7*, e1243. 10.1002/aps3.1243.

731 60. Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Marande, W., Callot, C., Langlade, N., and Muños,
732 S. (2016). Extraction of high-molecular-weight genomic DNA for long-read sequencing of single
733 molecules. BioTechniques *61*, 203–205. 10.2144/000114460.

734 61. Merges, D., Dal Grande, F., Greve, C., Otte, J., and Schmitt, I. (2021). Virus diversity in metagenomes
735 of a lichen symbiosis (*Umbilicaria phaea*): complete viral genomes, putative hosts and elevational
736 distributions. Environ Microbiol *23*, 6637–6650. 10.1111/1462-2920.15802.

737 62. Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using
738 repeat graphs. Nat Biotechnol *37*, 540–546. 10.1038/s41587-019-0072-8.

739 63. Boetzer, M., and Pirovano, W. (2014). SSPACE-LongRead: scaffolding bacterial draft genomes using
740 long read sequence information. BMC Bioinformatics *15*, 211. 10.1186/1471-2105-15-211.

741 64. Huson, D.H., Albrecht, B., Bağcı, C., Bessarab, I., Górska, A., Jolic, D., and Williams, R.B.H. (2018).
742 MEGAN-LR: new algorithms allow accurate binning and easy interactive exploration of metagenomic
743 long reads and contigs. Biol Direct *13*, 6. 10.1186/s13062-018-0208-7.

744 65. Girgis, H.Z. (2015). Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the
745 genomic scale. BMC Bioinformatics *16*, 227. 10.1186/s12859-015-0654-5.

746 66. Brůna, T., Hoff, K.J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: automatic
747 eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database.
748 NAR Genomics and Bioinformatics *3*, lqaa108. 10.1093/nargab/lqaa108.

749 67. Brůna, T., Lomsadze, A., and Borodovsky, M. (2020). GeneMark-EP+: eukaryotic gene prediction with
750 self-training in the space of genes and proteins. NAR Genomics and Bioinformatics *2*, lqaa026.
751 10.1093/nargab/lqaa026.

752 68. Barnett, D.W., Garrison, E.K., Quinlan, A.R., Stromberg, M.P., and Marth, G.T. (2011). BamTools: a C++
753 API and toolkit for analyzing and managing BAM files. Bioinformatics *27*, 1691–1692.
754 10.1093/bioinformatics/btr174.

755 69. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.,
756 and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format
757 and SAMtools. Bioinformatics *25*, 2078–2079. 10.1093/bioinformatics/btp352.

758 70. Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S.O., and Grau, J. (2018). Combining RNA-seq data
759 and homology-based gene prediction for plants, animals and fungi. BMC Bioinformatics *19*, 189.
760 10.1186/s12859-018-2203-5.

761 71. Keilwagen, J., Wenk, M., Erickson, J.L., Schattat, M.H., Grau, J., and Hartung, F. (2016). Using intron
762 position conservation for homology-based gene prediction. Nucleic Acids Res *44*, e89–e89.
763 10.1093/nar/gkw092.

764 72. Lomsadze, A., Burns, P.D., and Borodovsky, M. (2014). Integration of mapped RNA-Seq reads into
765 automatic training of eukaryotic gene finding algorithm. Nucleic Acids Research *42*, e119–e119.
766 10.1093/nar/gku557.

767 73. Gotoh, O. (2008). A space-efficient and accurate method for mapping and aligning cDNA sequences
768 onto genomic sequence. Nucleic Acids Research *36*, 2630–2638. 10.1093/nar/gkn105.

769 74. Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006). Gene prediction in eukaryotes
770 with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics
771 *7*, 62. 10.1186/1471-2105-7-62.

772 75. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.
773 EMBnet j. *17*, 10. 10.14806/ej.17.1.200.

774 76. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment
775 and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol *37*, 907–915. 10.1038/s41587-019-
776 0201-4.

77. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. 10.1093/gigascience/giab008.

78. Cabau, C., Escudié, F., Djari, A., Guiguen, Y., Bobe, J., and Klopp, C. (2017). Compacting and correcting Trinity and Oases RNA-Seq *de novo* assemblies. PeerJ *5*, e2988. 10.7717/peerj.2988.

79. Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics *28*, 1086–1092. 10.1093/bioinformatics/bts094.

80. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc *8*, 1494–1512. 10.1038/nprot.2013.084.

81. Eddy, S.R. (1998). Profile hidden Markov models. Bioinformatics *14*, 755–763. 10.1093/bioinformatics/14.9.755.

82. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: The protein families database in 2021. Nucleic Acids Research *49*, D412–D419. 10.1093/nar/gkaa913.

83. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212. 10.1093/bioinformatics/btv351.

84. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240. 10.1093/bioinformatics/btu031.

85. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol *20*, 238. 10.1186/s13059-019-1832-y.

86. R Core Team (2013). R: A language and environment for statistical computing.

87. Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol *20*, 275. 10.1186/s13059-019-1905-y.

88. Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Research *35*, W265–W268. 10.1093/nar/gkm286.

89. Ou, S., and Jiang, N. (2018). LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. Plant Physiol. *176*, 1410–1422. 10.1104/pp.17.01310.

90. Su, W., Gu, X., and Peterson, T. (2019). TIR-Learner, a New Ensemble Method for TIR Transposable Element Annotation, Provides Evidence for Abundant New Transposable Elements in the Maize Genome. Molecular Plant *12*, 447–460. 10.1016/j.molp.2019.02.008.

813 91. Xiong, W., He, L., Lai, J., Dooner, H.K., and Du, C. (2014). HelitronScanner uncovers a large overlooked
814 cache of *Helitron* transposons in many plant genomes. Proc. Natl. Acad. Sci. U.S.A. *111*, 10263–10268.
815 10.1073/pnas.1410068111.

816 92. Shi, J., and Liang, C. (2019). Generic Repeat Finder: A High-Sensitivity Tool for Genome-Wide De Novo
817 Repeat Detection. Plant Physiol. *180*, 1803–1815. 10.1104/pp.19.00386.

818 93. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020).
819 RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl.
820 Acad. Sci. U.S.A. *117*, 9451–9457. 10.1073/pnas.1921046117.

821 94. Zhang, R.-G., Wang, Z.-X., Ou, S., and Li, G.-Y. (2019). TEsorter: lineage-level classification of
822 transposable elements using conserved protein domains (Bioinformatics) 10.1101/800177.

823 95. Smit, A., Hubley, R., and Green, P. (2013). RepeatMasker Open-4.0.

824 96. Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., and
825 Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. Nat
826 Biotechnol *38*, 276–278. 10.1038/s41587-020-0439-x.

827 97. Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., and Notredame, C. (2017).
828 Nextflow enables reproducible computational workflows. Nat Biotechnol *35*, 316–319.
829 10.1038/nbt.3820.

830 98. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic
831 features. Bioinformatics *26*, 841–842. 10.1093/bioinformatics/btq033.

832 99. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-
833 aware quantification of transcript expression. Nat Methods *14*, 417–419. 10.1038/nmeth.4197.

834 100. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and
835 Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
836 10.1093/bioinformatics/bts635.

837 101. Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R., Salzberg, S.L., and Pertea, M. (2019).
838 Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol *20*, 278.
839 10.1186/s13059-019-1910-1.

840 102. Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed:
841 enabling browsing of large distributed datasets. Bioinformatics *26*, 2204–2207.
842 10.1093/bioinformatics/btq351.

843 103. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for
844 differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.
845 10.1093/bioinformatics/btp616.

846 104. Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression
847 analysis of RNA-seq data. Genome Biol *11*, R25. 10.1186/gb-2010-11-3-r25.

105. Rohart, F., Gautier, B., Singh, A., and Lê Cao, K.-A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput Biol *13*, e1005752. 10.1371/journal.pcbi.1005752.

106. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421. 10.1186/1471-2105-10-421.

107. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research *32*, 1792–1797. 10.1093/nar/gkh340.

108. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol *7*, 539. 10.1038/msb.2011.75.

109. Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973. 10.1093/bioinformatics/btp348.

110. Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Molecular Biology and Evolution *26*, 1641–1650. 10.1093/molbev/msp077.

111. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution *37*, 1530–1534. 10.1093/molbev/msaa015.

112. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods *14*, 587–589. 10.1038/nmeth.4285.

113. 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. Systematic Biology *59*, 307–321. 10.1093/sysbio/syq010.

114. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution *35*, 518–522. 10.1093/molbev/msx281.

115. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research *49*, W293–W296. 10.1093/nar/gkab301.

116. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. 10.1093/bioinformatics/bty191.

117. Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution *30*, 772–780. 10.1093/molbev/mst010.

118. Johnson, L.S., Eddy, S.R., and Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics *11*, 431. 10.1186/1471-2105-11-431.

119. Haon, M., Grisel, S., Navarro, D., Gruet, A., Berrin, J.-G., and Bignon, C. (2015). Recombinant protein production facility for fungal biomass-degrading enzymes using the yeast Pichia pastoris. Front Microbiol *6*, 1002. 10.3389/fmicb.2015.01002.

120. Bennati-Granier, C., Garajova, S., Champion, C., Grisel, S., Haon, M., Zhou, S., Fanuel, M., Ropartz, D., Rogniaux, H., Gimbert, I., et al. (2015). Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by Podospora anserina. Biotechnol Biofuels *8*, 90. 10.1186/s13068-015-0274-3.

Supplementary Data

Supplementary Data S1: Newick trees

Supplementary Table S1: List of species used in this study

Supplementary Table S2: Long-read DNA sequencing statistics

Supplementary Table S3: Illumina RNASeq stastitics

Supplementary Table S4: Detailed percentage of TEs per species and per TE familiy

Supplementary Table S5: Significant IPR and orthogroups related to lichenization, and phylogenetic analysis

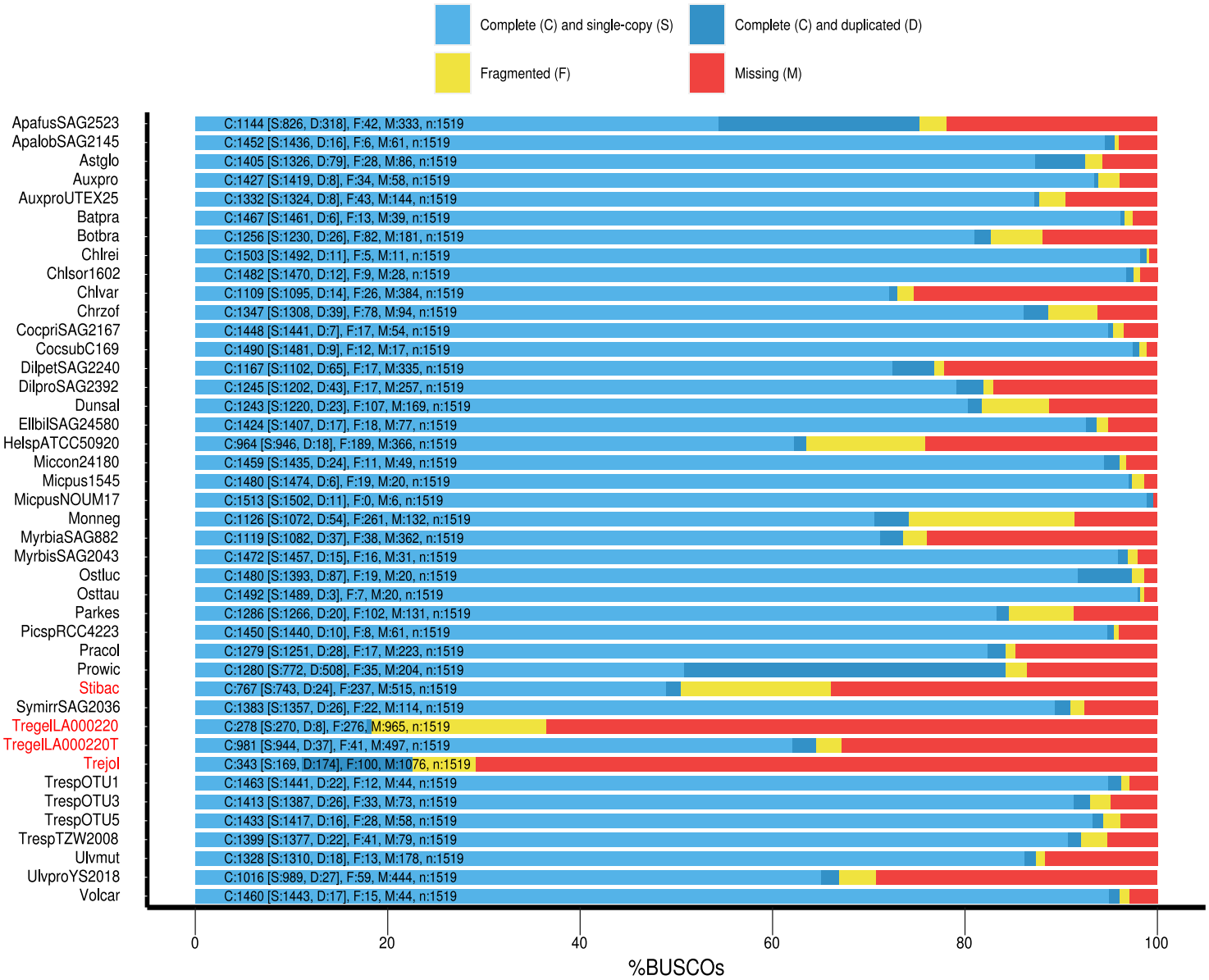
Supplementary Table S6: Re-estimation of differentially expressed genes of Trebouxia sp. TZW2008.

Supplementary Table S7: Number of IPR and orthogroups phylogenetically analysed and the number of expansions in the N15 (Trebouxiophcean) clade

Supplementary Table S8: List of sequences used for GH8 phylogenetic analysis

Supplementary Table S9: Phylogenetic hierarchical orthogroups

Supplementary Figure S1: BUSCO assessment results. Abbreviations in red indicate discarded species



Supplementary Figure S2: Genomic comparisons between algae classes (left) and LFA s nLFA (right). Statistical comparisons have been performed using a Wilcoxon test and p-value threshold of 0.05. For all comparions, the outgroup species *Prasinoderma coloniale* has been excluded.

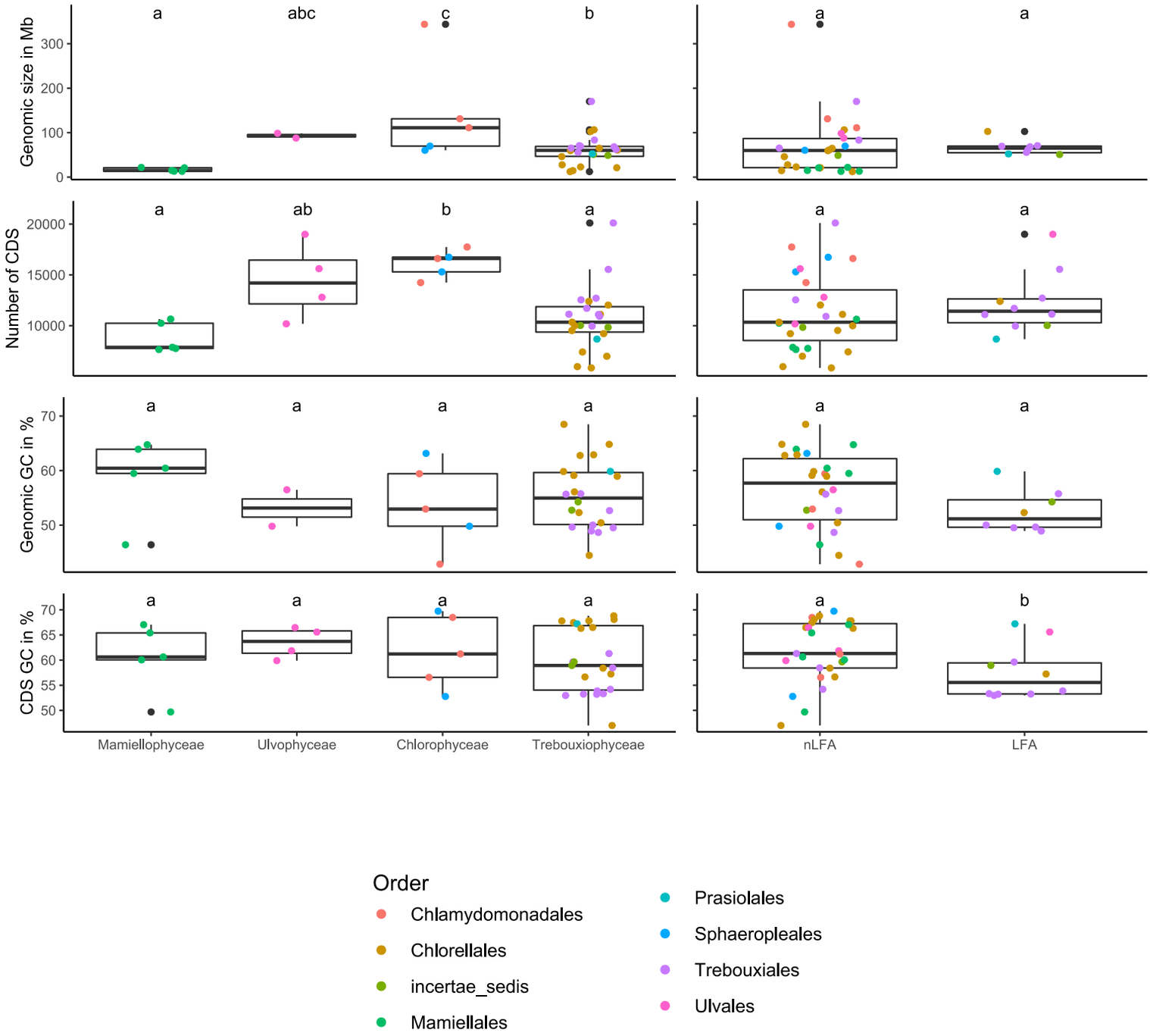


Figure S3: Comparison of TEs percentage between algae classes (left column) and LFA vs nLFA (right column). Statistical comparisons have been performed using a Wilcoxon test and p-value threshold of 0.05. For all comparions, the outgroup species Prasinoderma coloniale has been excluded.

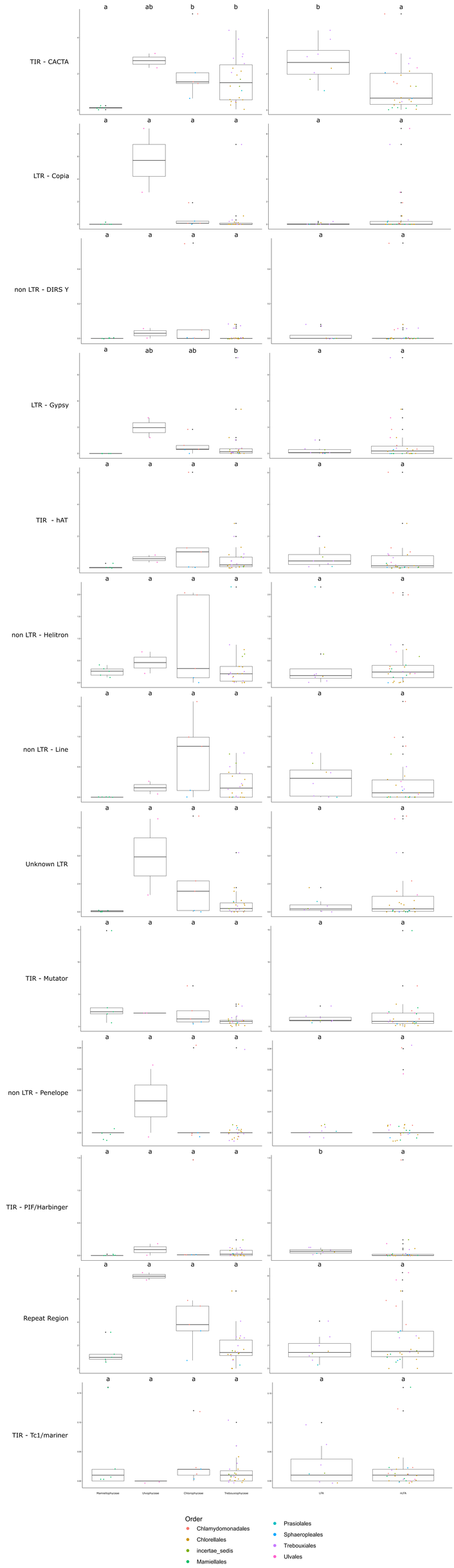
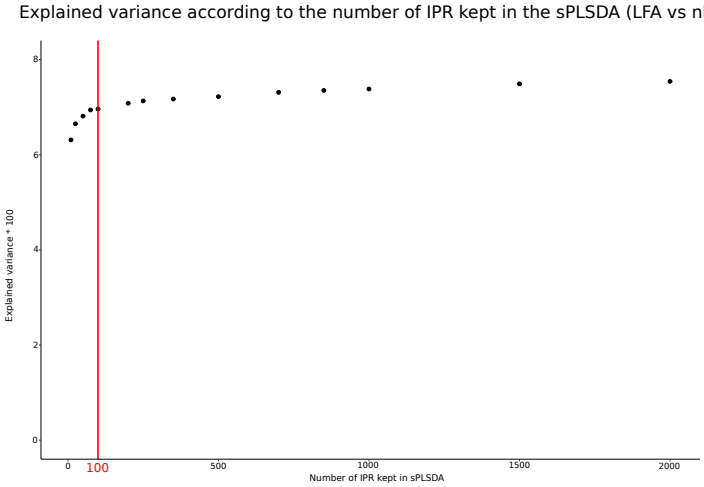
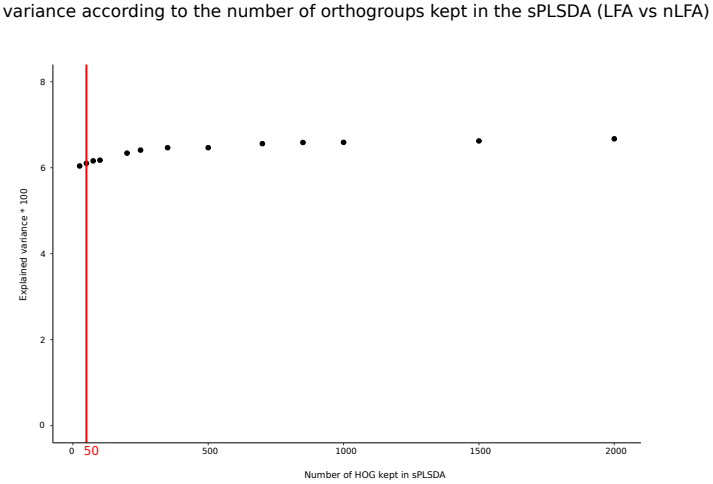
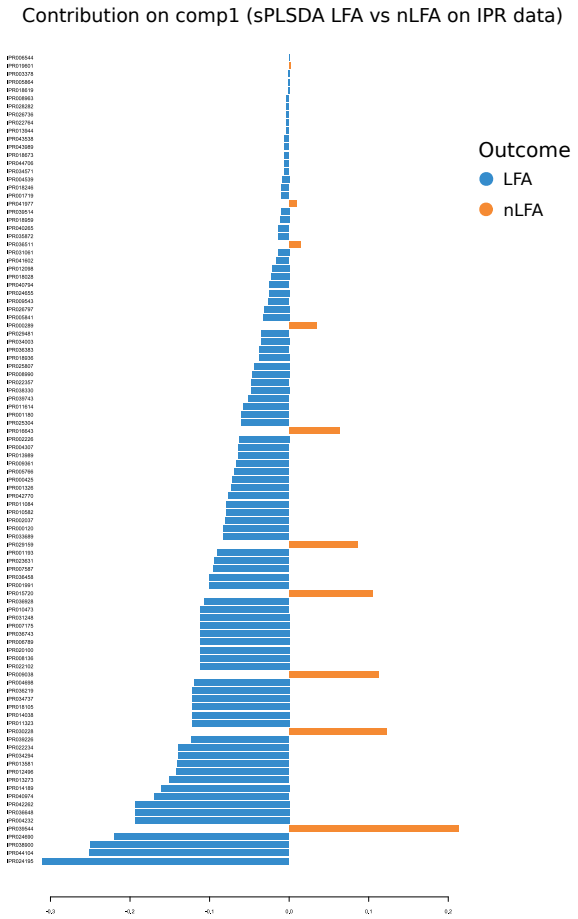


Figure S3: explained variance according to the number of IPR or orthogroups kept in the sPLSDA and the contribution of the IPR or orthogroups on the first component







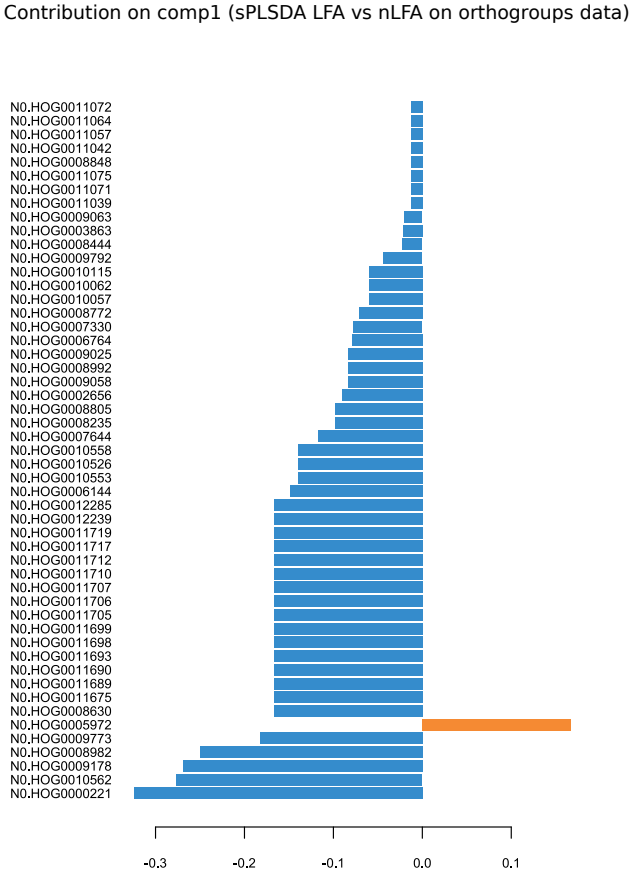
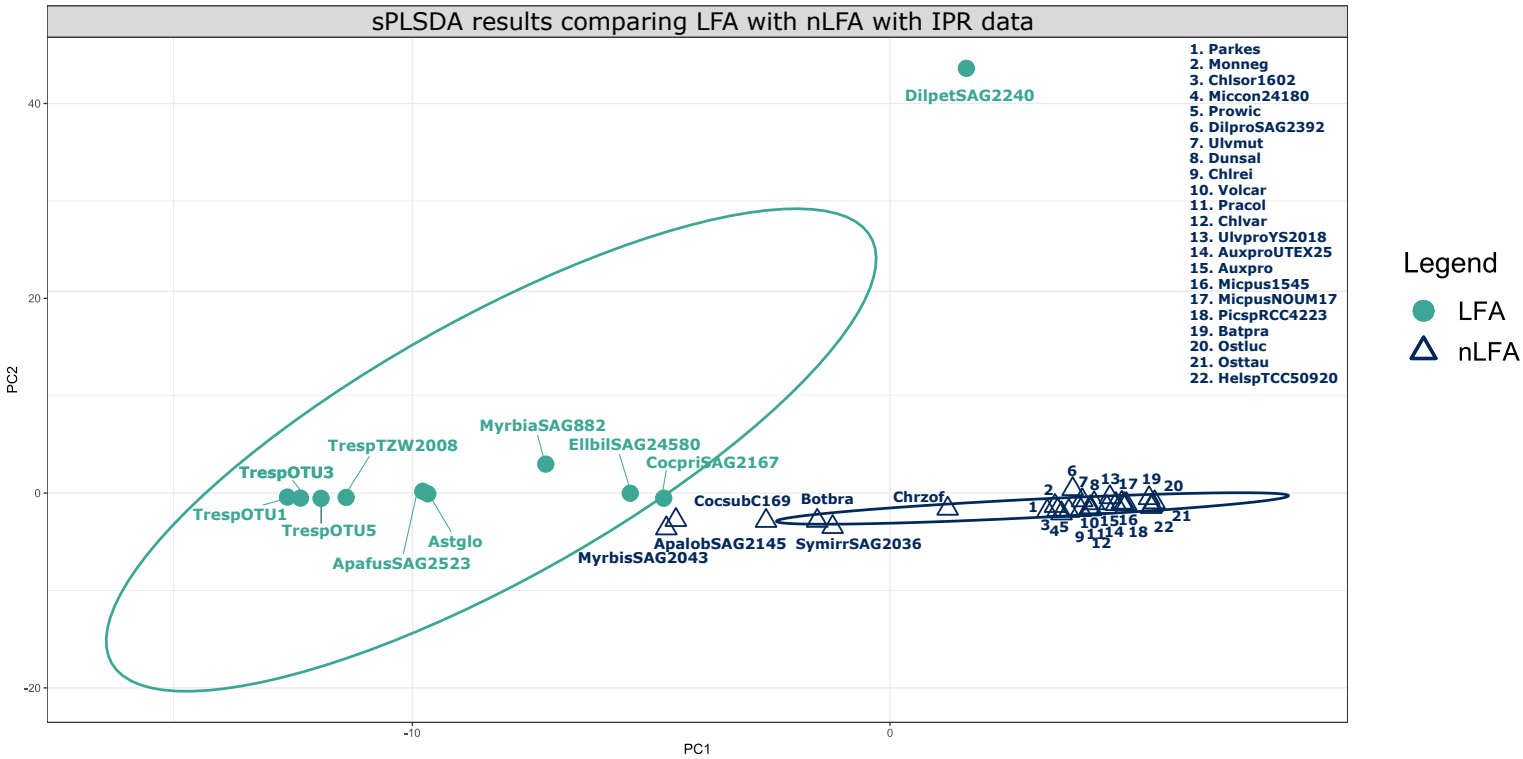


Figure S5: Individual plots of the sPLSDA conducted on the IPR and orthogroups data



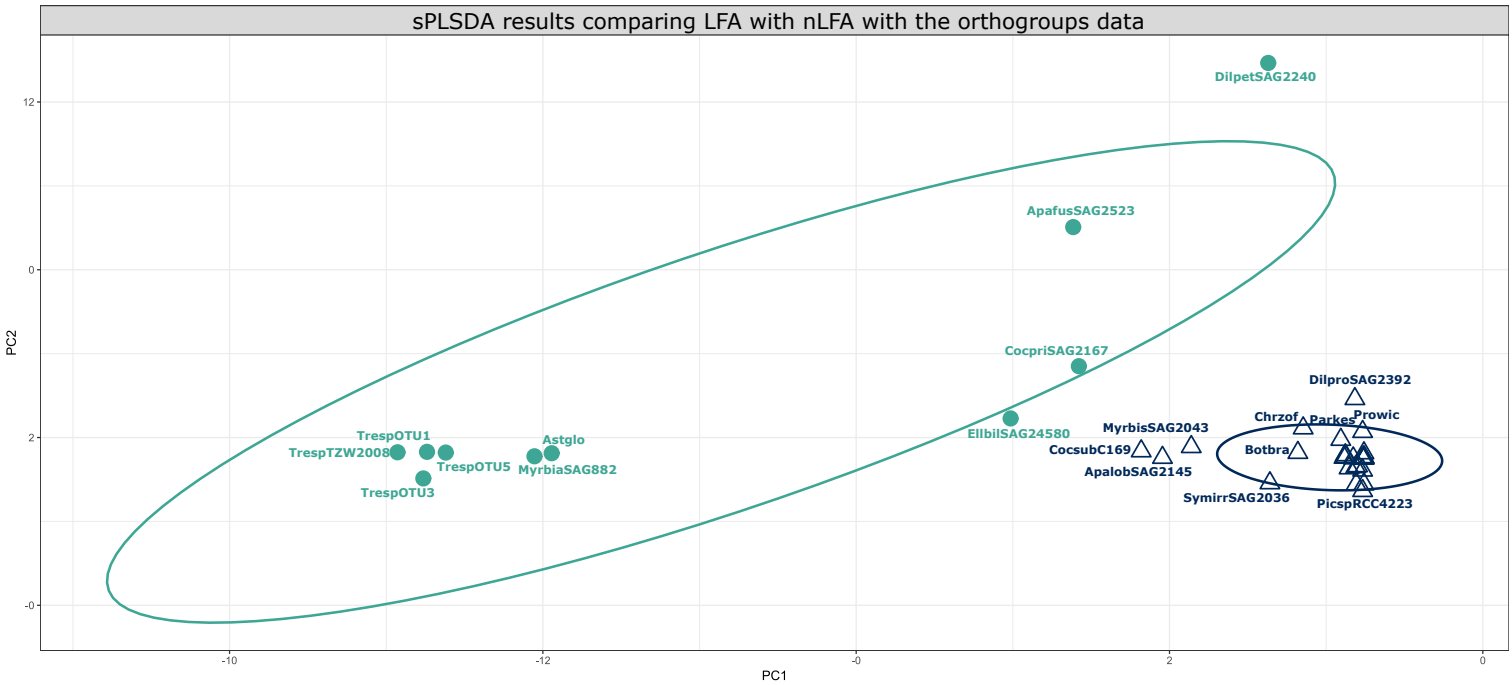
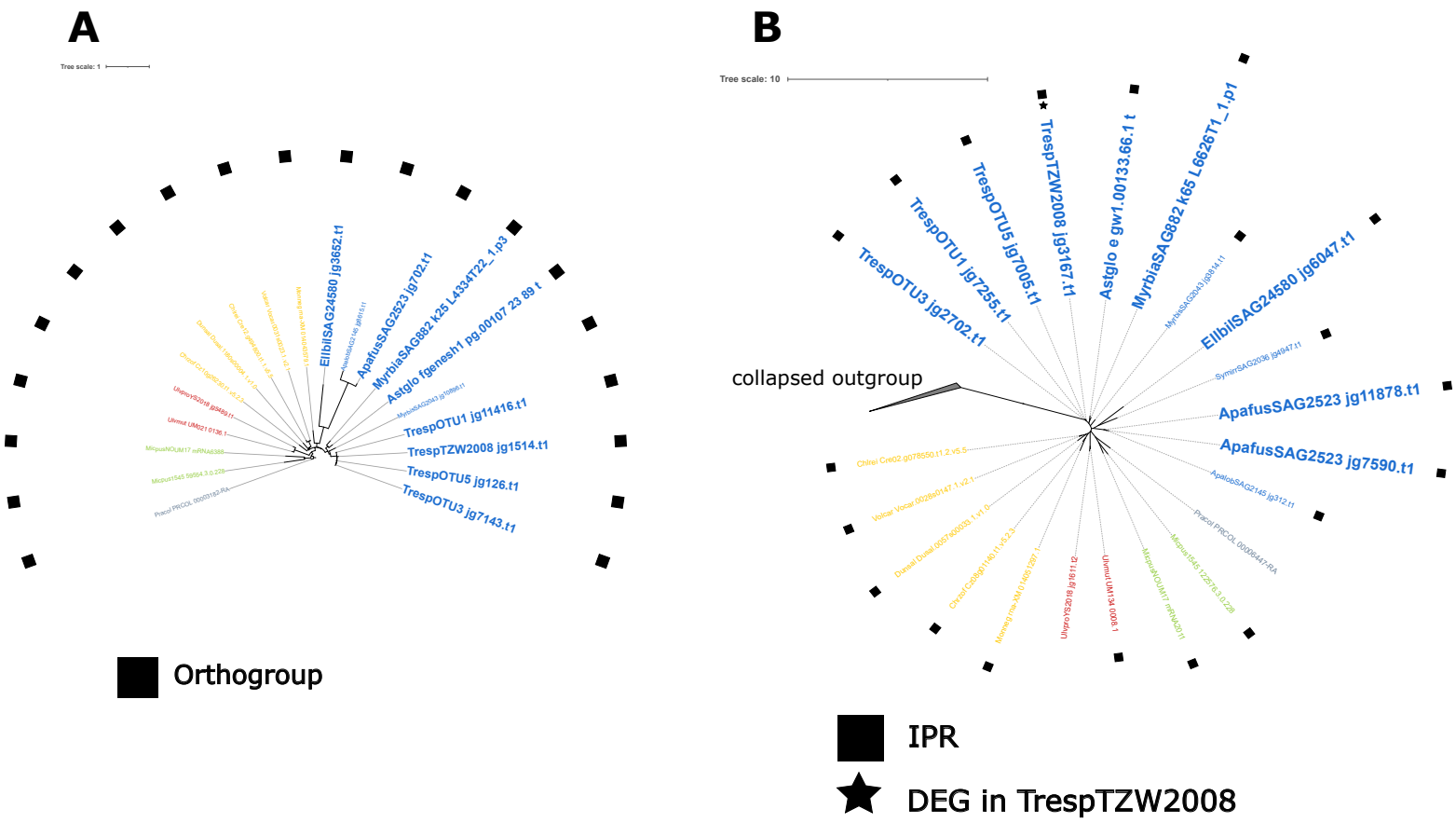
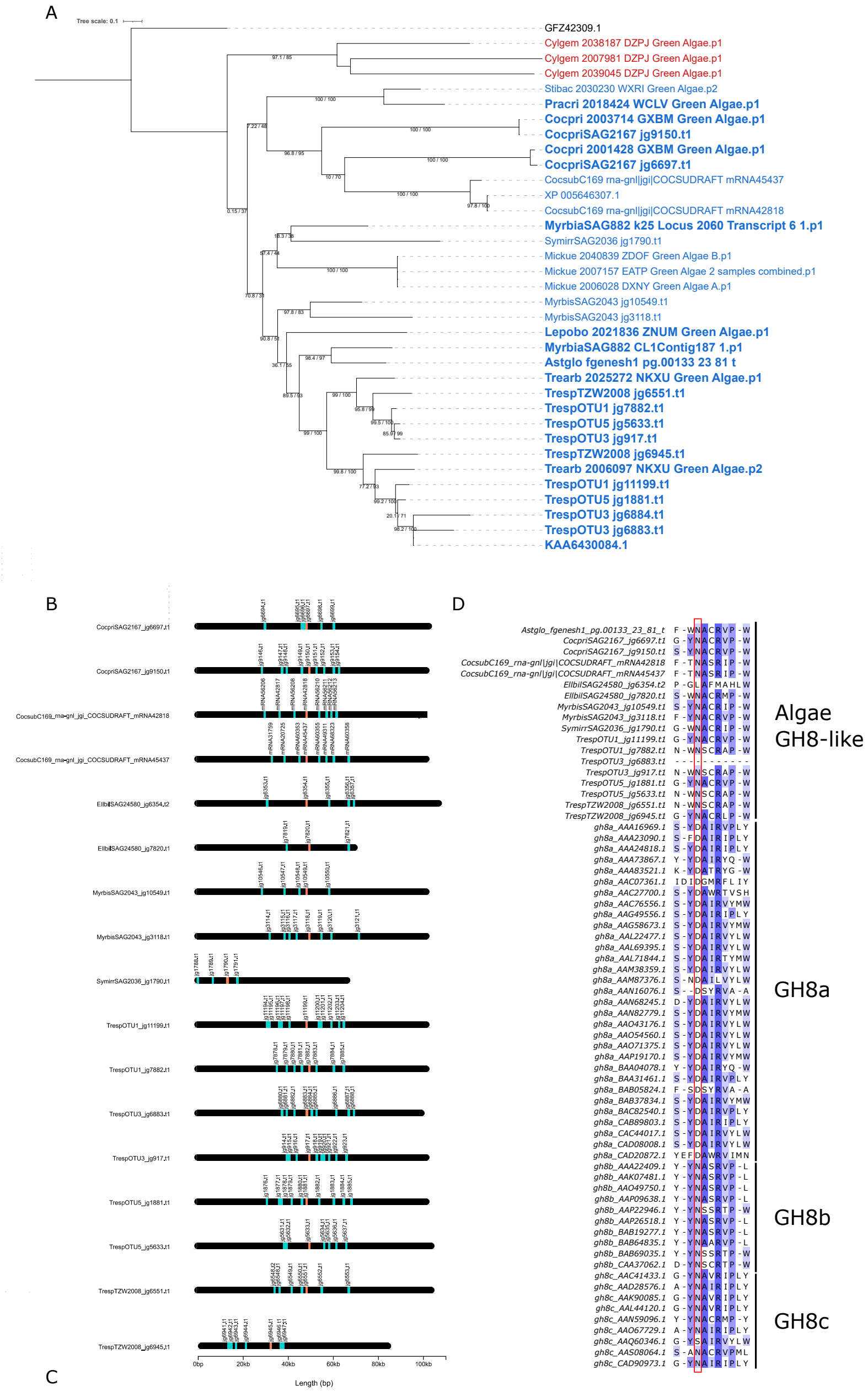


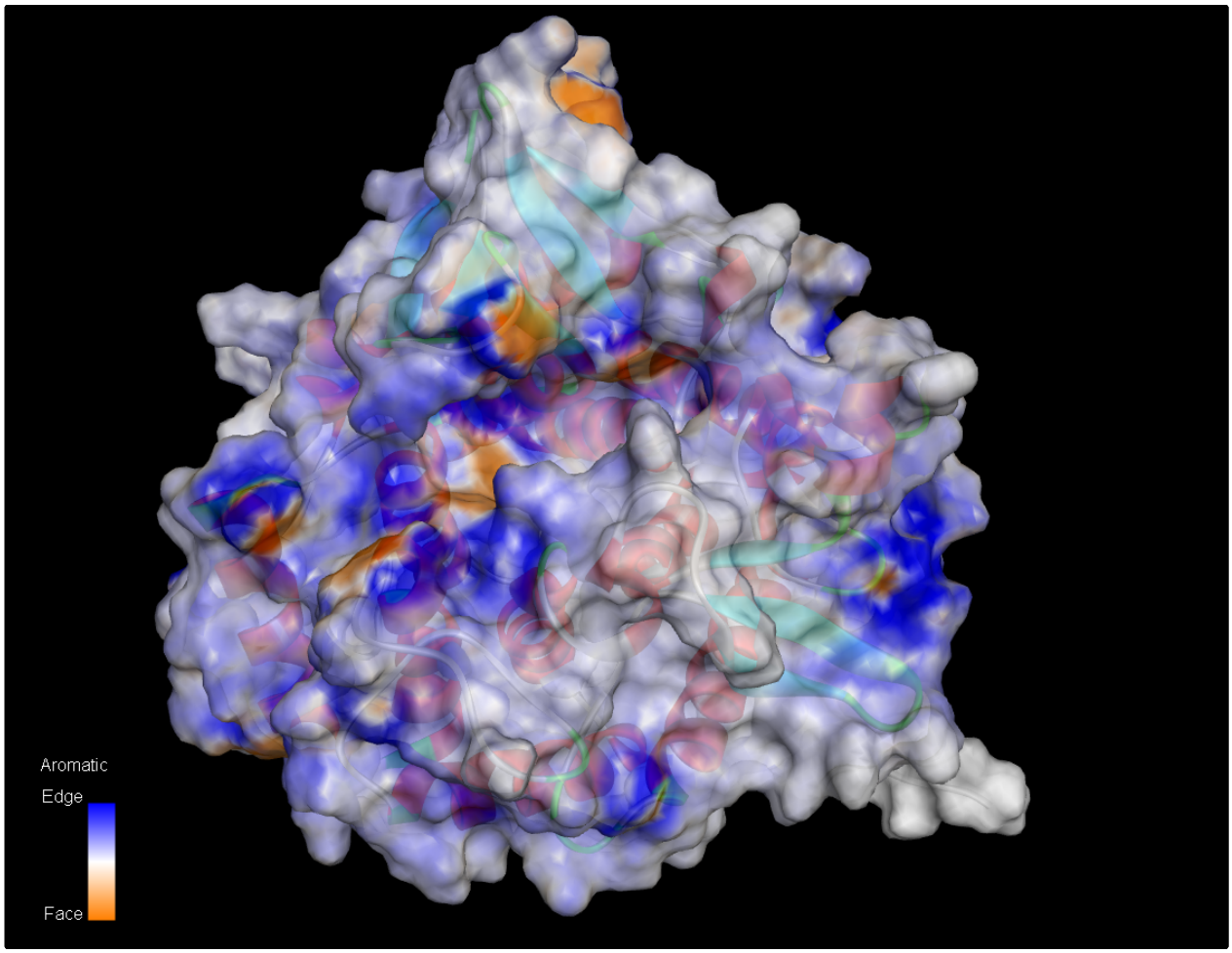
Figure S6: Phylogenetic trees of orthogroups/IPR present in all classes of the sampling but lost in the N16 clade (conserved in the Trebouxiophcyeae clade N15). A. Maximum likelihood phylogenetic tree of the orthogroup N0.HOG0007818 rooted on Prasinoderma coloniale. B. Maximum likelihood phylogenetic tree of the IPR039986 (Coiled-coil domain-containing protein 173).
On each tree, the lichen-forming species are indicated in bold. Black squares mark the proteins from the orthogroup as reconstructed by OrthoFinder for tree concerning orthogroups while they indicate the presence of the functional domain (IPR) on trees on functional categories. Black stars mean that the gene is up-regulated during the lichen formation of Trebouxia sp. TZW2008.



Supplementary Figure S7: A. Maximum likelihood tree (WAG+I+G4) of the GH8 subclade from tree presented in Figure 4A. LFA species are indicated in bold. Blue and red colored branches indicate Trebouxiophyceae and Ulvophyceae respectively. Tree is rooted on a close relative fungi (in bmack). B. Genomic context on a 100kb region centered on algae GH8 (coral bars). Other genes in these regions are marked with cyan bars. C. Reads mapping on a 100kb region of *Trebouxia sp.* OTU5 and *Coccomyxa pringsheimii*. D. Alignment of GH8 subfamilies catalytic site highlighted by the red square). Bacterial sequences are from Adachi et al., 2004.



Supplementary Figure S8: AlphaFold prediction of the *A. glomerata* GH8. A. 3D structure of the GH8 with the active catalytic site indicated with a white arrow. B. Aromatic structure of the GH8 with the putative binding pouch containing the active catalityc site.

A

B