bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Phylogenomics reveals the evolutionary origin of lichenization in chlorophyte algae

Jean Keller[1], Camille Puginier[1], Cyril Libourel[1], J. Otte[2], P. Skaloud[3], Pierre-Marc Delaux[1]*, Francesco Dal Grande[2,4]*

[1] Laboratoire de Recherche en Sciences Végétales (LRSV), Université de Toulouse, CNRS, UPS, INP Toulouse, 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] LOEWE Centre for Translational Biodiversity Genomics (TBG), Senckenberganlage 25, 60325 Frankfurt am Main, Germany

*To whom correspondence may be addressed: pierre-marc.delaux@lrsv.ups-tlse.fr; francesco.dalgrande@senckenberg.de

Author Contributions:

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

JK, FDG, CP, CL, PS and JO conducted experiments

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

P-M.D 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

Abstract

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 15 lichen-forming and closely-related non-lichen-forming algae and performed comparative phylogenomics with 22 genomes previously generated. We identified more than 350 functional categories significantly enriched in chlorophyte green algae able to form lichens. Among them, functions such as light perception or resistance to dehydration were shared between lichenizing and other terrestrial algae but lost in non-terrestrial ones, indicating that the ability to live in terrestrial habitats is a prerequisite for lichens to evolve. We detected lichen-specific expansions of glycosyl hydrolase gene families known to remodel cell walls, including the glycosyl hydrolase 8 which 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. We conclude that the

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

lichen-forming ability evolved multiple times in chlorophyte green algae, following a two-step mechanism which involves an ancestral adaptation to terrestrial lifestyle and molecular innovations to modify the partners' cell walls.

Significance Statement

Mutualistic symbioses have contributed to major transitions in the evolution of life and are at the center of extant ecosystems. How these symbiotic associations evolve and function are central questions in biology. Here, we sequenced and compared the genomes of green algal symbionts of the emblematic lichen symbiosis. We discovered functional features specifically expanded in lichen-forming algae suggesting the evolution of a terrestrial lifestyle as a prerequisite for lichen evolution. Projecting the lichen-specific function on the green algae phylogeny support the independent gain of the ability to form lichens in algae, through gene-family expansions and ancient horizontal gene transfer.

Main Text

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 85% 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 lichens 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). 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, 10). The comparison of genomes in a defined phylogenetic context, an approach known as comparative phylogenomics, has successfully unraveled the evolution of plant mutualistic symbioses with complex evolutionary histories (11–13). In addition, such approaches have the potential to shed light into the molecular mechanisms associated with major innovations, including symbioses (14–17). 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, 11, 12, 15), or multiple independent and convergent gains. The limited availability of LFA genomes constrained so far molecular analyses to single algal species such as *Asterochloris glomerata* and *Trebouxia sp.* TZW2008 (18, 19). The evolutionary history of lichens and the molecular mechanisms linked to the ability to form lichens in green algae remain thus elusive. In this manuscript, we sequenced the genomes of LFA and closely related non-Lichen-Forming Algae (nLFA) and used comparative phylogenomics to discover molecular processes

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

associated with the ability to lichenize in algae. 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. Furthermore, our results support that the ability to live in terrestrial environments might have predisposed LFA to lichenization.

Results

General genome characteristics do not differ between lichen-forming and non-lichen-forming chlorophyte algae. Using the PacBio platform we sequenced the genomes of six LFA belonging to the Trebouxiales, Botryococcus- and Apatococcus-clades (Fig. 1 and *SI Appendix,*

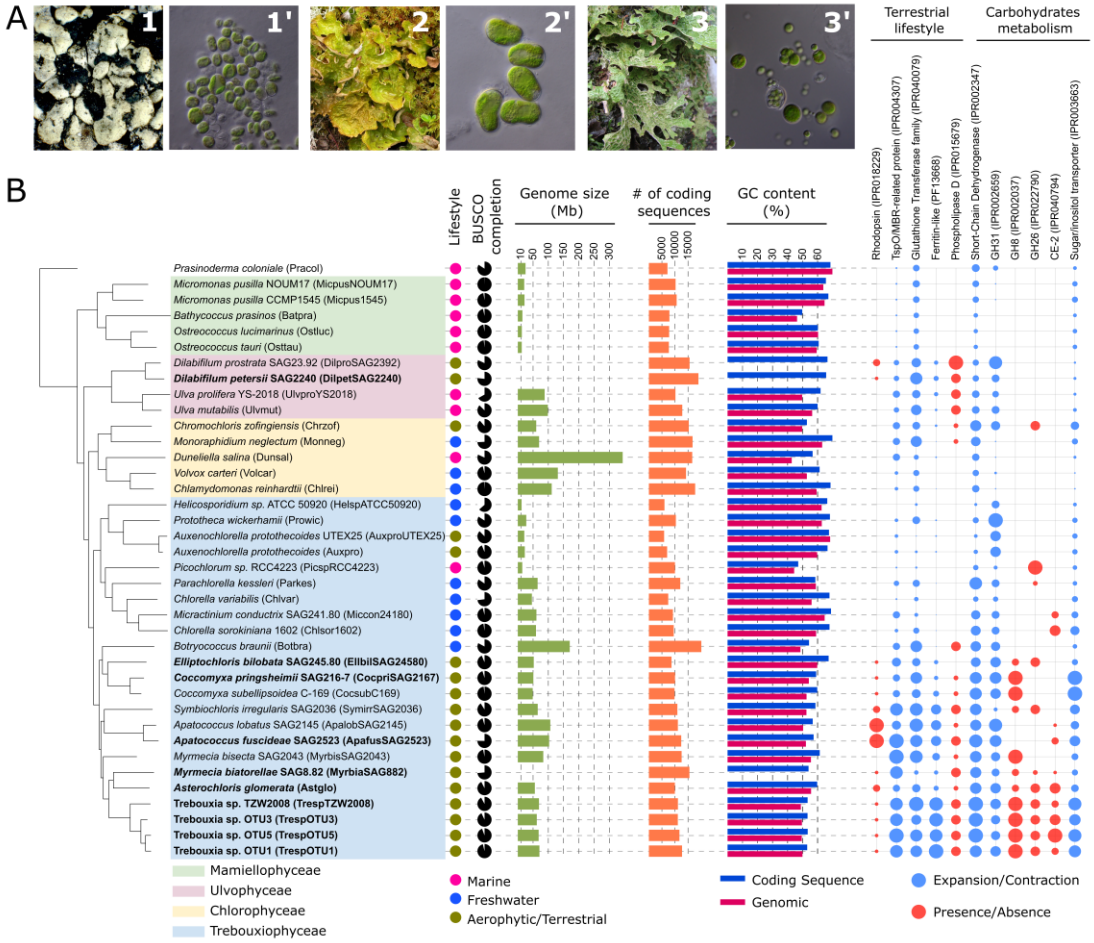


Figure 1. Lichens sampling, evolution, and distribution of habitats, genomic parameters and functional families related to lichen evolution. A. Pictures of three lichens and their algal partner: *Catolechia wahlenbergii* (1) and *Elliptochloris bilobata* SAG245.80 (1'), *Lobaria linita* (2) and *Myrmecia 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% completion level), size of the genome in Mb (empty rows are for species with only transcriptomic data), number of protein-coding genes, GC content for protein-coding genes (blue bars) and genomes (red bars), mapping of functional family sizes significantly more present in LFA compared to nLFA. Size is normalized for each column and are not relative between columns. Bold: lichen-forming algae.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S1). In parallel, we sequenced three closely-related nLFA including species from the *Symbiochloris* and *Myrmecia* genera for which no genomes were available (Fig. 1 and *SI Appendix,* Table S1). Assemblies for eight of the nine species displayed high quality statistics with an average N50 of almost 2Mb, and an average of only 143 scaffolds (Fig. 1 and *SI Appendix,* Table S2). The ninth assembly (*Apatococcus fuscidae* SAG2523) displayed a slightly lower assembly N50 (50 kb) and a higher number of scaffolds (2,319, Fig. 1 and *SI Appendix,* Table S2). To complete this dataset, the transcriptome of six additional species, including the LFA Ulvophyceae *Dilabifilum* (*Paulbroadya) petersii*, were sequenced on an Illumina NovaSeq platform yielding an average of 40 million reads (details of sequencing and assembly statistics in *SI Appendix,* Table S3). Finally, we collected already available genome assemblies for 22 species covering other classes of chlorophyte green algae, and re-annotated six of them (*SI Appendix,* Table S1). Completion of the predicted proteomes estimated for all species using BUSCO ranged from 18.3% to 99.6% (Fig. 1 and *SI Appendix,* Fig. 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 (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 (20–23). We compared LFA and nLFA genome size, GC content, coding sequence number and the number of putatively secreted proteins. A significant difference was observed for the GC content of coding sequences which was lower in LFA compared to nLFA (*SI Appendix,* Fig. 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 to the lichenization habit (*SI Appendix,* Fig. S2). We did not detect significant differences between LFA and nLFA in the other parameters (*SI Appendix,* Fig. 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 (*SI Appendix,* Fig. S2). Altogether this indicates that general genomic features are similar between LFA and nLFA.

The lichen-forming ability is linked to adaptations to a terrestrial lifestyle. To identify the molecular mechanisms associated with the lichen-forming ability, we compared the abundance of all described InterPro (IPR) and PFAM terms, reflecting biological processes and molecular functions, between LFA and nLFA species. We sorted functional terms in two categories based on their distribution in LFA and nLFA species (see Materials and Methods). The first category encompasses terms specific to LFA or nLFA species (Wilcoxon p-value ≤ 0.01 and Fisher p-value ≤ 0.01), while the second category corresponds to functional terms expanded or contracted in LFA species compared to nLFA species (Wilcoxon p-value ≤ 0.01 and Fisher p-value ≥ 0.01). Based on these statistical comparisons, we identified 302 and 234 IPR terms in significant expansion/contraction (in at least 50% LFA and 50% nLFA) or specific to LFA species, respectively (*SI Appendix,* Table S4). Similarly, 78 and 81 PFAM terms were respectively significantly associated with LFA or in expansion/contraction in these species (*SI Appendix,* Table S5). Together, this indicates that LFA species share specific functional features not shared by nLFA. To further identify among these features those directly associated with the ability to lichenize, we cross-referenced the genes containing the discriminating IPR and PFAM terms with the genes found up-regulated in the LFA species *Trebouxia sp.* TZW2008 (*SI Appendix,* Table S6) in association with its fungal symbiont *Usnea hakonensis* (19). After this cross-referencing, we retained 35 IPR mainly present in LFA, and 115 IPR significantly expanded in LFA (*SI Appendix,* Table S4). Similarly, 16 PFAM categories were almost specific to LFA and up-regulated in the symbiotic state in *Trebouxia* sp. TZW2008, while 27 were expanded in LFA (*SI Appendix,* Table

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

S5). These features are shared by most LFA and thus represent functional modules linked with lichenization, not playing a role at the species-specific level, but rather for lichenization in general.

Several seemingly unrelated IPR and PFAM terms were identified as expanded in or mostly specific to LFA species. Those include, for instance, functions linked to the adaptation to high light intensity, such as gluthatione S-transferase (PF13417, IPR040079, respectively 1.53 and 1.73 times more represented in LFA than nLFA; 20, 21), or rhodopsin (IPR018229 (26)) that was present in all LFA species investigated and only four of the 28 nLFA algae (*SI Appendix,* Table S4/S5). Similarly, we identified several functions linked to drought tolerance. This includes members of the TspO/MBR (PF03073/ IPR004307) family which were identified with a two-fold increase in abundance in LFA compared to nLFA, and has been found induced in drought stress condition as well as essential for adaptation to salt stress in the model moss *Physcomitrella patens* (27). We also identified the PFAM domain PF13668 corresponding to Ferritin-like domain known to improve drought tolerance and previously found expanded in the *T. gelatinosa* genome (28). This functional category is absent from 10 of the 28 nLFA while conserved in all LFA, and is also four times more abundant in LFA than nLFA. The ferritin-like superfamily category was also among the most significantly enriched IPR term (IPR009078) in putatively secreted proteins of LFA species compared to nLFA (Wilcoxon p-value 6.3e-05, *SI Appendix,* Table S4.). Lastly, we detected a significant enrichment of the Phospholipase D (IPR015679) in LFA (9/10 species) compared to nLFA (10/28 species), a function previously linked to the stabilization of membrane during desiccation in the LFA *Asterochloris erici* (29). High light radiations and drought are two of the main challenges faced by organisms transitioning from an aquatic to a terrestrial environment (3). The identification of expansion of gene families and biological functions linked to the terrestrial lifestyle as one of the hallmarks of the LFA genomes suggests that the ability to live in terrestrial habitats could be a major prerequisite for the evolution of the ability to form lichens. To determine the evolutionary origin of the terrestrial lifestyle-related features, we conducted targeted phylogenetic analyses of these five genes (*SI Appendix,* Table S7). Among the five phylogenies, two are unresolved (TspO/MBR and Glutathione transferase) while the other three (Rhodopsin, Ferritin-like and the Phospholipase D) display a similar phylogenetic pattern. All three gene-families were gained in the ancestor of the core chlorophytes and lost in nLFA and aquatic species (Rhodopsin, present in only 3/22 core chlorophytes nLFA; present in 0/14 core non-terrestrial chlorophytes species, Fig. 2A) or expanded in LFA Trebouxiophyceae (Ferritin-like and Phospholipase D). The Ferritin-like evolution is marked by two additional duplications, one with species-specific expansion in the core trebouxiophyceans (Fig. 2B bottom clade) while the other (Fig. 2B upper clade) is only constituted of terrestrial algal species and also exhibits species-specific expansion in the core trebouxiophyceans. The Phospholipade D family shows an expansion in the Ulvophyceae clade, while the Trebouxiophyceae clade only contains species from the core Trebouxiophycean clade (Fig. 2C).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

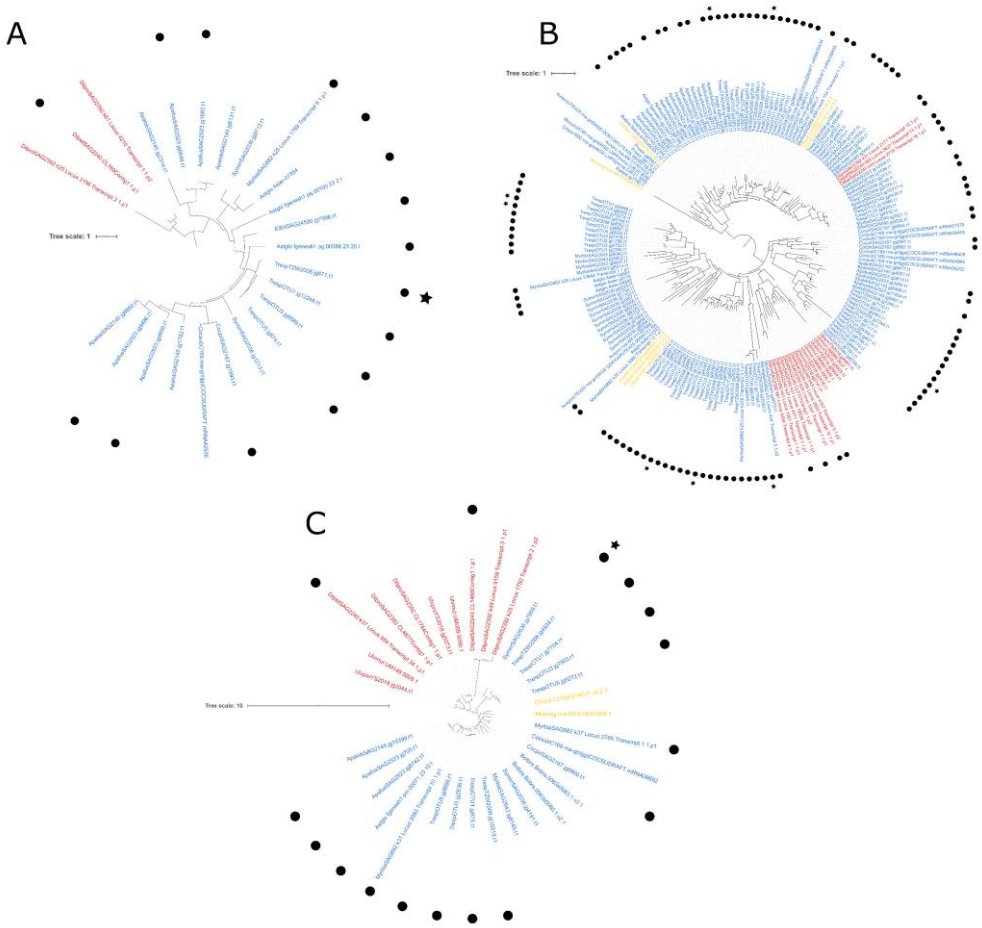


Figure 2. Phylogenies of the terrestrial lifestyle-associated genes. A. Maximum likelihood tree of the Rhodopsin family (IPR018229) rooted on the Ulvophyceae. B. Maximum likelihood tree of the Ferritin-like family (PF13668). Maximum likelihood tree of the PhospholipaseD family (IPR015679) rooted on the Ulvophyceae. Maximum likelihood tree of the GH26 family (IPR0022790). Branches are colored as follow: the Ulvophyceae are in red, the Chlorophyceae in yellow, and the Trebouxiophyceae in blue. The black dots indicate LFA, while the black stars indicate the up-regulated genes during the symbiosis from *Trebouxia* TZW2008.

Altogether, our results indicate that the LFA genomes encompass suits of genes facilitating the terrestrial lifestyle, genes that have an ancient origin in core chlorophyte green algae and have been lost independently in non-terrestrial species.

Gene family expansions and horizontal gene transfer shaped the carbohydrate metabolism and catabolism in LFA species. Manual inspection of the lists of LFA-enriched IPRs and PFAMs identified six main functional categories related to carbohydrate metabolism as significantly expanded (3/6) or present (3/6) in LFA compared to nLFA (*SI Appendix,* Table S4/S5). Among them, we identified a short-chain dehydrogenase (IPR002347/PF00106, in average 1.5-1.6 times more present in LFA) corresponding to a family containing enzymes involved in ribitol biosynthesis. Ribitol is a linear pentose alcohol previously identified as the major sugar produced in diverse lichens such as in *Peltigera aphtosa* (*Coccomyxa* photobiont), *Xanthoria aureola* or *Fulgensia bracteata* (*Trebouxia* spp. photobionts) (30, 31). The addition of exogenous ribitol to culture of

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 this functional category across LFA thus indicates that the expanded ability to produce ribitol is not a species-specific feature but is rather linked with the ability to lichenize itself. Following a similar pattern was the IPR003663 (1.8 times more present in LFA than in nLFA), associated with Sugar/inositol transporter. It is tempting to speculate that such transporters could play a role in the export of ribitol to the fungal partner.

Three Glycosyl Hydrolase gene families were found either expanded in LFA (GH31 IPR002659, 1.6 more times more present in LFA compared to nLFA) or almost specific to LFA (GH8 IPR002037/ PF01270, 8/10 LFA and 3/28 nLFA, Wilcoxon and Fisher p-values: 6.4.10-5 and 0.00013 respectively; GH26 IPR022790, 7/10 LFA and 4/28 nLFA, Wilcoxon and Fisher p-values: 0.0015 and 0.0022 respectively). Glycosyl Hydrolases from these families are known to hydrolase large diversities of polysaccharides including those forming the plant and fungal cell walls (33). In particular, the GH8 enzymes are known to degrade the fungal polysaccharide chitosan and even lichenin, a carbohydrate found in lichens (34). Cell wall remodeling occurs in all the symbioses between plants and microorganisms, facilitating the progression of the microbial partner in the plant tissues in the case of endosymbiosis (35, 36), and observed at the contact sites between symbionts in the case of lichens (37). The occurrence of an expanded and specific repertoire of Glycosyl Hydrolases in LFA echoes the peculiar patterns of retention and losses of cell wall degrading enzymes linked to symbiotic abilities in fungi (14, 17, 38, 39). Finally, a carbohydrate esterase 2 (CE2) N-terminal domain was restricted to LFA (IPR040794/PF17996; 7/10 LFA and 3/28 nLFA, Wilcoxon and Fisher p-values: 0.00027 and 0.00087 respectively) and is known, in bacteria, to be involved in deacetylation of plant polysaccharides (40). Together with the absence of specificities in the Glycosyl Hydrolases repertoire from lichen-forming fungi (41), this indicates that LFA may directly contribute to the remodeling of the partners' cell wall during lichenization.

Phylogenetic analyses of carbohydrate-related IPRs revealed two evolutionary patterns. First, the increased abundance of GH31 and GH26 is linked to gene-family expansions in the Trebouxiophyceae and/or core Trebouxiophycean clade (Figure 3A and B). Such expansion was not observed in the LFA Ulvophyceae *Dilabifilum* (*Paulbroadya) petersii* indicative of clade-specific innovations in the Trebouxiophyceae. The phylogeny of the Glycosyl Hydrolase 8 (GH8) displayed a more unusual pattern indicative of a putative horizontal gene transfer (HGT) that would have occurred from bacteria before the radiation of the Trebouxiophyceae (Figure 3C). In addition to the phylogenetic analysis, this HGT event is supported by the gene positions, anchored in large scaffolds (Fig.3D, *SI Appendix,* Fig. S3A) and by reads mapping that excluded any chimeric assembly (*SI Appendix,* Fig. S3B). In bacteria the GH8 gene family is divided in three subfamilies based on the position of the proton acceptor (42). Alignment of reference proteins for the three subfamilies with the LFA GH8-like sequences revealed that these latter shared the asparagine catalytic site with GH8b, the subfamily that encompasses, among others, licheninase enzymes degrading the lichenin polymer (*SI Appendix,* Fig. S3C, (34, 42). In addition to the Trebouxiophyceae, our phylogenetic analysis supports another HGT event of the bacterial GH8 which has also been transferred to fungi. However, no link has been identified between fungal species possessing the GH8-like gene and species forming lichen associations (*SI Appendix,* Table S8).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

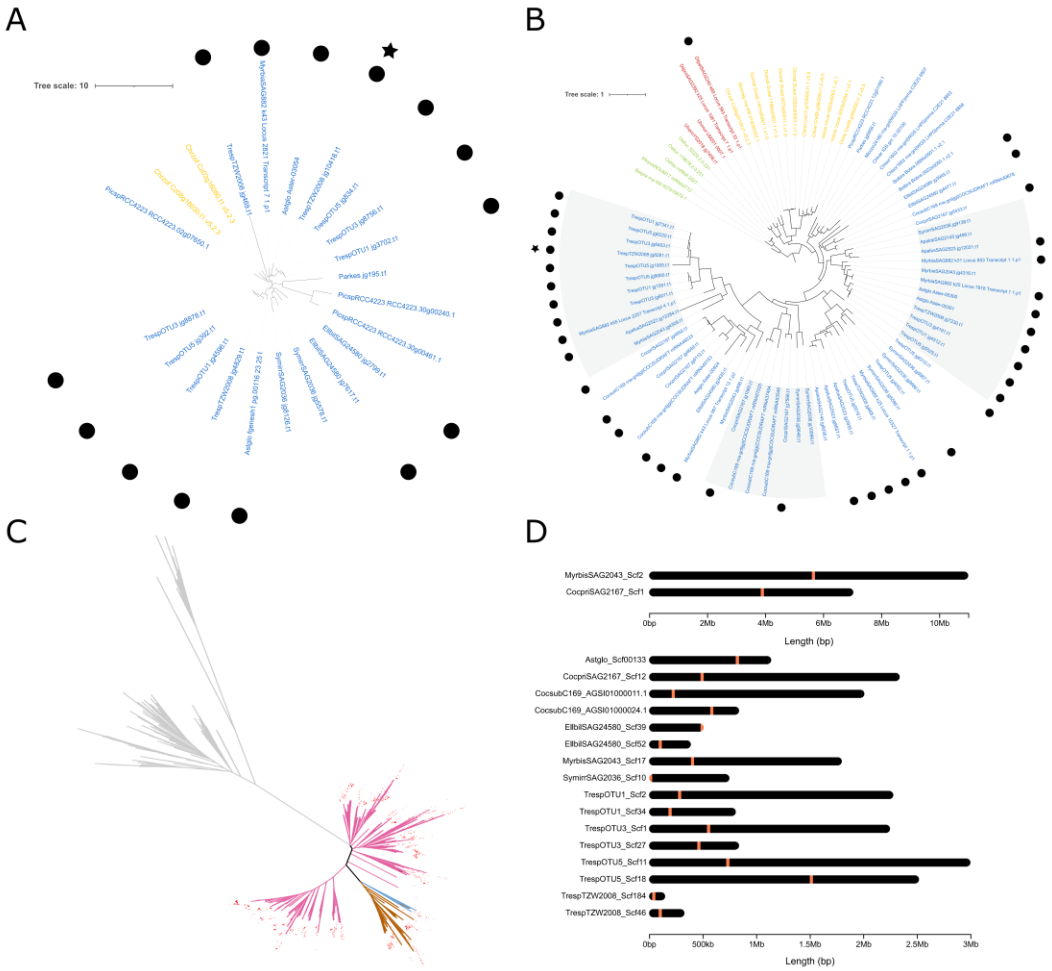


Figure 3. Phylogenie of expanded or LFA-specific functional categories and demonstration ofothe GH8 horizontal gene transfer. A. Maximum likelihood tree of the IPR031155 (Urea active transporter) rooted on the Mamiellophyceae. B. Maximum likelihood tree of the GH26 family (IPR0022790). Branches are colored as follow: the Mamiellophyceae are in green, the Ulvophyceae in red, the Chlorophyceae in yellow, and the Trebouxiophyceae in blue. The grey ranges represent the gene-family expansions. The black dots indicate LFA, while the black stars indicate the up-regulated genes during the symbiosis from Trebouxia TZW2008. C. Unrooteed maximum likelihood tree of GH8 family. Branches are colored as follow: outgroup sequences without the GH8 domains are in grey, while bacteria, fungi and algae GH8 are colored in pink, orange and blue respectively. The red dots indicate the presence of the GH8 functional domain (PF01270). Sequence names have been hidden for readability reasons and can be found in the Supplementary Table S8. D. Position of GH8-like coding-genes of algae on chromosome. For readability reasons, chromosomes longer than 5Mb are displayed separately. The position of the genes is indicated by the coral-colored bar.

Phylogenomics supports at least two independent gains of the lichen-forming lifestyle in chlorophyte algae. The genomic features associated with the ability to form lichens and shared among LFA might either be the result of convergent evolution or the sign of ancestral gain(s) before the diversification of LFA lineages. Both evolutionary patterns seem to emerge from the targeted phylogenies conducted on the terrestrial lifestyle-related genes (ancestral gains, multiple losses) and on the carbohydrate-related genes (clade-specific innovations). To further test these two scenarios, we conducted two additional analyses. First, we ran phylogenetic analyses of all the

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

features (IPR and PFAM) discriminating LFA from nLFA after cross-referencing the functional category enrichment with the genes transcriptionally up-regulated during lichenization in *Trebouxia* sp. TZW2008 (*SI Appendix,* Table S6). All the generated phylogenies were mined to identify either gene-family expansion in LFA or convergent gene losses in nLFA from the Ulvophyceae, Chlorophyceae and Trebouxiophyceae. Out of the 89 resolved phylogenies, a large majority indicated gene-family expansions at the Trebouxiophyceae (9/89), core Trebouxiophycean (43/89) or both (3/89) nodes. Eight other phylogenies displayed a gain in the core Trebouxiophycean clade. By contrast with these clade-specific innovations, only two phylogenies showed convergent gene losses in nLFA from the Trebouxiophyceae, the Ulvophyceae and the Chlorophyceae (*SI Appendix,* Table S7 and Data S1). This first layer of analysis thus does not support the ancestral gain hypothesis of the ability to lichenize in chlorophytes.

To further test the ancestral gain hypothesis, we reconstructed orthogroups for all 38 genomes. A total of 28,454 phylogenetic hierarchical orthogroup (HOG, *SI Appendix,* Table S9) containing 86.9% of the 434,648 proteins from the 38 species were obtained. The phylogenetic tree calculated by OrthoFinder based on 626 orthogroups containing all species was re-rooted on the outgroup species *Prasinoderma coloniale* (43) and was consistent with the already published phylogenies (43, 44). We parsed the orthogroups to sort those with a Wilcoxon p-value ≤ 0.01 and a Fisher p-value ≤ 0.01 and extracted those containing genes transcriptionally up-regulated during lichenization in *Trebouxia* sp. TZW2008 (*SI Appendix,* Table S6). Except for the terrestrial lifestyle-related genes previously identified, all other phylogenetic analyses on the hOG do not support the hypothesis of a single ancestral gain of the ability to lichenize in chlorophytes (*i.e.* in the most recent common ancestor of all the chlorophyte LFA). By contrast, the analyses rather support at least two independent gains, one at the base of the Trebouxiophyceae or core Trebouxiophycean clade, and one in the Ulvophyceae.

The lichen-forming lifestyle is not linked to metabolic streamlining of the chlorophyte symbiont. Co-evolution between symbiotic partners may lead to metabolic streamlining, where the symbiont entirely relies on the other partner for the biosynthesis of primary metabolites leading to relaxed selection and loss of the associated enzymes in the symbiont genome. Such examples have been described in vertically transmitted symbionts of animals (22) or plants (23), but also in the case of horizontally transmitted symbionts such as the arbuscular mycorrhizal fungi in land plants (45, 46). Our results support that the ability to lichenize is ancestral in the Trebouxiophyceae. To test whether the ability of establishing lichen symbiosis in that clade is associated with shared losses, we searched for functional categories significantly contracted or absent in LFA species. A total of 10 IPR and 42 PFAM functional categories are significantly contracted in LFA, while 15 PFAM and 74 and IPR terms present in nLFA are absent from most LFA (*SI Appendix,* Table S4/S5). Among these functional categories, no clear evidence for a shared metabolic streamlining could be identified across the LFA species. The absence of convergent losses, even in the *Trebouxia* genus which contains only LFA species, suggests that the lichen association is not obligate, a fact supported by the ability of the LFA species to grow in absence of lichen-forming fungal hosts.

Discussion

Lichens evolution has been mainly unsymmetrically studied from the fungal partner side. These studies have revealed the convergent evolution of lichens ability resulting in over 20,000 fungal species able to form this symbiosis (9, 47). 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

chlorophyte green algae and possibly other eukaryotic symbionts (48, 49). The chlorophyte green algae encompass more than 8,650 species divided in 13 classes (50). LFA have been described in two of these classes, the Trebouxiophyceae and the Ulvophyceae (AlgaeBase, accessed 12/16/2021). 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. This later evolutionary pattern has been observed in embryophytes for the arbuscular mycorrhizal symbiosis and the nitrogen-fixing root nodule symbiosis (11, 12). 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 (51).

A common feature of lichenizing Ulvophyceae and Trebouxiophyceae is their ability to thrive in terrestrial habitats, which we found mirrored by the expansion of molecular mechanisms linked to drought adaptation. When projecting these processes on the phylogeny, it appeared that other chlorophyte green algae with a terrestrial lifestyle, such as *Chromochloris zofingiensis*, also shared these molecular adaptations which have been lost in non-terrestrial chlorophyte algae. Besides their impact on our understanding of lichenization, these results lay the ground for a completely novel and provocative hypothesis: the most recent common ancestor of extant chlorophyte green algae may have been equipped for the terrestrial lifestyle. This hypothesis will require comparative phylogenomic approaches on a dedicated set of species to be further explored. Thus, we propose that the adaptation to a terrestrial lifestyle is a prerequisite, a potentiation (52), for lichens to evolve.

Following potentiation, trait evolution is predicted to be followed by actualization, the actual expression of the trait, even at a low-efficiency rate (52). Here, comparing a large diversity of LFA and close relative nLFA species we can propose molecular mechanisms associated with the actualization of lichen formation (Fig. 4). We discovered shared expansions of gene families in LFA, and LFA-specific genes. Among them, we identified the glycosyl hydrolase GH26, GH31 and GH8 which are possibly involved in cell wall remodeling process, a critical step of plant–microbe interactions (36). 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 (53). 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 identified enzymes, especially the GH8, are required to conclude on the nature of the targeted polysaccharide. Along with glycosyl hydrolases, we also identified the expansions of the short-chain dehydrogenase gene family putatively involved in the biosynthesis and excretion 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* (18, 19).

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 reinhardtiii.*

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

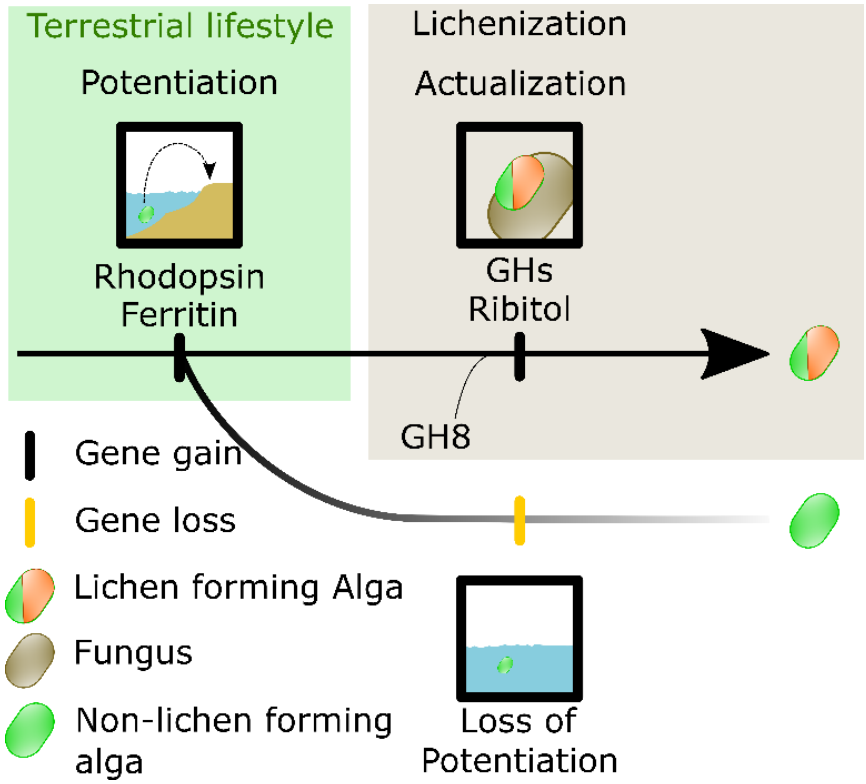


Figure 4. Model of lichens evolution in chlorophyte algae

The molecular functions and biological processes that we found recruited here for the lichen symbiosis in chlorophyte algae are very different from the ones identified for the evolution of other plant – fungi symbioses (1, 15, 56). However, a number of general rules seems to emerge, such as the production of chemical signals as symbiotic cue. While LFA have increased their ability to produce (Short-chain dehydrogenase) and possibly secrete (Sugar/inositol transporter) ribitol, embryophytes have evolved strigolactones to activate the fungal metabolism (57, 58). Similar to the expanded repertoire of glycosyl hydrolases in LFA, embryophytes have evolved ways to remodel cell walls (35). The last commonality between fungi – embryophytes and lichen symbioses is in the actual evolutionary mechanisms behind the evolution of the traits. Indeed, echoing our finding that horizontal gene transfer participated in shaping the *lichenization toolkit* (*e.g.* GH8), the evolution of the GRAS family of transcription factors, which are well-known regulators of the arbuscular mycorrhizal symbiosis in embryophytes, stems from a HGT (59). The repeated occurrence of such major transfers between species during the evolution of mutualistic symbioses supports the emerging concept that synthetic biology approaches to transfer entire gene pathways pave the road toward the successful deployment of novel symbiotic option in agriculture (60).

Materials and Methods

Algal cultures

Three *Trebouxia* photobionts with varied ecologies (61) were isolated from thalli of the lichen *Umbilicaria pustulata* using a micro-manipulator as described in (62). Cultures were grown on solid 3N BBM + V medium (Bold's Basal Medium with vitamins and triple nitrate (63, 64) under a 30 µmol/m−2s−1 photosynthetic photon flux density with a 12 h photoperiod at 16°C. The identity of

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the isolated photobionts was validated by comparing the ITS sequence to those from (61). 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 (65). Green algae were transferred to fresh agar plates two days before nuclei isolation. For this, we used a modified protocol by Nishi et al. (2019) 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% ß-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% ß-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 (41) with modifications described in (42) 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 (66). 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 'continuous long read' (i.e., CLR) 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).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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) subreads set using the genome assembler Flye (version 2.7-b1587) (67) 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 (68) with default parameters.

The received scaffolds were taxonomically binned via BLASTx against the NCBI nr database using DIAMOND (--more-sensitive) in MEGAN v.6.7.7 (69), using Max Expected set to 1E-10 and the MEGAN-LR algorithm (Huson et al. 2018) and 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 green algal genomes already published without structural annotations (*SI Appendix,* Table S1).

Genome assemblies were softmasked using Red (70) and annotated using BRAKER2 pipeline (71–79). 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 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). The raw fastq reads were cleaned from adapters and the low-quality sequences using cutadapt v2.1, (80) and TrimGalore v0.6.5, (https://github.com/FelixKrueger/TrimGalore) with -q 30 --length 20 options. The cleaned reads were mapped against the corresponding genomes using HISAT2 v2.1.0 (81) with --score-min L,-0.6,-0.6 --max-intronlen 10000 --dta options. Duplicated reads were removed using SAMtools v1.10 (82) , markdup command. 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 (83). runDrap was first used on the unique samples applying the Oases RNAseq assembly software (84). 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) and hits from BLASTp against the Swissprot database (downloaded on September 2021) and HMMER search against the Pfam v34 database (85, 86). Completeness of newly sequenced and annotated genomes and transcriptomes was assessed using BUSCO V4.1.4 (87) 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 (88) with the following analysis enabled: PFAM, ProSite profiles and patterns, Panther, TIGERFAMS, CATH-Gene3D, CDD, HAMAP, PIRSF, PRINTS, SFLD, SMART and SUPERFAMILY.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Identification of putative secreted proteins

Identification of putative secreted proteins were conducted using a combination of approaches. First, proteins shorter than 300 amino acids were selected for each species. Then, this set of proteins were subjected to a series of analysis. Signal peptides were predicted with Signalp5 (89), Phobius v1.01 (90) and TargetP v2.0 (91). In addition, transmembrane domains were identified using TMHMM 2.0c (92). Finally, the presence of the endoplasmic reticulum signal sequence (KDEL/HDEL) was verified using ps_scan script v1.86 from the ProSite suite (93) to search the PS00014 motif. All programs were used with default parameters. To be considered as putatively secreted, short proteins should have a predicted signal peptide predicted by the three different methods used, not harbor a signal sequence for endoplasmic reticulum and do not have transmembrane domain predicted.

Orthogroups reconstruction

Orthogroups reconstruction was performed using OrthoFinder v2.5.2 (94) 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-ran with this correctly rooted tree and the MSA option to improve the orthogroups reconstruction.

Assessing the ancestral gain hypothesis

One of the evolutionary hypotheses is that the ability to form lichens was acquired in the common ancestor of Trebouxiophyceae and Ulvophyceae. To identify orthogroups associated with the lichen-forming ability that could represent molecular evidence in favour the ancestral gain hypothesis, each orthogroup was submitted to a statistical approach: a Wilcoxon test was coupled with a Fisher exact test. Orthogroups with a Wilcoxon p-value < 0.01 and a Fisher p-value < 0.01 were analysed. To empower this statistical approach, only orthogroups with the 10 LFA and a differentially expressed gene (DEG) between the symbiotic state (natural and/or co-culture) and the aposymbiotic state of the photobiont *Trebouxia sp* TZW2008 were analysed.

Identification of contraction or expansion of molecular functions

Each PFAM and IPR domains were submitted to the same statistical approach to detect those for which there is a significant difference in the number of sequences between the LFA and the nLFA (Wilcoxon) and for which this difference is due to the number of gene copies associated with the PFAM/IPR (Fisher exact test). Thus, PFAM/IPR with a Wilcoxon pvalue < 0.01, a Fisher pvalue > 0.01 and present in at least half of LFA and nLFA species were analysed. To empower the statistical approach, transcriptomic data from *Trebouxia sp.* (TZW2008 isolate, (19)) was used to analyse significative Pfam/IPR that also corresponds to a DEG in *Trebouxia sp.* TZW2008.

Identification of functions associated with lichen-forming ability

The same statistical approach is used to detect PFAM or IPR domains linked to the lichenization that could have been commonly lost in nLFA or commonly gained in LFA. To do so, PFAM/IPR with a Wilcoxon p-value < 0.01 and a Fisher p-value < 0.01 were analysed. Here again, transcriptomic data from *Trebouxia* sp TZW2008 was used to identify IPR/Pfam that contain a DEG in a symbiotic state compared to the aposymbiotic state.

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 (19) with results of our contraction/expansion analysis. To do so, the raw reads were downloaded (*SI Appendix,* Table S1)

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and submitted to the nf-core/rnaseq v3.4 (95) workflow in nextflow v21.04 (96) using *-profile debug,genotoul --skip_qc --aligner star_salmon* options. Nextflow nf-core rnaseq workflow used *bedtools* v2.30.0 (97), *cutadapt* v3.4 (80) implemented in *TrimGalore!* v 0.6.7 (https://github.com/FelixKrueger/TrimGalore), *picard* v2.25.7 (https://broadinstitute.github.io/picard), *salmon* v1.5.2 (98), *samtools* v1.13 (74), *star* v2.6.1d and v2.7.6a (99), *stringtie* v2.1.7 (100) and *UCSC tools* v377 (101).

The counted data were analysed using *edgeR* package v3.36.0 (102) with R v4.1.1 (103). 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.

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 (105) and an e-value threshold of 1e-30. Then, retained sequences were aligned using MUSCLE v3.8.1551 (106) or Clustal Omega (107) with default parameters and obtained alignments cleaned with trimAl v1.4.1 (108) 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 v2.1.10 (109) 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 (110). Prior to tree reconstruction, the best fitting evolutionary model was tested using ModelFinder (111). Branch supports were tested using 10,000 replicates of both SH-aLRT (112) and ultrafast bootstrap (113). Trees were visualized and annotated in the iTOL v6 platform (114).

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 has been 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 (115) and default parameters. Finally, a phylogenetic analysis has been conducted to place algal GH8 in an evolutionary context. Using the BLASTp v2.10.1+ algorithm (105) with an e-value threshold of 1e-30 homologs of algal GH8-like were searched against three different database: the JGI fungal resources (accessed in February 2020, contains more than 1600 fungal genomes), the non-redundant protein database from NCBI (BLAST db v5) and the SymDB database covering Viridiplantae species (15). Obtained sequences were subjected to phylogenetic analysis as described above and using Clustal Omega for the alignment step and the fast tree reconstruction option from IQ-TREE. Presence of the GH8 functional domain (PF01270) was determined with hmmsearch function from the HMMER v3.3 package (116) with default parameters.

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(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.

All data were deposited in NCBI under the BioProject PRJNA790449.

References

1. M. K. Rich, *et al.*, Lipid exchanges drove the evolution of mutualism during plant terrestrialization. *Science* **372**, 864–868 (2021).

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

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

4. D. Beerling, *The Emerald Planet: How plants changed Earth's history* (Oxford University Press, 2007) https:/doi.org/10.1093/oso/9780192806024.001.0001 (December 13, 2021).

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

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

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

8. M. L. Berbee, *et al.*, Genomic and fossil windows into the secret lives of the most ancient fungi. *Nat Rev Microbiol* **18**, 717–730 (2020).

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

10. F. Lutzoni, *et al.*, Contemporaneous radiations of fungi and plants linked to symbiosis. *Nat Commun* **9**, 5451 (2018).

11. P.-M. Delaux, *et al.*, Comparative Phylogenomics Uncovers the Impact of Symbiotic Associations on Host Genome Evolution. *PLoS Genet* **10**, e1004487 (2014).

12. M. Griesmann, *et al.*, Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* **361**, eaat1743 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

13. R. van Velzen, *et al.*, Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proc Natl Acad Sci USA* **115**, E4700–E4709 (2018).

14. A. Kohler, *et al.*, Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat Genet* **47**, 410–415 (2015).

15. G. V. Radhakrishnan, *et al.*, An ancestral signalling pathway is conserved in intracellular symbioses-forming plant lineages. *Nat. Plants* **6**, 280–289 (2020).

16. E. Kiss, *et al.*, Comparative genomics reveals the origin of fungal hyphae and multicellularity. *Nat Commun* **10**, 4080 (2019).

17. F. Mesny, *et al.*, Genetic determinants of endophytism in the Arabidopsis root mycobiome. *Nat Commun* **12**, 7227 (2021).

18. D. Armaleo, *et al.*, The lichen symbiosis re-viewed through the genomes of Cladonia grayi and its algal partner Asterochloris glomerata. *BMC Genomics* **20**, 605 (2019).

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

20. S. Noh, Linear paths to genome reduction in a defensive symbiont. *Proc Natl Acad Sci USA* **118**, e2106280118 (2021).

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

22. T. Y. Nechitaylo, *et al.*, Incipient genome erosion and metabolic streamlining for antibiotic production in a defensive symbiont. *Proc Natl Acad Sci USA* **118**, e2023047118 (2021).

23. L. Ran, *et al.*, Genome Erosion in a Nitrogen-Fixing Vertically Transmitted Endosymbiotic Multicellular Cyanobacterium. *PLoS ONE* **5**, e11486 (2010).

24. Á. Gallé, *et al.*, Plant Glutathione Transferases and Light. *Front. Plant Sci.* **9**, 1944 (2019).

25. K. Balarinová, M. Barták, J. Hazdrová, J. Hájek, J. Jílková, Changes in photosynthesis, pigment composition and glutathione contents in two Antarctic lichens during a light stress and recovery. *Photosynt.* **52**, 538–547 (2014).

26. M. Awasthi, K. Sushmita, M. S. Kaushik, P. Ranjan, S. Kateriya, Novel Modular Rhodopsins from Green Algae Hold Great Potential for Cellular Optogenetic Modulation Across the Biological Model Systems. *Life* **10**, 259 (2020).

27. W. Frank, *et al.*, A mitochondrial protein homologous to the mammalian peripheral-type benzodiazepine receptor is essential for stress adaptation in plants: Knock-out of a TSPO homologue in moss. *The Plant Journal* **51**, 1004–1018 (2007).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

28. F. C. Carniel, *et al.*, New features of desiccation tolerance in the lichen photobiont Trebouxia gelatinosa are revealed by a transcriptomic approach. *Plant Mol Biol* **91**, 319–339 (2016).

29. F. Gasulla, *et al.*, The Role of Phospholipase D and MAPK Signaling Cascades in the Adaption of Lichen Microalgae to Desiccation: Changes in Membrane Lipids and Phosphoproteome. *Plant Cell Physiol* **57**, 1908–1920 (2016).

30. K. Palmqvist, Carbon economy in lichens. *New Phytologist* **148**, 11–36 (2000).

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

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

33. Z. Minic, L. Jouanin, Plant glycoside hydrolases involved in cell wall polysaccharide degradation. *Plant Physiology and Biochemistry* **44**, 435–449 (2006).

34. A. S. Perlin, S. Suzuki, The Structure of Lichenin: Selective Enzymolysis studies. *Can. J. Chem.* **40**, 50–56 (1962).

35. R. Balestrini, P. Bonfante, Cell wall remodeling in mycorrhizal symbiosis: a way towards biotrophism. *Front. Plant Sci.* **5** (2014).

36. N. J. Brewin, Plant Cell Wall Remodelling in the Rhizobium–Legume Symbiosis. *Critical Reviews in Plant Sciences* **23**, 293–316 (2004).

37. R. L. Chapman, Ultrastructural investigation on the foliicolous pyrenocarpous lichen Strigula elegans (Fée) Müll. Arg. *Phycologia* **15**, 191–196 (1976).

38. E. Tisserant, *et al.*, Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proceedings of the National Academy of Sciences* **110**, 20117–20122 (2013).

39. E. Martino, *et al.*, Comparative genomics and transcriptomics depict ericoid mycorrhizal fungi as versatile saprotrophs and plant mutualists. *New Phytol* **217**, 1213–1229 (2018).

40. C. Montanier, *et al.*, The Active Site of a Carbohydrate Esterase Displays Divergent Catalytic and Noncatalytic Binding Functions. *PLoS Biol* **7**, e1000071 (2009).

41. P. Resl, *et al.*, "Large differences in carbohydrate degradation and transport potential in the genomes of lichen fungal symbionts" (Genomics, 2021) https:/doi.org/10.1101/2021.08.01.454614 (December 13, 2021).

42. W. Adachi, *et al.*, Crystal Structure of Family GH-8 Chitosanase with Subclass II Specificity from Bacillus sp. K17. *Journal of Molecular Biology* **343**, 785–795 (2004).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

43. L. Li, *et al.*, The genome of Prasinoderma coloniale unveils the existence of a third phylum within green plants. *Nat Ecol Evol* **4**, 1220–1231 (2020).

44. F. Leliaert, "Green Algae: Chlorophyta and Streptophyta" in *Reference Module in Life Sciences*, (Elsevier, 2019), p. B978012809633820890X.

45. M. Malar C, *et al.*, The genome of Geosiphon pyriformis reveals ancestral traits linked to the emergence of the arbuscular mycorrhizal symbiosis. *Current Biology* **31**, 1570-1577.e4 (2021).

46. V. Wewer, M. Brands, P. Dörmann, Fatty acid synthesis and lipid metabolism in the obligate biotrophic fungus *Rhizophagus irregularis* during mycorrhization of *Lotus japonicus*. *Plant J* **79**, 398–412 (2014).

47. A. Gargas, P. T. DePriest, M. Grube, A. Tehler, Multiple Origins of Lichen Symbioses in Fungi Suggested by SSU rDNA Phylogeny. *Science* **268**, 1492–1495 (1995).

48. M. P. Nelsen, R. Lücking, C. K. Boyce, H. T. Lumbsch, R. H. Ree, The macroevolutionary dynamics of symbiotic and phenotypic diversification in lichens. *Proc Natl Acad Sci USA* **117**, 21495–21503 (2020).

49. P. K. Divakar, *et al.*, Evolution of complex symbiotic relationships in a morphologically derived family of lichen-forming fungi. *New Phytol* **208**, 1217–1226 (2015).

50. E. R. Hanschen, S. R. Starkenburg, The state of algal genome quality and diversity. *Algal Research* **50**, 101968 (2020).

51. C. Lemieux, C. Otis, M. Turmel, Chloroplast phylogenomic analysis resolves deep-level relationships within the green algal class Trebouxiophyceae. *BMC Evol Biol* **14**, 211 (2014).

52. Z. D. Blount, J. E. Barrick, C. J. Davidson, R. E. Lenski, Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature* **489**, 513–518 (2012).

53. E. F. Y. Hom, A. W. Murray, Niche engineering demonstrates a latent capacity for fungal-algal mutualism. *Science* **345**, 94–98 (2014).

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

55. H. Tatara, *et al.*, A method for the preparation of electrocompetent cells to transform unicellular green algae, Coccomyxa (Trebouxiophyceae, Chlorophyta) strains Obi and KJ. *Algal Research* **48**, 101904 (2020).

56. P.-M. Delaux, *et al.*, Algal ancestor of land plants was preadapted for symbiosis. *Proc Natl Acad Sci USA* **112**, 13390–13395 (2015).

57. A. Besserer, *et al.*, Strigolactones Stimulate Arbuscular Mycorrhizal Fungi by Activating Mitochondria. *PLoS Biol* **4**, e226 (2006).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

58. K. Kodama, *et al.*, "An Ancestral Function of Strigolactones as Symbiotic Rhizosphere Signals" (Plant Biology, 2021) https:/doi.org/10.1101/2021.08.20.457034 (December 13, 2021).

59. S. Cheng, *et al.*, Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant Evolution. *Cell* **179**, 1057-1067.e14 (2019).

60. C. Rogers, G. E. D. Oldroyd, Synthetic biology approaches to engineering the nitrogen symbiosis in cereals. *Journal of Experimental Botany* **65**, 1939–1946 (2014).

61. F. Dal Grande, *et al.*, Environment and host identity structure communities of green algal symbionts in lichens. *New Phytol* **217**, 277–289 (2018).

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

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

64. R. C. Starr, J. A. Zeikus, UTEX-The Culture Collection of Algae at the University of Texas at Austin. *J Phycol* **29**, 1–106 (1993).

65. K. Bethune, *et al.*, Long-fragment targeted capture for long-read sequencing of plastomes. *Appl Plant Sci* **7**, e1243 (2019).

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

67. M. Kolmogorov, J. Yuan, Y. Lin, P. A. Pevzner, Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37**, 540–546 (2019).

68. M. Boetzer, W. Pirovano, SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* **15**, 211 (2014).

69. D. H. Huson, *et al.*, MEGAN-LR: new algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. *Biol Direct* **13**, 6 (2018).

70. H. Z. Girgis, Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. *BMC Bioinformatics* **16**, 227 (2015).

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

74. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

75. J. Keilwagen, F. Hartung, M. Paulini, S. O. Twardziok, J. Grau, Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics* **19**, 189 (2018).

76. J. Keilwagen, *et al.*, Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res* **44**, e89–e89 (2016).

77. A. Lomsadze, P. D. Burns, M. Borodovsky, Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Research* **42**, e119–e119 (2014).

78. O. Gotoh, A space-efficient and accurate method for mapping and aligning cDNA sequences onto genomic sequence. *Nucleic Acids Research* **36**, 2630–2638 (2008).

79. M. Stanke, O. Schöffmann, B. Morgenstern, S. Waack, Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62 (2006).

80. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* **17**, 10 (2011).

81. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).

82. P. Danecek, *et al.*, Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).

83. C. Cabau, *et al.*, Compacting and correcting Trinity and Oases RNA-Seq *de novo* assemblies. *PeerJ* **5**, e2988 (2017).

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

85. S. R. Eddy, Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).

86. J. Mistry, *et al.*, Pfam: The protein families database in 2021. *Nucleic Acids Research* **49**, D412–D419 (2021).

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

88. P. Jones, *et al.*, InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

89. J. J. A. Armenteros, *et al.*, SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**, 420–423 (2019).

90. L. Käll, A. Krogh, E. L. L. Sonnhammer, A Combined Transmembrane Topology and Signal Peptide Prediction Method. *Journal of Molecular Biology* **338**, 1027–1036 (2004).

91. J. J. A. Armenteros, *et al.*, Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* **2**, e201900429 (2019).

92. A. Krogh, B. Larsson, G. von Heijne, E. L. L. Sonnhammer, Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. *Journal of Molecular Biology* **305**, 567–580 (2001).

93. E. de Castro, *et al.*, ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research* **34**, W362–W365 (2006).

94. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238 (2019).

95. P. A. Ewels, *et al.*, The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* **38**, 276–278 (2020).

96. P. Di Tommaso, *et al.*, Nextflow enables reproducible computational workflows. *Nat Biotechnol* **35**, 316–319 (2017).

97. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

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

99. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

100. S. Kovaka, *et al.*, Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol* **20**, 278 (2019).

101. W. J. Kent, A. S. Zweig, G. Barber, A. S. Hinrichs, D. Karolchik, BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).

102. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

104. M. D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25 (2010).

105. C. Camacho, *et al.*, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

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

107. F. Sievers, *et al.*, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).

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

109. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **5**, e9490 (2010).

110. B. Q. Minh, *et al.*, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530–1534 (2020).

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

112. S. Guindon, *et al.*, New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59**, 307–321 (2010).

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

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

115. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

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

Figures

Figure 1. A. Pictures of three lichens and their algal partner: *Catolechia wahlenbergii* (1) and *Elliptochloris bilobata* SAG24580 (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 the distant species

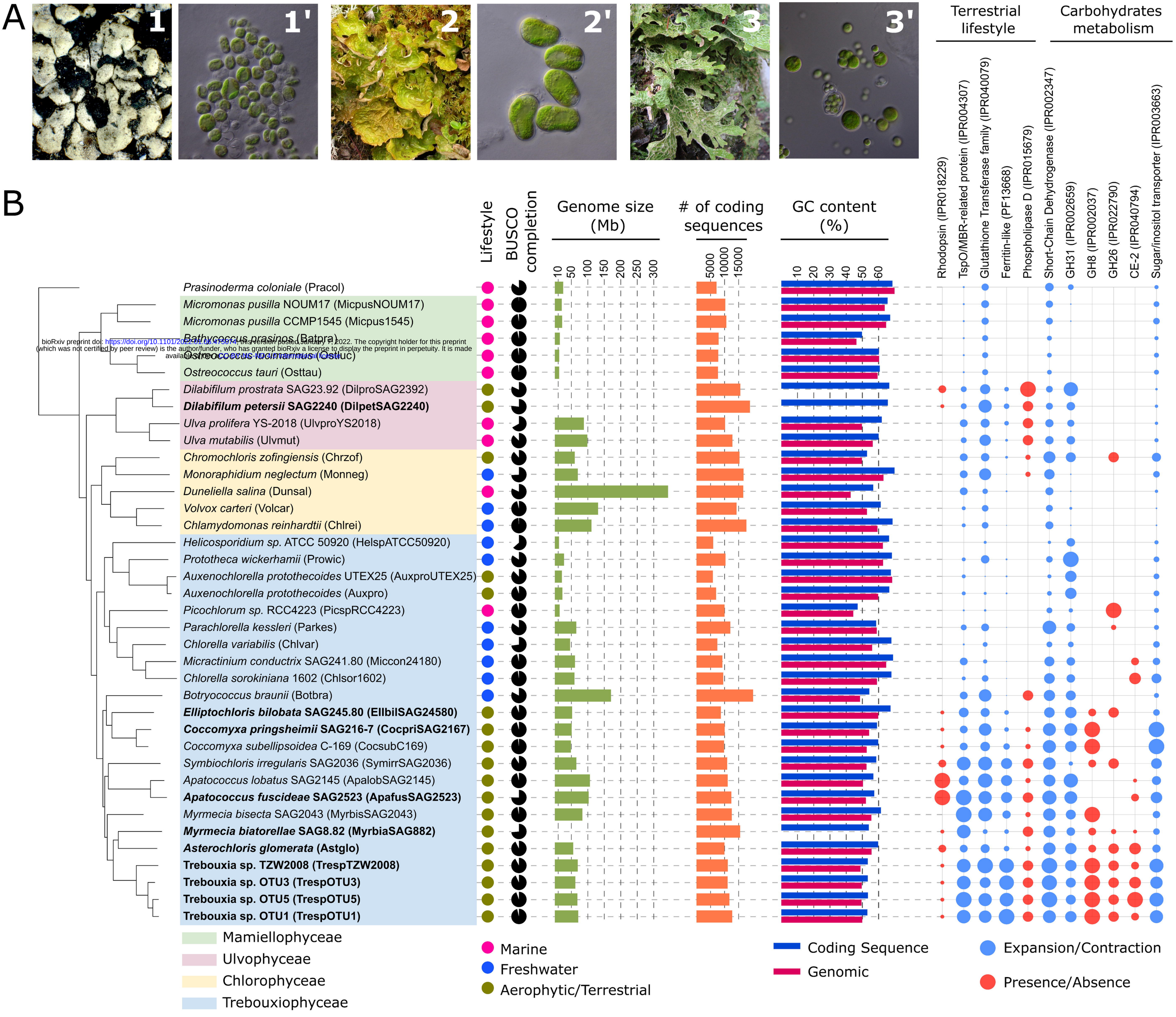
bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Prasinoderma coloniale), lifestyle of the different algae species investigated, BUSCO completion of proteomes (full circle indicates 100% completion level), size of the genome in Mb (empty rows are for species with only transcriptomic data), number of protein-coding genes, GC content for protein-coding genes (blue bars) and genomes (red bars), mapping of functional family sizes significantly more present in LFA compared to nLFA. Size is normalized for each column and are not relative between columns.

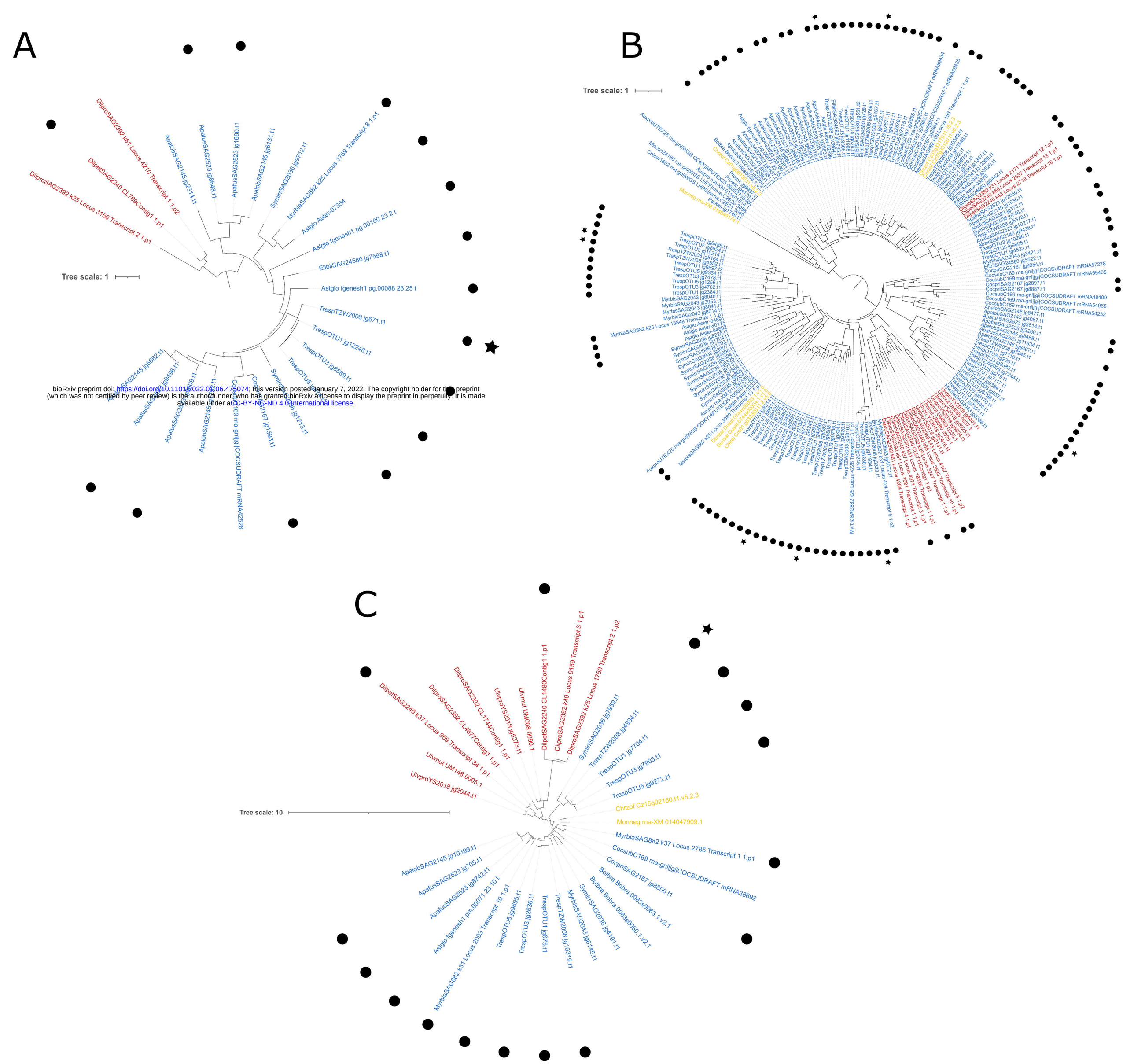
Figure 2. Phylogenies of the terrestrial lifestyle-associated genes. A. Maximum likelihood tree of the Rhodopsin family (IPR018229) rooted on the Ulvophyceae. B. Maximum likelihood tree of the Ferritin-like family (PF13668). Maximum likelihood tree of the PhospholipaseD family (IPR015679) rooted on the Ulvophyceae. Maximum likelihood tree of the GH26 family (IPR0022790). Branches are colored as follow: the Ulvophyceae are in red, the Chlorophyceae in yellow, and the Trebouxiophyceae in blue. The black dots indicate LFA, while the black stars indicate the up-regulated genes during the symbiosis from Trebouxia TZW2008.

Figure 3. A. Maximum likelihood tree of the IPR031155 (Urea active transporter) rooted on the Mamiellophyceae. B. Maximum likelihood tree of the GH26 family (IPR0022790). Branches are colored as follow: the Mamiellophyceae are in green, the Ulvophyceae in red, the Chlorophyceae in yellow, and the Trebouxiophyceae in blue. The grey ranges represent the gene-family expansions. The black dots indicate LFA, while the black stars indicate the up-regulated genes during the symbiosis from Trebouxia TZW2008. C. Unrooteed maximum likelihood tree of GH8 family. Branches are colored as follow: outgroup sequences without the GH8 domains are in grey, while bacteria, fungi and algae GH8 are colored in pink, orange and blue respectively. The red dots indicate the presence of the GH8 functional domain (PF01270). Sequence names have been hidden for readability reasons and can be found in the Supplementary Table S8. D. Position of GH8-like coding-genes of algae on chromosome. For readability reasons, chromosomes longer than 5Mb are displayed separately. The position of the genes is indicated by the coral-colored bar.

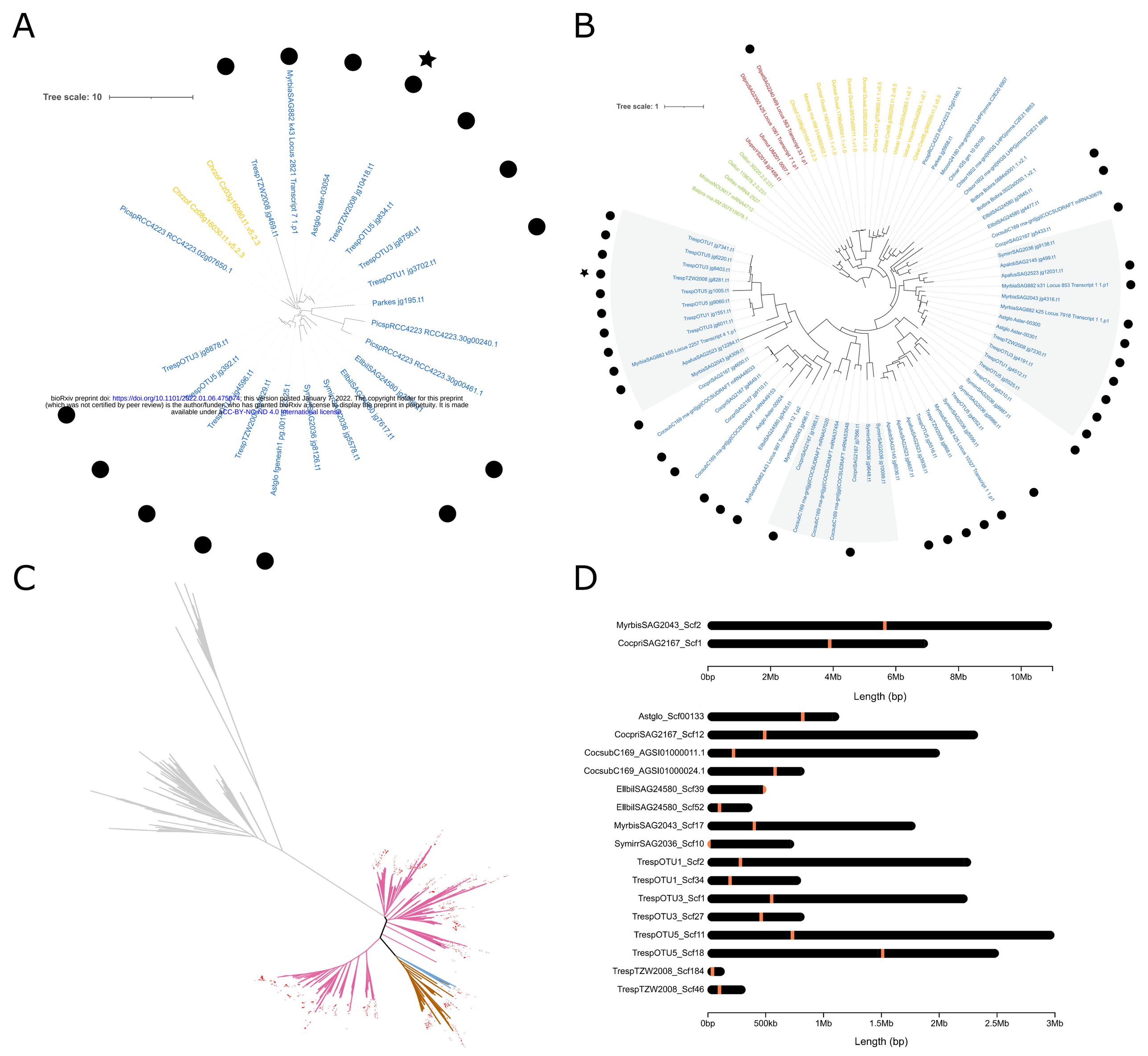
Figure 4. Model of lichens evolution in chlorophyte algae



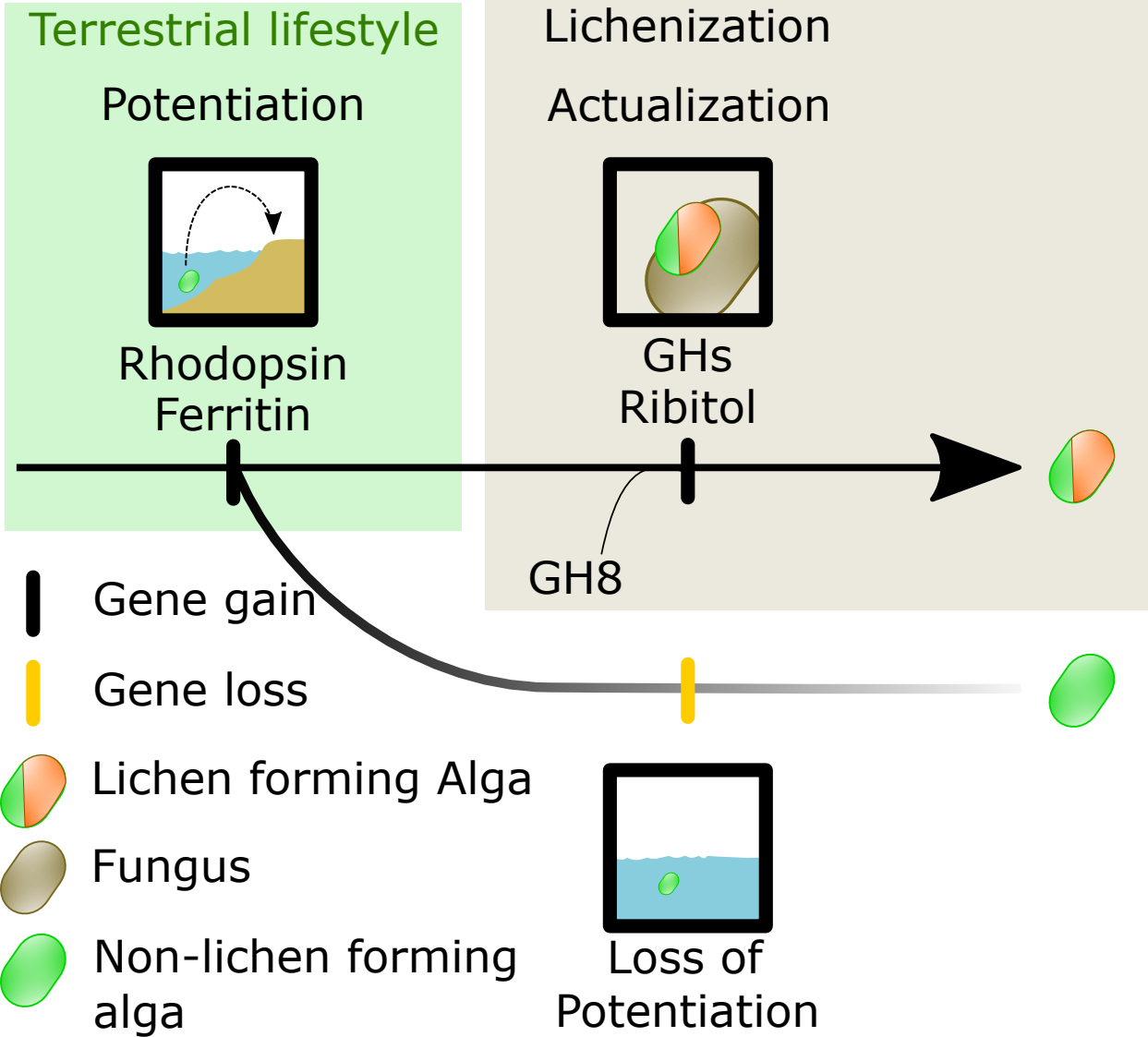
bioRxiv preprint doi: https://doi.org/10.1101/2022.01.04.475074 ; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475074; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2022.01.06.475674; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


Terrestrial lifestyle
Potentiation
Rhodopsin
Ferritin
Lichenization
Actualization
GHs
Ribitol
GH8
Gene gain
Gene loss
Lichen forming Alga
Fungus
Non-lichen forming alga
Loss of Potentiation