Investigating the Principles Governing the Interactions Between Terrestrial Green Algae and Their Associated Microbial Communities

Doctoral thesis

for

the award of the doctoral degree

of the Faculty of Mathematics and Natural Sciences

of the University of Cologne

submitted by

Magdalena Wiktoria Slawinska

accepted in the year 2026

Abstract

In terrestrial environments, soil algae take up carbon in amounts corresponding to approximately 31% of the global anthropogenic carbon emissions. In nature, algae recruit microbes from their environment to their phycosphere, a region closely surrounding algal cells, analogous to the rhizosphere of land plants. The phycosphere is enriched in algal exudates, which stimulate bacterial colonization. In exchange for carbon and metabolites provided by the algae, microbial communities supply their photosynthetic hosts with micronutrients, alleviate stresses, or provide protection against pathogens. However, despite their importance for the global carbon cycle and mitigating climate change, and the dependence of many algae on bacterial presence for optimal growth, the interactions of soil-borne microscopic algae with their associated phycosphere communities remain underexplored. Recent work showed that subaerial green algae, including the model alga *Chlamydomonas reinhardtii*, can recruit and assemble their phycosphere microbiota from soil-derived bacteria. However, how widespread this ability is among diverse green algae and whether host identity shapes community composition remained unclear. Whether soil-borne fungi, in addition to bacteria, can be recruited to the phycosphere was also unknown. To address these questions, we combined ecological characterization with molecular analysis. First, we characterized the bacterial and fungal communities associated with subaerial algae in a terrestrial environment and explored the temporal dynamics of bacterial communities in controlled liquid-culture experiments. Second, we characterized the *C. reinhardtii* transcriptional response to its phycosphere microbiota to identify processes and genes possibly important for the interaction.

Using the first approach, we show that subaerial green algae ranging from Chlorophyta to Streptophyta assemble distinct phycosphere communities from soil-derived bacteria and fungi, and that host identity is the main driver of phycosphere composition. In soil, the bacterial communities were strongly reshaped by algal hosts, whereas the fungal communities showed weaker and less consistent responses. To reduce environmental variability, we also examined bacterial community assembly in a controlled liquid-based system, where we detected a significant correlation between host phylogeny and bacterial community composition – a hallmark of phylosymbiosis.

Among the bacterial and fungal lineages present in algal phycospheres, we detected taxa known to associate with land plants, including Rhizobiales and Hypocreales, further supporting the finding that some microbial taxa can associate with a wide range of photosynthetic hosts.

With the second approach, we provide the first characterization of the model chlorophyte *Chlamydomonas reinhardtii* transcriptional response to its phycosphere microbiota and define processes potentially important for this interaction. We show that the phycosphere community triggers a coordinated shift in *C. reinhardtii* gene expression, suggesting a transition from catabolism toward growth- and biosynthesis-oriented metabolism. This change in gene expression is possibly supported by bacterial contributions (e.g. CO_2 and other metabolites) and involves upregulation of biosynthetic pathways, translation machinery, and metabolic exchange genes. Central to the proposed interaction are changes in amino acid oxidation mediated by L-amino acid oxidase (LAO1), together with SynCom-dependent upregulation of ammonium transport and assimilation, agmatine iminohydrolase-mediated putrescine biosynthesis, and amino acid exchange mediated by a bidirectional amino acid transporter (BAT1). Parts of this response are conserved with evolutionarily distant land plants, suggesting their possible ancestral origin. The identified candidate genes provide a starting point for functional validation using available *C. reinhardtii* mutants.

Together, these results demonstrate that the capacity to assemble distinct phycosphere communities is shared between subaerial green algae, possibly predating the divergence of Chlorophyta and Streptophyta. The conservation of transcriptional responses between *C. reinhardtii* and land plants suggests that principles of host–microbe interactions are conserved across evolutionarily distant photosynthetic hosts. This thesis deepens our understanding of principles governing the interactions of terrestrial algae with their phycosphere microbiota, as well as of evolutionary origins of interactions between photosynthetic organisms and bacteria, and may inform microbiome-based strategies for biotechnology applications.

Zusammenfassung

In terrestrischen Ökosystemen fixieren Bodenalgen Kohlenstoff in einer Größenordnung, die etwa 31 % der globalen anthropogenen Kohlenstoffemissionen entspricht. In natürlichen Habitaten rekrutieren Algen Mikroorganismen aus ihrer Umgebung in die sogenannte Phycosphäre, einen die Algenzellen unmittelbar umgebenden Bereich, der funktionell der Rhizosphäre von Landpflanzen entspricht. Die Phycosphäre ist durch algenstämmige Ausscheidungen angereichert, welche die mikrobielle Besiedlung, insbesondere durch Bakterien, fördern. Im Gegenzug für von den Algen bereitgestellten Kohlenstoff und Metabolite versorgen mikrobielle Gemeinschaften ihre photosynthetischen Wirte mit Mikronährstoffen, tragen zur Abmilderung von Stress bei oder bieten Schutz vor Pathogenen. Trotz ihrer Bedeutung für den globalen Kohlenstoffkreislauf und die Abschwächung des Klimawandels sowie der Abhängigkeit vieler Algen von bakteriellen Partnern für ein optimales Wachstum sind die Interaktionen mikroskopisch kleiner Bodenalgen mit ihren assoziierten Phycosphären-Gemeinschaften bislang nur unzureichend erforscht. Jüngere Studien haben gezeigt, dass subaeriale Grünalgen, darunter die Modellalge *Chlamydomonas reinhardtii*, ihre Phycosphären-Mikrobiota aus bodenbürtigen Bakterien rekrutieren und strukturieren können. Unklar blieb jedoch bisher, wie verbreitet diese Fähigkeit innerhalb der Grünalgen ist und inwieweit die Identität des Wirts die Zusammensetzung der Gemeinschaft bestimmt. Ebenso war bislang unbekannt, ob neben Bakterien auch bodenbürtige Pilze in die Phycosphäre rekrutiert werden. Zur Beantwortung dieser Fragestellungen kombinierten wir ökologische Charakterisierungen mit molekularbiologischen Analysen. Zunächst untersuchten wir die mit subaerialen Algen in einer terrestrischen Umgebung assoziierten Bakterien- und Pilzgemeinschaften und analysierten die zeitliche Dynamik bakterieller Gemeinschaften in kontrollierten Flüssigkulturexperimenten. Anschließend charakterisierten wir die transkriptionelle Antwort von *C. reinhardtii* auf seine Phycosphären-Mikrobiota, um Prozesse und Gene zu identifizieren, die für diese Interaktion potenziell relevant sind.

Mithilfe des ersten Ansatzes zeigen wir, dass subaeriale Grünalgen aus verschiedenen Linien, von Chlorophyta bis Streptophyta, jeweils spezifische Phycosphären-Gemeinschaften aus bodenbürtigen Bakterien und Pilzen assemblieren und dass die

Wirtsidentität den wichtigsten Einflussfaktor für die Zusammensetzung dieser Gemeinschaften darstellt. Während die bakteriellen Gemeinschaften im Boden durch die Algenwirte stark umgeformt wurden, reagierten die Pilzgemeinschaften schwächer und weniger konsistent. Um umweltbedingte Variabilität zu minimieren, untersuchten wir zudem die Assemblierung bakterieller Gemeinschaften in einem kontrollierten, flüssigkeitsbasierten System. Dabei konnten wir eine signifikante Korrelation zwischen der Phylogenie des Wirts und der Zusammensetzung der bakteriellen Gemeinschaft nachweisen, die ein charakteristisches Merkmal der Phylosymbiose ist.

Unter den in den Algen-Phycosphären nachgewiesenen bakteriellen und pilzlichen Taxa identifizierten wir auch Gruppen, die auch als Assoziationspartner von Landpflanzen bekannt sind, darunter Rhizobiales und Hypocreales. Dies stützt die Annahme, dass bestimmte mikrobielle Taxa mit einer Vielzahl photosynthetischer Wirte interagieren können.

Mit dem zweiten Ansatz liefern wir die erste Charakterisierung der transkriptionellen Antwort des Modellchlorophyten *C. reinhardtii* auf seine Phycosphäre-Mikrobiota und identifizieren Prozesse, die für diese Interaktion von Bedeutung sein könnten. Wir zeigen, dass die Phycosphären-Gemeinschaft eine koordinierte Umprogrammierung der Genexpression in *C. reinhardtii* auslöst, die auf einen Übergang von einem katabolen zu einem wachstums- und biosyntheseorientierten Stoffwechsel hinweist. Diese veränderte Genexpression wird möglicherweise durch bakterielle Beiträge, wie etwa CO_2 und andere Metabolite, unterstützt und umfasst eine Hochregulation biosynthetischer Stoffwechselwege, der Translationsmaschinerie sowie von Genen, die am metabolischen Austausch beteiligt sind. Zentral für die postulierte Interaktion sind die Veränderungen in der Aminosäureoxidation, vermittelt durch die L-Aminosäureoxidase (LAO1), in Kombination mit der SynCom-abhängigen Hochregulation des Ammoniumtransports und der Ammoniumassimilation, der durch Agmatiniminohydrolase vermittelten Putrescinbiosynthese sowie des Aminosäureaustausch über einen bidirektionalen Aminosäuretransporter (BAT1). Teile dieser Antwort sind zwischen *C. reinhardtii* und evolutionär weit entfernten Landpflanzen konserviert, was auf einen möglichen gemeinsamen evolutionären Ursprung hindeutet. Die identifizierten Kandidatengene stellen einen Ausgangspunkt für eine funktionelle Validierung mithilfe verfügbarer *C. reinhardtii*-Mutanten dar.

Zusammenfassend zeigen diese Ergebnisse, dass die Fähigkeit zur Assemblierung spezifischer Phycosphären-Gemeinschaften bei subaerialen Grünalgen weit verbreitet ist und möglicherweise bereits vor der Aufspaltung von Chlorophyta und Streptophyta existierte. Die Konservierung bestimmter transkriptioneller Antworten zwischen *C. reinhardtii* und Landpflanzen legt nahe, dass grundlegende Prinzipien der Wirt-Mikroben-Interaktion über evolutionär weit entfernte photosynthetische Organismen hinweg erhalten geblieben sind. Diese Arbeit erweitert unser Verständnis der Mechanismen, die die Interaktionen terrestrischer Algen mit ihrer Phycosphären-Mikrobiota steuern, sowie der evolutionären Ursprünge der Interaktionen zwischen photosynthetischen Organismen und Bakterien und könnte mikrobiombasierter Strategien für biotechnologische Anwendungen beitragen.

Acknowledgments

First and foremost, I would like to thank my advisor, Ruben Garrido-Oter, for giving me the opportunity to pursue this PhD project and for providing an environment in which I could grow as an independent researcher. From the very beginning, he showed me tremendous patience and encouragement. The enthusiasm with which he approaches science has been a constant source of motivation. I have learnt so much from him: how to ask the right questions, analyze data rigorously, and develop scientific arguments. I deeply admire how quickly he can spot patterns in data, get to the core of a problem, and draw connections that others might miss. I am profoundly grateful for the trust he placed in me.

I also want to thank José (Pepe) Flores-Uribe, who trained me patiently and from whom I learnt a great deal about performing wet-lab experiments and bioinformatics. The joy with which he approaches research, his balance of rigor and creativity, and his unfailing kindness have been invaluable. He was always generous with his time, willing to explain, teach, and discuss. I am deeply thankful for all he has invested in my development.

I am grateful to Paloma Durán, who has always been open to discussion and taught me so much about working with algae. Collaborating with you has been a true pleasure. I also thank my Thesis Advisory Committee — Stanislav Kopriva and Stéphane Hacquard — for their feedback and guidance at key stages of this project, and Michael and Barbara Melkonian for isolating the algae strains and advice on cultivating them. To all my group members, whom I am lucky to call friends: thank you for the wonderful discussions and your unwavering support.

I would not have come this far without my family. I am deeply thankful to my parents for always supporting my endeavors and encouraging my curiosity, and to my brother, who inspired me to move abroad and choose paths that would allow me to grow – advice that led me here. I also thank all my friends for being there for me. Finally, I would like to thank my husband, Maciek, for his love, patience, and constant support. He has always encouraged my ideas and given me space to grow, and in these final months he has carried so much so that I could focus on finishing this work. I am truly lucky to have you by my side.

Table of contents

Table of Figures

List of abbreviations

ADC	arginine decarboxylase
ADP	adenosine diphosphate
AGF	ancestral green flagellate
AGO	Argonaute
AM	arbuscular mycorrhizal
AMP	adenosine monophosphate
ASV	amplicon sequence variant
ATP	adenosine triphosphate
BAT1	bidirectional amino acid transporter
BCAA	branched-chain amino acids
BP	Biological Process
CAP	Constrained Analysis of Principal Coordinates
CAS	Cologne Agricultural Soil
CDPK	calcium-dependent protein kinase
CLDs	compact letter displays
CLiP	Chlamydomonas Library Project
CO_2	carbon dioxide
COs	chitooligosaccharides
CPCoA	constrained Principal Component Analysis
DCL	Dicer-like
DEGs	differential expressed genes
DNAJC2	DnaJ homolog subfamily C member 2
dpi	days post-inoculation
EGF	epidermal growth factor
ETI	effector-triggered immunity
FDR	false discovery rate
G6PD	glucose-6-phosphate 1-dehydrogenase
GA	gibberelic acid
GAF	Gene Annotation File
GMP	guanosine monophosphate
GO	Gene Ontology
GOMAP	Gene Ontology Meta Annotator for Plants
GSEA	Gene Set Enrichment Analysis
HGT	horizontal gene transfer
IAA	indole-3-acetic acid
IMP	inosine 5'-monophosphate
IMPDH	IMP dehydrogenase
IPyA	indole-3-pyruvic acid
ISR	induced systemic resistance
JA	jasmonic acid
KCM-grade	Klebsormidiophyceae, Chlorokybophyceae, and Mesostigmatophyceae
KEGG	Kyoto Encyclopedia of Genes and Genomes

KEGG Orthology	KO
LAO1	L-amino acid oxidase
LCOs	lipochitooligosaccharides
log2FC	log2(fold change)
LRR	leucine-rich repeatrs
LysM	lysine motif
MAMPs	microbe-associated molecular patterns
MAPK	mitogen-activated protein kinase
MF	Molecular Function
MLD	malectin-like domain
Mya	million years ago
NADH	nicotinamide adenine dinucleotide, reduced
NLR	nucleotide-binding leucine-rich repeat
NOP56	nucleolar protein 56
NPR	nonexpresser of PR genes
OrgDb	organism annotation database
PCoA	Principal Component Analysis
PCoA	Principal Component Analysis
PRMT3	protein arginine methyltransferase 3
PRPP	5-phosphoribosyl-1-pyrophosphate
PRRs	pattern recognition receptors
PTI	pattern-triggered immunity
RA	relative abundance
RBH	Reciprocal Best Hit
RIO2	RIO kinase 1=2
RLCK	receptor-like cytoplasmic kinase
RLK	receptor-like kinase
RLP	receptor-like protein
RNA-seq	RNA sequencing
RNAi	RNA interference
RT	room temperaturs
SA	salicylic acid
SAR	systemic acquired resistance
siRNAs	small interfering RNAs
SynCom	synthetic community
TAG	tricacylglycerol
TCA	tricarboxylic acid
TF	transcription factor
TUC	Trebuxiophyceae, Ulvophyceae, Chlorophyceae
XMP	xanthosine 5'-monophosphate
ZCC-grade	Zygnematophyceae, Coleochaetophyceae and Charophyceae

1 Introduction

1.1 The evolution of photosynthetic organisms

The emergence of photosynthesis has altered Earth's ecosystems, allowing aerobic organisms, including humans, to evolve. It is generally accepted that the first organisms capable of oxygenic photosynthesis were ancestors of modern Cyanobacteria[1–3]. These organisms transformed Earth's atmosphere from reducing to oxidizing and later became the evolutionary source of plastids in photosynthetic eukaryotes[4].

1.1.1 The diversification of green algae

The green lineage originated from a primary endosymbiosis event, in which a heterotrophic eukaryote engulfed a cyanobacterium, which was subsequently integrated into the cell as a plastid[4–6]. The cyanobacterial genome was later reduced, and some genes were transferred to the host nucleus via endosymbiotic gene transfer[5,7]. Three Archaeplastida lineages originated from this primary endosymbiosis event and contain primary plastids: green plants, red algae, and glaucophytes[4,7,8]. Their plastids have two membranes, which are believed to correspond to the two membranes (plasma membrane and outer membrane) of the engulfed cyanobacteria[5,8,9]. Subsequently, the plastids of red and green algae plastids spread to other eukaryotes in the events of secondary (eukaryotic) endosymbiosis, in which a heterotrophic eukaryotic host ingested the primary endosymbiont and retained only its plastid[8]. As a result, plastids resulting from the secondary endosymbiosis are surrounded by three or four membranes[5,8]. In some cases, secondary endosymbionts were engulfed by other eukaryotes, giving rise to tertiary endosymbionts[5–7]. Eukaryotic lineages that harbor plastids of green algal ancestry, resulting from secondary endosymbiosis events, include euglenids and chlorarachniophytes. Euglenids, protists found in marine and freshwater habitats, have plastids surrounded by three membranes, while chlorarachniophytes have plastids surrounded by four membranes and possess a nucleomorph, a reduced remnant of the eukaryotic endosymbiont's nucleus[5,8].

The diversity of plastids is greater among organisms that harbor plastids of red algal ancestry. This group includes: haptophytes and stramenophiles, which contain chlorophyll *a* + *c* and have plastids enclosed by four membranes[8,10], cryptophytes that have chlorophyll *a* + *c* pigmentation and the nucleomorph[11], and dinoflagellates, whose plastids are most diverse among all eukaryotic phototrophs[8]. The most common plastids among dinoflagellates are plastids acquired via secondary endosymbiosis events that contain both chlorophyll *a* + *c*, and the carotenoid pigment peridinin[8,9]. Other dinoflagellates, such as *Karlodinium* and *Karenia*, harbor haptophyte-derived plastids that originated through tertiary endosymbiosis[6,12].

Although a wider diversity of photosynthetic organisms acquired plastids through secondary or tertiary endosymbiosis, green plants (Viridiplantae), whose plastids originated from primary endosymbiosis, have played an important role in Earth's ecosystem[7]. Viridiplantae include green algae, which are prominent in aquatic and terrestrial ecosystems, and land plants, which dominate terrestrial habitats[4,7]. Despite their immense morphological diversity, ranging from unicellular to complex multicellular forms, green algae and land plants share several common characteristics[7]. Their chloroplasts are enclosed by double membranes and contain chlorophyll *a* + *b* along with other pigments such as carotenoids and xanthophylls[4]. Their thylakoids are lamellae-grouped, and in some chloroplasts, pyrenoids surrounded by starch are present[4]. Most green plants have cell walls composed of cellulose, and most green algae possess two or four flagella[4,13]. These flagella are isokont, meaning they are similar in structure, but may differ in length, and the flagellar transition zone usually has a stellate structure (nine-pointed star linking nine pairs of microtubules)[14].

Green plants comprise two lineages: Chlorophyta, which diversified from oceanic plankton and later split into modern planktonic prasinophytes and the core chlorophytes, and Streptophyta, which evolved in freshwater and terrestrial habitats and include algae and land plants[7,15]. Most prasinophytes inhabit marine environments, although some species can be found in freshwater environments[16]. Prasinophytes are mostly unicellular algae with differently shaped cells with or without flagella, contain different pigments and show diverse mitotic and cytokinetic processes[13,15,17]. Based on their morphological features and their phylogenetic position,

prasinophytes are believed to resemble the ancestral green flagellate (AGF), which gave rise to the green plant lineage 700-1500 million years ago (Mya)[7]. AGF is inferred to have been a marine, planktonic, unicellular flagellate with some traits typical of prasinophytes[7].

The core Chlorophyta lineage evolved from one of the ancestral prasinophytic lineages and includes marine and freshwater Pedinophyceae and Chlorodendrophyceae, as well as the larger, diverse clades Trebouxiophyceae, Ulvophyceae, and Chlorophyceae, collectively referred to as the TUC clade, which inhabit marine, freshwater, and terrestrial habitats[13,15,18]. Most Chlorophyta are characterized by the presence of a phycoplast comprised of microtubules lying in the plane of cell division; however, the phycoplast was lost in the Ulvophyceae[7,13,19]. Ulvophyceae exhibit the greatest diversity in cellular morphology among Chlorophyta, ranging from unicellular organisms to macroscopic multicellular forms, including macroscopic green seaweeds found in marine environments. However, some members of Ulvophyceae inhabit freshwater and damp subaerial habitats[18]. Morphological forms of Ulvophyceae include non-motile uninucleate cells, uninucleate cells arranged into filaments or blades, siphonocladous organisms with multicellular bodies composed of multinucleate cells, and siphonous organisms, whose entire body is a single giant tubular cell containing millions of nuclei[13,20–24]. Similarly, Trebouxiophyceae display a wide range of morphologies, including both motile and non-motile forms: unicellular flagellates, coccoids, colonies, multicellular filaments, and blades[4,13]. While some Trebouxiophyceae are free-living, others form symbiotic lichen formations with fungi or live as parasites[4,25,26]. The final group within the TUC clade, Chlorophyceae, is also highly diverse and comprises motile and non-motile unicellular organisms, colonies, filaments, and thalli[13]. Members of Chlorophyceae reproduce both asexually and sexually and are commonly found in freshwater and terrestrial habitats[13]. One of the extensively studied members of Chlorophyceae is the model alga *Chlamydomonas reinhardtii*, a biflagellate, facultatively heterotrophic alga[13]. *C. reinhardtii* can reproduce both sexually and asexually[13]. During cell division, cells first enlarge without division, then divide rapidly in several rounds with little growth between rounds, generating multiple unicellular cells[27]. The sexual phase of the *C. reinhardtii* life cycle is triggered by nitrogen starvation, which induces differentiation of haploid vegetative cells into isogamous gametes[27,28]. Gametes of

opposite mating type fuse together to form a diploid zygote, which then undergoes meiosis to generate four haploid vegetative cells[13,28]. *C. reinhardtii* is a well-established model organism with a fully sequenced genome[29], for which a variety of molecular techniques[30,31], as well as barcoded mutant libraries representing the majority of *C. reinhardtii* genes, are available, making it an attractive model to investigate gene functions in photosynthetic algae[32–34].

1.1.2 Streptophyta and the colonization of land

The second lineage of green plants, Streptophyta, split from the Chlorophyta approximately one billion years ago[15]. This clade includes both streptophyte algae with diverse morphologies, encompassing unicellular flagellates and filaments, and complex land plants, which colonized land approximately 476-432 Mya[7,13]. The paraphyletic Streptophyta can be differentiated into two grades: the basal-branching KCM-grade, including Klebsormidiophyceae, Chlorokybophyceae, and Mesostigmatophyceae, and the higher-branching ZCC-grade (Zygnematophyceae, Coleochaetophyceae, and Charophyceae), which represents the sister group to land plants[15,19]. Within the KCM-grade, freshwater Mesostigmatophyceae and terrestrial Chlorokybophyceae are small algal classes that diverged before the filamentous Klebsormidiophyceae[19,35]. Klebsormidiophyceae diverged more than 830 Mya and successfully colonized diverse habitats owing to physiological traits, such as the ability to tolerate prolonged periods of dessication[36]. During cell division in KCM-grade algae, cytokinesis occurs via centripetal cleavage, whereas ZCC-grade algae evolved cytokinesis via a centrifugal phragmoplast, formed by microtubules perpendicular to the cell division plane[37].

Charophyceae, which belong to the ZCC-grade, are well represented in the fossil record and comprise mainly freshwater streptophyte algae, whose bodies are complex macroscopic thalli[19,37]. Their cell walls contain plasmodesmata which facilitate cytoplasmic communication between neighboring cells[19,37]. Another class of multicellular streptophytes, Coleochaetophyceae, inhabits terrestrial and freshwater habitats and exhibits a wide range of body plans: from branched filaments to discoid parenchymatous thalli[19,38]. The most species-rich class within the ZCC-grade is Zygnematophyceae, which can be found in freshwater and terrestrial habitats,

including extreme environments such as acidic waters and soils, snow and ice, and desert crusts[19,39–41]. Zygnematophyceae encompass unicellular, colonial, and filamentous algae, and are divided into five orders: Spirogloeales, Desmidiales, Zygnematales, Spirogyrales, and Serritaeniales[42,43]. Interestingly, Zygnematophyceae, rather than the structurally more complex Charophyceae and Coleochaetophyceae, have been identified as the sister group to land plants[43,44]. Recent phylogenomic analyses placed one of the Zygnematophyceae algae, *Spirogloea muscicola*, on a branch distinct from other Zygnematophyceae and closest to the branch point separating the land plants from Zygnematophyceae[43]. This finding suggests that *S. muscicola* represents a lineage that is the closest extant relative of the most recent common ancestor of streptophyte algae and land plants[43]. The common ancestor of land plants and Zygnematophyceae was already subaerial or terrestrial and had acquired some genes important for plant growth regulation and responses to biotic and abiotic stresses from soil bacteria through horizontal gene transfer (HGT)[43].

1.2 Plant-microbiota interactions

While studying the evolution of photosynthetic organisms, we must consider that they have established long-term ecological interactions with bacteria since the early stages of evolution, as bacteria appeared on Earth before the emergence of algae[45,46]. According to the hologenome theory of evolution, the plant, together with its associated microorganisms, constitutes a holobiont[47,48]. Within the holobiont, the symbiotic microorganisms can be transmitted between generations, the plant-microbe associations influence the fitness of the holobiont, and the holobiont's evolution can result in changes to both the host and the microbes' genomes[47]. While pathogens challenged plants and led to the evolution of defense mechanisms, mutually beneficial symbiotic interactions with microbes helped land plants to colonize land[49]. Notably, studies have revealed that the ability of the earliest land plants to engage in symbiotic interactions with fungi was among the innovations that facilitated plant colonization of land[50].

1.2.1 Evolution of plant-microbiota interactions

The preadaptation for symbiosis was most likely present in the algal ancestor of land plants[50]. Potential homologs of two genes controlling steps of arbuscular mycorrhizal

(AM) symbiosis, a mutualistic interaction between land plants and soil fungi, were found in both chlorophytes and charophytes[50]. One gene encodes a nuclear envelope-localized potassium channel DMI1, and the other encodes a calcium- and calmodulin-dependent protein kinase CCaMK[50]. This indicates genes might have appeared before the divergence of chlorophytes and the land colonization by plants[50].

The establishment of AM symbiosis begins with an exchange of chemical signals between the plant and the fungal partner. The plant emits strigolactone signals, which are perceived by the fungus, and the fungus, in turn, initiates the production of Myc factors[51–53]. These Myc factor signals are perceived by the plant at the plasma membrane, where lysin motif receptor-like kinase (LysM-RLK) proteins and a malectin-like domain (MLD)-RLK (DIM2) activate DMI1[54,55]. DMI1 induces calcium oscillations in and around the nucleus that are decoded by CCaMK[54,56–60]. CCAMK phosphorylates the substrate protein CYCLOPS, activating downstream transcription factors (TFs), and thereby initiating the symbiosis program in the plant[54,56–60]. Among these TFs, GRAS genes were found in *S. muscicola*, the extant relative of the most recent common ancestor of streptophyte algae and land plants[43]. These genes may have been acquired by the common ancestor of Zygnematophyceae and land plants via HGT from soil bacteria, as they are found only in bacteria, Zygnematophyceae, and land plants[43]. In contrast to this early signaling module, the components of later AM symbiosis steps, e.g., VAPYRIN, were not found in Zygnematophyceae, indicating that they evolved first in embryophytes[43,50].

In contrast to beneficial symbiotic interactions, encounters with harmful or pathogenic microbes drove the evolution of a complex plant immune system, whose key components evolved at different timepoints[46,49]. Microbe-associated molecular patterns (MAMPs) include molecules such as chitin, flagellin, chitooligosaccharides (COs), and lipochitooligosaccharides (LCOs)[46,49,61]. While COs and LCOs include specialized molecules that promote the establishment of AM symbiosis, the majority of MAMPs are perceived by pattern recognition receptors (PRRs), triggering the initiation of pattern-triggered immunity (PTI)[62]. PRRs include receptor-like kinases (RLKs) that possess an extracellular ligand-binding domain, a transmembrane domain, and an intracellular kinase domain, as well as receptor-like proteins (RLPs), which lack an intracellular kinase domain[63,64]. The extracellular domains of PRRs are

highly diverse and may include leucine-rich repeats (LRR), lysine motifs (LysM), epidermal growth factors (EGF), or lectin domains[49,65]. Ligand binding to the extracellular PRR domains triggers association and activation of receptor-like cytoplasmic kinases (RLCKs), which in turn activate mitogen-activated protein kinase (MAPK) and calcium-dependent protein kinases (CDPKs) cascades, ultimately leading to PTI through transcriptional reprogramming[63–65].

To prevent pathogen infection, plants possess intracellular disease-resistance proteins such as members of the nucleotide-binding leucine-rich repeat (NLR) family, which detect pathogen-derived effectors inside the cells[62]. Recognition of the effectors induces effector-triggered immunity (ETI), leading to programmed cell death that limits pathogen spread from the infection site, and to production of mobile immune signals that are transported to uninfected tissues[46,62,66]. These mobile immune signals include methyl salicylic acid, azelaic acid and glycerol-3-phosphate, which induce systemic acquired resistance (SAR) mechanism, in which the perception of the these signals triggers accumulation of salicylic acid (SA), which binds to nonexpresser of PR genes (NPR) proteins, resulting in transcriptional reprogramming and production of antimicrobial pathogenesis-related proteins that protect plants from secondary infections[46,66–68]. In addition to SA-mediated SAR against biotrophic pathogens, other phytohormones are linked to plant immunity: jasmonic acid (JA) mediates induced systemic resistance (ISR), which activates response against necrotrophic pathogens, and ethylene, abscisic acid, and gibberellic acid (GA)[19,69]. Some components of the JA signaling pathway, such as the oxygenation of α-linolenic acid to the oxylipin 12-oxo-phytodienoic acid, which is subsequently converted to JA, emerged before the split of Klebsormidiophyceae and ZCC-grade algae, while others predated or accompanied terrestrialization[70,71].

Different parts of the complex plant immune system evolved at distinct timepoints. As noted previously, the functional signaling module involved in early AM symbiosis evolved in the common ancestor of streptophytes and land plants, whereas components of later AM symbiosis stages first appeared in land plants[43,50]. In general, the immune systems of algae, including model alga *C. reinhardtii,* differ significantly from those of land plants[46]. The earliest component of the plant immune system to evolve is RNA interference (RNAi), which defends plants against viral infections[63,72].

While not all Chlorophyta possess the full RNAi system, *C. reinhardtii* has both Dicer-like (DCL) proteins, which generate viral small interfering RNAs (siRNAs) from double-stranded RNA intermediates of RNA viruses, and Argonaute (AGO) proteins, which mediate RNA-based antiviral immunity using siRNAs[46,63,72]. Although RLK homologs are present in the *C. reinhardtii* genome, they do not encode canonical RLK domain architectures[46]. The immune system of Streptophyta algae is more complex, with the Klebsormidiophyceae alga *Klebsormidium nitens* already possessing functional RLKs and NLRs[46]. While RLKs and NLRs are present in *K. nitens*, only NPRs homologs lacking canonical domain architectures were found in its genome[46]. NPR proteins are present only in land plants, indicating that they evolved after the origin of land plants[46]. Taken together, these observations indicate that both symbiosis-related and immunity-related programs are present in land plants and algae; however, in algae, these programs are often represented by simpler gene sets or by evolutionary precursors of the more complex pathways found in plants.

1.2.2 The root microbiota

The immune system plays an important role in how plants respond to microbes, thereby influencing the assembly of complex microbial communities associated with the plant. While all plant tissues are colonized by microbes that improve the host fitness by growth promotion[73,74], nutrient mobilization[75–77], enhanced pathogen resistance[78,79] , and stress tolerance[73,80,81], this section focuses on the root microbiota. The root-associated microbiota is recruited from soil. The first microhabitat influenced by the plant is the rhizosphere, a thin soil layer surrounding the roots, in which the local environment is modified by plant exudates[48,82]. These exudates stimulate bacterial chemotaxis and can serve as energy sources for microbes[83]. While bulk soil is rich in diverse microorganisms, the rhizosphere microbial community has lower richness and differs from that of the surrounding soil[84–86]. Beyond the rhizosphere, root-associated microbial habitats can be further subdivided. The rhizoplane is the root tissue surface itself that is colonized by microbes firmly attached to it[48]. Microbes can also colonize the interior of the root, referred to as the endosphere, where microbial diversity is greatly reduced compared to the external root compartments[85,86].

The root microbiota is dominated by several bacterial phyla, including Acidobacteria, Actinobacteria, Bacteroidetes, Firmicutes, Planctomycetes, and Proteobacteria[83], as well as by diverse fungal phyla: Ascomycota, Basidiomycota, Chytridiomycota, Mortiellomycota, Mucoromycota, Zoopagomycota[87], and Glomeromycota, the latter of which comprises AM fungi[88]. The overall structure of the root-associated microbiota is largely shaped by soil microbiota composition, but is also influenced by the plant host genotype, species identity, and evolutionary history[85,86,89–91]. While bacterial root-associated communities are strongly shaped by the host[89], fungal communities are influenced more by the soil than by host species identity[90].

Several factors are important for root microbiota assembly, including edaphic factors, which are soil properties influencing microbial community structures, and host-dependent factors[83]. The plant immune system, which protects plants against pathogens, can be modulated by root commensals, thereby enabling their colonization and shaping the root microbiota[92,93]. Differentiation of the root microbiota from the surrounding soil community can be described by a two-step selection process. In the first step, the growth of selected bacteria is promoted by rhizodeposits, which include a wide range of secreted compounds, such as sugars, amino acids, purines, and organic acids, and host cell wall features. In the second step, host-genotype selection near the root fine-tunes the root microbiota[83].

1.2.3 Root exudates shape plant-microbiota interactions

Root exudates are crucial players in shaping the root microbiota[83,94] and include a diverse array of compounds such as polyamines (e.g. putrescine[77,95]), isoflavonoids[96], purines[97], coumarins[75], amino acids[95,98,99], organic acids[82], sugars[77], and vitamins[77,99]. These metabolites can influence microbiota assembly by providing nutrients, mediating signaling, regulating the local environment, inhibiting pathogens, and attracting beneficial microbes[94,100–102]. Plant-derived nutrients, such as sugars, amino acids, and organic acids, provide the microbes with energy and growth substrates[82,100]. Other root exudates function as signaling molecules, stimulating chemotaxis and promoting microbial colonization[94]. These exudates can attract specific microbial groups that provide beneficial services to the plant[99]. For example, flavonoids and isoflavonoids secreted by legume plants facilitate the establishment of symbiosis with

nitrogen-fixing bacteria[96]. Another root exudate, putrescine, was suggested to act as a signaling molecule in the rhizosphere, as a *Pseudomonas fluorescens* mutant unable to metabolize putrescine was impaired in its ability to colonize *Arabidopsis thaliana* roots[103]. Putrescine was also shown to be important for the plant growth-promoting interaction between the endophytic root fungus *Piriformospora indica* and tomato as well as *Arabidopsis thaliana*[95]. Additionally, plants can modify their rhizosphere environment under stress conditions to attract beneficial microbes. For instance, coumarins secreted by *A. thaliana* under iron limitation attract bacteria that facilitate iron uptake by the plant[75]. The recruited root-associated microbiota fulfills multiple functions, including plant growth promotion[73,74], nutrient mobilization[75–77], enhanced pathogen resistance[78,79], and stress tolerance[73,80,81].

1.3 Parallels between the root microbiota and the phycosphere microbiota

In algae, there is an analogous compartment to the plant rhizosphere, referred to as the phycosphere[104]. The phycosphere is a thin layer surrounding an algal cell, and is enriched in algal exudates, which stimulate bacterial colonization[104]. Although the term was originally applied primarily to aquatic environments, the phycosphere community of *C. reinhardtii* cultivated in soil was recently described[105]. Several important parallels exist between the rhizosphere and the phycosphere. At both interfaces, exudates modify the physicochemical environment, stimulating bacterial chemotaxis and enabling colonization of the photosynthetic host[77,94,106,107]. Some of the exchanged compounds are identical at both interfaces[104,108]. In addition, certain microbial taxa are found in both the rhizosphere and the phycosphere. For instance, some plant growth-promoting bacteria colonize the phycospheres of green algae[45], and six bacterial taxa, Caulobacterales, Hyphomicrobiales (Rhizobiales), Sphingomonadales, Burkholderiales, Lysobacterales (Xanthomonadales), and Chitinophagales (Bacteroidetes) colonize both the *C. reinhardtii* phycosphere and land plant roots[105].

Similarly, in aquatic environments, various metabolites, such as organic carbon[109,110], vitamins[106,111], and other micronutrients[112], are exchanged, influencing algal growth and resembling processes in the plant rhizosphere[104]. Beneficial bacteria can mobilize

nutrients[75,76], mitigate the effects of abiotic stresses[73,80,81], and protect plants from pathogens[78,79]. Their growth is supported by the secretion of compounds that act as attractants while simultaneously restricting the growth of non-beneficial bacteria[83,106,111,113,114]. Historically, algal-bacterial research focused on pairwise interactions between individual bacterial species and algae[111,115–118]. However, recent studies have shown that the model alga *C. reinhardtii* is capable of assembling a community of heterotrophic, soil-borne bacteria in a manner similar to land plants[105], despite lacking a complex immune system[46,105], which has been shown to play a role in the root-associated microbiota assembly in land plants[84]. It is assumed that the primary driver of the assembly of soil-derived microbial communities associated with unicellular algae is the secretion of photoassimilates[82,83,119]. Nevertheless, recent findings suggest that the provision of organic carbon compounds cannot be the only cue for microbiota assembly, and that the exchange of metabolites and/or molecular signals is most likely necessary to establish and support the growth of the phycosphere microbiota[105]. However, the principles governing the establishment of phycosphere microbiota in subaerial algae have not yet been elucidated.

1.3.1 Algae-bacteria interactions in aquatic environments

The signaling molecules and nutrients exchanged in the phycosphere of marine algae largely overlap with those in the rhizosphere and serve similar functions, including improving nutrient availability, acting as chemoattractants, and stimulating bacterial chemotaxis[120]. For instance, the interaction between *Pseudo-nitzschia multiseries*, a diatom found in the Pacific Ocean and the Atlantic Ocean, and the bacterium *Sulfitobacter* sp., involves a complex nutrient exchange of diatom-produced tryptophan and organosulfur metabolites, and ammonium and indole-3-acetic acid (IAA) produced by the bacterium[117]. The diatom *P. multiseries* attracts *Sulfitobacter* sp. with dimethylsulfoniopropionate and tryptophan. Subsequently, *Sulfitobacter* sp. metabolizes the diatom-secreted and endogenous tryptophan to produce IAA, thereby promoting *P. multiseries* growth[117].

Similar to what has been shown in plants[75], bacteria can also enhance iron nutrition in algae[112]. For example, *Marinobacter* produces vibrioferrin, a dicitrate siderophore

that forms a complex with iron. Upon photolysis, the iron-vibrioferrin chelates release iron that can assimilated by a dinoflagellate[112]. Sulfur compounds can similarly mediate the alga-bacteria interactions, as shown in the interaction between the marine *Roseobacter* and the diatom *Thalassiosira pseudonana*, in which the diatom produces dihydroxypropane-1-sulfonate, which is catabolized by *Roseobacter*[108]. Another example of a metabolic exchange is the provision of vitamin B_{12} (cobalamin) by bacteria to cobalamin auxotrophic algae[121].

In addition to metabolic exchange, marine bacteria can protect their associated algae from pathogenic bacteria. For instance, the oceanic unicellular alga *Emiliania huxleyi* can be protected from the pathogen *Phaeobacter inhibiens* by *Sulfitobacter* bacteria[118].

While algae-bacteria interactions in aquatic environments have been studied extensively, the phycosphere of terrestrial algae has only recently become a research focus.

1.3.2 Algae-bacteria interactions in terrestrial environments

Although algae-bacteria interactions in aquatic environments have been studied and their importance is generally known[104,120], in terrestrial environments, they have become a subject of study only recently, when the phycosphere microbiota of *C. reinhardtii*, originally isolated from soil[122], was analyzed in a series of experiments[105]. *C. reinhardtii* is capable of recruiting and sustaining the growth of soil-borne bacteria, and some of the bacterial lineages present in its microbiota are commonly found in root communities of various land plants[89,105]. Interestingly, this ability to associate with soil-borne bacteria was also observed in several other chlorophyte and streptophyte algae isolated from a common terrestrial habitat[105].

1.3.3 *C. reinhardtii* as a model for phycosphere research

C. reinhardtii is a convenient model organism for phycosphere research. It has a sequenced and annotated genome[29], grows well in the laboratory environment, and can be genetically transformed[30,31]. Moreover, publicly available *C. reinhardtii* mutant collections provide resources for functional characterization of *C. reinhardtii* genes[32–34]. The first Chlamydomonas Library Project (CLiP) mutant library (CLiP1) was

constructed in *C. reinhardtii* CC-4533 (CMJ030, identical to CC-5325) background strain[32,33,123] and used to characterize genes necessary for photosynthesis[32], CO_2-concentrating mechanism, and DNA repair[33]. Recently, a second *C. reinhardtii* mutant library (CLiP2) was released[34]. CLiP2 was constructed in a different background strain, *C. reinhardtii* CC-5415, which, unlike CC-4533, has a cell wall, making it attractive for some studies[34]. Mutants from the two libraries can be genetically crossed if necessary, as the background strains have opposite mating types[32–34]. All these resources and features make *C. reinhardtii* an attractive model organism for experimental studies.

Apart from the study of *C. reinhardtii* phycosphere microbiota assembled from soil-derived bacteria[105], the research on *C. reinhardtii*-bacteria interactions has been focused on the binary interactions between *C. reinhardtii* and individual bacterial species. In a recent study, a mutualistic interaction between *C. reinhardtii* and the plant growth-promoting bacterium *Methylobacterium aquaticum* under N-limiting conditions was described[116]. Under N-limiting conditions, *C. reinhardtii* expresses an L-amino acid oxidase (LAO1), which deaminates a broad range of amino acids, producing ammonium that is taken up by *C. reinhardtii* cells, and α-keto acids that remain extracellular[116,124]. During the interaction with *M. aquaticum*, LAO1 mediates the first step of IAA biosynthesis – deamination of L-tryptophane into indole-3-pyruvic acid (IPyA). IAA was accumulated extracellularly in an LAO1-dependent manner[116]. While low IAA concentrations can promote *C. reinhardtii* growth[125], high IAA concentrations inhibit *C. reinhardtii* growth[116]. However, in co-cultures with *M. aquaticum*, *M. aquaticum* degrades IAA, facilitating *C. reinhardtii* growth[116].

C. reinhardtii was also used to study cobalamin-responsiveness in algae using a reporter gene approach[126]. Recently, the tripartite interaction involving *C. reinhardtii*, the mutualistic bacterium *Mycetocola lacteus* and the antagonistic bacterium *Pseudomonas protegens* was investigated[115]. *P. protegens* secretes an algicidal cyclic lipopeptide, deflagellating, blinding, and lysing *C. reinhardtii* cells[115,127,128]. *M. lacteus* cleaves ester bonds of this cyclic lipopeptide, enhancing *C. reinhardtii* growth and protecting *C. reinhardtii* against *P. protegens*. In return, *C. reinhardtii* supplies methionine and vitamins B_1, B_3, and B_5 to the bacterium[115].

To date, research on the *C. reinhardtii* phycosphere has focused primarily on binary and tripartite interactions with individual bacterial strains. The ability of *C. reinhardtii* to assemble phycosphere microbiota from soil-borne bacteria was recently described, and it was suggested that both physical proximity of algal and bacterial cells and bidirectional exchange of chemical compounds may be needed for the recruitment and sustained growth of the *C. reinhardtii* phycosphere microbiota[105]. Yet, the principles governing the interaction between *C. reinhardtii* and its phycosphere microbiota remain unknown.

1.4 Patterns of phylosymbiosis in host-associated microbiomes

In host-associated microbiomes, a recurring pattern has been described in multiple studies: more closely related species tend to be associated with more similar microbiomes than more distantly related species[129–131]. This pattern, referred to as "phylosymbiosis", was introduced in 2013 and can be defined as "microbial community relationships that recapitulate the phylogeny of their host"[132,133]. Phylosymbiosis may arise from both stochastic and deterministic evolutionary and ecological forces[132]. Stochastic forces include microbe dispersal between environments, which influences the microbes a host encounters, niche variation across host lineages, and ecological drift, which influences the gain or loss of microbiota members[89,132,134,135]. Deterministic forces encompass selection, including microbial preference for a given host or host traits regulating microbial colonization, co-speciation of microbes with their hosts, and interactions such as priority effects – the host might be initially associated with a specific microbe, exerting selective forces on other potential colonizers[132,134].

Phylosymbiosis can result from host-microbe coevolution, codiversification, or cospeciation, but it can also emerge from rather short-term changes in microbiome composition[136]. Microbes can have an impact on their host by inducing host genomic changes that further affect phylosymbiosis[132,136]. Studying phylosymbiosis requires the collection of two data types: host and microbiome data[132]. Then, genetic distances between hosts are computed, and a phylogenetic tree is constructed for the hosts, while for the microbiome, microbial beta diversity is calculated and compared between hosts, and dendrograms or beta-diversity distance matrices are generated[132].

Comparisons between host phylogenies and the composition of microbial communities can be quantified using topological or matrix-based approaches. Topological comparisons, such as the Robinson-Foulds metric, analyze the distance between the host and microbial trees to determine the smallest number of operations necessary to convert one tree topology into the other[137]. Matrix comparisons, such as the Mantel test, use permutation to evaluate the linear correlation between corresponding elements of two independent matrices[138].

Patterns of phylosymbiosis have been observed across a wide range of host species and habitats[132]. In insects, examples include the gut microbiome of closely related *Nasonia* wasps[133], fig wasps[139] , and social bees[140]. Phylosymbiosis has also been found in animals, including gut microbiota in mammals[129] and the skin microbiota in Appalachian salamanders, where the skin microbiota provides pathogen-protective functions[141]. In aquatic environments, phylosymbiosis was observed in the skin microbiome of wild fish[142] and coral microbiomes[143]. Investigations of phylosymbiosis in plants have often been conducted to disentangle the effects of soil and host phylogeny on the microbiome[132]. Evidence of phylosymbiosis were found in maize and other *Poaceae* species[130,144], phylogenetically diverse angiosperm species[145], and a broad range of plants, including lycopods, ferns, gymnosperms and angiosperms along a tropical soil chronosequence[89].

1.5 Synthetic communities (SynComs) in studying plant-microbiota interactions

Understanding plant-microbiota interactions is critical for elucidating the mechanisms shaping these interactions, and for advancing sustainable agriculture. A powerful tool that can be used to study plant-microbiota interactions is microbial communities of reduced complexity and defined composition, termed synthetic communities (SynComs)[146]. Well-designed SynComs balance reproducibility, tractability and representation of the relevant features of the natural communities, allowing reconstruction of host microbiota under laboratory conditions[146]. Designing SynComs involves three steps: (1) strain isolation and establishment of culture collections capturing the diversities of natural microbial communities[78,84,105,147], (2) designing representative SynComs balancing complexity with tractability, and (3)

combining the selected strains into an input SynCom for deployment in a gnotobiotic system[146,148].

Over the past decade, culture collections from different hosts and environments have been established and SynComs have been used to study a multitude of plant-microbiota interactions[78,84,105,147]. In a study focusing on the functional overlap of the *Arabidopsis* leaf and root microbiota, bacterial culture collections for both *A. thaliana* root and leaf microbiota were established and subsequently used to test capacity of bacteria to colonize different plant compartments, showing that certain bacteria can relocate between compartments[147]. In another study, mono-kingdom and multi-kingdom SynComs, including bacteria, fungi, and oomycetes, have been used to investigate plant-microbe interactions as well as microbe-microbe interactions in *A. thaliana*[78]. SynComs were also used to investigate hos preference and invasiveness of commensal bacteria in the root microbiota of *Lotus japonicus* and *A. thaliana*[84]. In this research, in addition to investigating the host preference of the bacterial communities, the transcriptomic responses of both plants to their native (*A. thaliana*-derived for *A. thaliana*, *L. japonicus*-derived for *L. japonicus*) and non-native SynCom (derived from the other plant) in a soil-based gnotobiotic system were analyzed, deepening our understanding of plant responses to the commensal microbiota[84]. Recently, multiple randomly assembled SynComs from the *A. thaliana* leaf culture collection[147] were analyzed to identify microbial community patterns playing a role in plant protection[149]. Beyond culture collections of microbes associating with land plants, a *C. reinhardtii* culture collection exists[105]. Interestingly, SynComs derived from *A. thaliana* and *C. reinhardtii* culture collections can colonize both photosynthetic hosts, suggesting that the assembly of *A. thaliana* root and *C. reinhardtii* phycosphere communities is driven by some shared ecological principles despite the evolutionary distance between the two hosts[105].

1.6 Thesis objectives

Recently, it was shown that *C. reinhardtii* and a subset of other subaerial chlorophyte and streptophyte algae can assemble phycosphere communities from heterotrophic, soil-borne bacteria in a manner similar to land plants[105].

The first part of this work (Chapter 2) investigates the phycosphere communities of different subaerial algae. Here, we asked how widespread the ability to associate with soil-derived bacteria and fungi is among diverse subaerial green algae, and whether host identity influences phycosphere community composition. To this end, we chose a panel of different subaerial green algae, including the model alga *C. reinhardtii* and microscopic algae originating from the same environment, encompassing Chlorophyta (Chlorophyceae, Ulvophyceae, Trebouxiophyceae) and Streptophyta algae (Klebsormidiophyceae, Zygnematophyceae), with *Spirogloea muscicola* representing the algal lineage believed to be the closest extant relative to the most recent common ancestor of streptophyte algae and land plants[43]. First, we characterized the bacterial and fungal communities associated with the subaerial algae in a terrestrial environment. Second, we explored the temporal dynamics of the bacterial communities in controlled experiments in liquid cultures.

The second part of this work (Chapter 3) focuses on elucidating the mechanisms governing phycosphere microbiota establishment in soil-borne algae. While it is generally assumed that the secretion of photoassimilates is the main driver of microbial community assembly around unicellular algae[82,83,119], recent findings suggest that the provision of organic carbon compounds alone cannot fully explain microbiota assembly. Instead, the exchange of metabolites and/or molecular signals is likely necessary to establish and support the growth of the phycosphere microbiota[105]. To investigate this, we used transcriptomics to compare gene expression in *C. reinhardtii* cultivated axenically versus in a co-culture with the *C. reinhardtii*-derived SynCom. This approach allowed us to characterize mechanisms important for the alga-bacteria interactions and candidate genes for future functional studies.

1.7 Thesis outline and author contributions

This work is a monograph dissertation that contains two main chapters, one of which is a manuscript before submission for scientific review. The code and data supporting this thesis are available on Zenodo at https://doi.org/10.5281/zenodo.18289930.

Chapter 1, as presented above, provides an overarching introduction to the evolution of photosynthetic organisms, plant-microbiota and algae-microbiota interactions, patterns of phylosymbiosis in host-associated microbiota and application of SynComs to studying plant-microbiota interactions.

Chapter 2 contains a manuscript before submission for scientific review and investigates the phycosphere communities of different subaerial algae. In this project, I conducted the soil experiment with the subaerial algae and performed data analysis. The experiment in the liquid system was conducted by Paloma Durán and José Flores-Uribe.

Chapter 3 contains the comparison of gene expression of the model alga *C. reinhardtii* cultivated axenically and in a co-culture with the *C. reinhardtii*-derived SynCom. In this chapter, the mechanisms potentially important for the algae-bacteria interactions are characterized and a list of candidate genes involved in algae-bacteria interactions for further testing is included. In this project, I conducted the transcriptomics experiment with the help of José Flores-Uribe during the photobioreactor setup, and analyzed the transcriptomics and amplicon sequencing data. The Appendix to this chapter describes the preliminary results of experiments with the *C. reinhardtii* background strains of the CLiP mutant libraries[32–34], which provide a foundation for candidate gene function characterization. The results presented in this chapter will be combined with candidate gene function characterization to constitute a research manuscript.

2 Chapter 2:
Assembly of Phycosphere Microbiota from Soil-Borne Microbes Across Diverse Green Algae

2.1 Abstract

Green algae associate with microbes recruited from their environment. These interactions occur in the phycosphere, a region closely surrounding algal cells, which is analogous to the rhizosphere of land plants. Recent work showed that subaerial green algae can recruit and assemble their phycosphere microbiota from soil-derived bacteria. However, how widespread this ability is among diverse green algae and whether host identity shapes community composition remained unclear. Whether soil-borne fungi, in addition to bacteria, can be recruited to the phycosphere was also unknown. Here, we show that subaerial green algae ranging from Chlorophyta to Streptophyta assemble distinct phycosphere communities from soil-derived bacteria and fungi, and that host identity is the primary driver of phycosphere composition. In soil, the bacterial communities were strongly reshaped by algal hosts, whereas the fungal communities showed weaker and less consistent responses. To reduce environmental variability, we also examined bacterial community assembly in a controlled liquid-based system, where we detected a significant correlation between host phylogeny and bacterial community composition – a hallmark of phylosymbiosis.

Among the bacterial and fungal lineages present in algal phycospheres, we detected taxa known to associate with land plants, including Rhizobiales and Hypocreales, further supporting the finding that some microbial taxa can associate with a wide range of photosynthetic hosts. Our findings demonstrate that the capacity to assemble distinct phycosphere communities is a feature shared between subaerial green algae, possibly predating the divergence of Chlorophyta and Streptophyta. The stronger effect of host identity on bacterial than fungal community mirrors patterns observed in land plant microbiota, suggesting that principles of host-microbe interactions are conserved across evolutionarily distant photosynthetic hosts.

2.2 Introduction

In natural conditions, algae recruit microbes from their environment to their phycosphere, a region closely surrounding algal cells[104]. The phycosphere is an environment analogous to the rhizosphere of land plants, and at both interfaces, exudates alter the physicochemical environment, stimulating bacterial colonization[104]. Additionally, some bacterial lineages can be found at both interfaces[104,105]. Furthermore, in exchange for carbon and metabolites provided by the algae[108,116], microbial communities supply their photosynthetic hosts with micronutrients[75,76,112,121,150], can alleviate stresses[80,81] or provide protection against pathogens[78,79,115].

In the past, the research on interactions between microscopic algae and their phycosphere communities focused mainly on aquatic environments and the interactions with their associated bacteria[104,114,150,151]. It was estimated that in aquatic environments, algae and bacteria engage in metabolic interactions, with bacteria consuming even up to 50% of the carbon fixed by algae, and the algae profiting from the micronutrients provided by bacteria[108,112,113,117,121]. Algal photosynthesis is not only important in aquatic environments, but also in terrestrial environments, where soil algae take up carbon in amounts equal to ca. 31% of the global anthropogenic carbon emissions[152,153]. However, despite their importance for the global carbon cycle and mitigating climate change, and the dependence of many algae on bacterial presence for optimal growth[112,153], the interactions of soil-borne microscopic algae with their associated microbial communities remain underexplored.

In our previous work, we showed that the soil-derived[122] model alga *Chlamydomonas reinhardtii* can recruit and assemble a phycosphere bacterial community from soil-borne bacteria[105]. This ability was not unique to *C. reinhardtii*, but was also present in other chlorophyte and streptophyte algae, aligning with the hypothesis that the common ancestor of land plants and algae was preadapted for symbiosis, and that the ability to associate with microbial communities facilitated land colonization by plants[49,50,154]. However, several questions about the ability of subaerial algae to associate with soil-derived microbes remain unanswered. Here, we asked how widespread the ability to associate with soil-derived bacteria and fungi is among different subaerial green algae, and whether the host identity influences the

phycosphere community composition. To address these questions, we chose a broader set of different subaerial green algae, including the model alga *C. reinhardtii* and microscopic algae originating from the same environment, encompassing Chlorophyta (Chlorophyceae, Ulvophyceae, Trebouxiophyceae) and Streptophyta algae (Klebsormidiophyceae, Zygnematophyceae) (Supplementary Table S2.1), with *Spirogloea muscicola* representing the algal lineage believed to be the closest extant relative to the most recent common ancestor of streptophyte algae and land plants[43]. First, we characterized the bacterial and fungal communities associated with the subaerial algae in a terrestrial environment. Secondly, we explored the temporal dynamics in the bacterial communities in controlled experiments in liquid cultures. Thereby, w show that host identity is the main driver of phycosphere composition and that the host effect is more pronounced for bacterial than for fungal communities. Furthermore, we show that host-driven differences in bacterial community composition are substantial and consistent between the two experimental setups used. Our findings indicate that the ability to associate with soil-derived bacteria and fungi is widely found among subaerial green algae ranging from Chlorophyta to Streptophyta, and that host identity is the main driver of phycosphere composition.

2.3 Results

2.3.1 Microbial diversity in the native environment and reference soil

We first aimed to determine whether subaerial algae associate with bacterial and fungal communities in natural soil, and whether the bacterial and fungal community composition is correlated with host phylogeny. To this end, we selected a subset of subaerial algal strains originating from a shared terrestrial environment in the Eifel National Park. The collection includes chlorophyte and streptophyte algae belonging to different classes (Supplementary Table S2.1). Among them, *Spirogloea muscicola* represents a lineage believed to be the closest extant relative to the most recent common ancestor of streptophyte algae and land plants[43]. To determine whether the terrestrial algae associate with soil-derived fungi and whether we can observe a correlation between their phycosphere microbiome composition and host taxonomy, we designed experiments with natural Cologne Agricultural Soil (CAS), sourced near the Eifel National Park, which we used before[105].

We first compared the bacterial and fungal communities from the native environment of the algae (Eifel site) and CAS to assess the similarity between the communities. Since the Eifel site samples were taken from a site inhabited by various types of organisms, and CAS samples were taken from an unplanted soil, we expected the samples to be characterized by high within-sample diversities. In line with our hypothesis, alpha diversity indices were high and not significantly different between the two sites (Fig. 2.1a, b), indicating high biodiversity of the two sites. Despite similar alpha diversities, the microbiome structure differed between the sites for both bacteria and fungi, with site explaining 15.0% and 12.4% of variance in community dissimilarity (constrained Principal Component Analysis (PCoA) and permutation test, $P < 0.001$), respectively (Fig. 2.1c-d). We observed higher heterogeneity among the environmental Eifel samples than among the CAS samples. This can be attributed to the complexity of the environment, which is inhabited by multiple organisms, whereas the analyzed CAS samples were all taken from uniform, unplanted CAS.

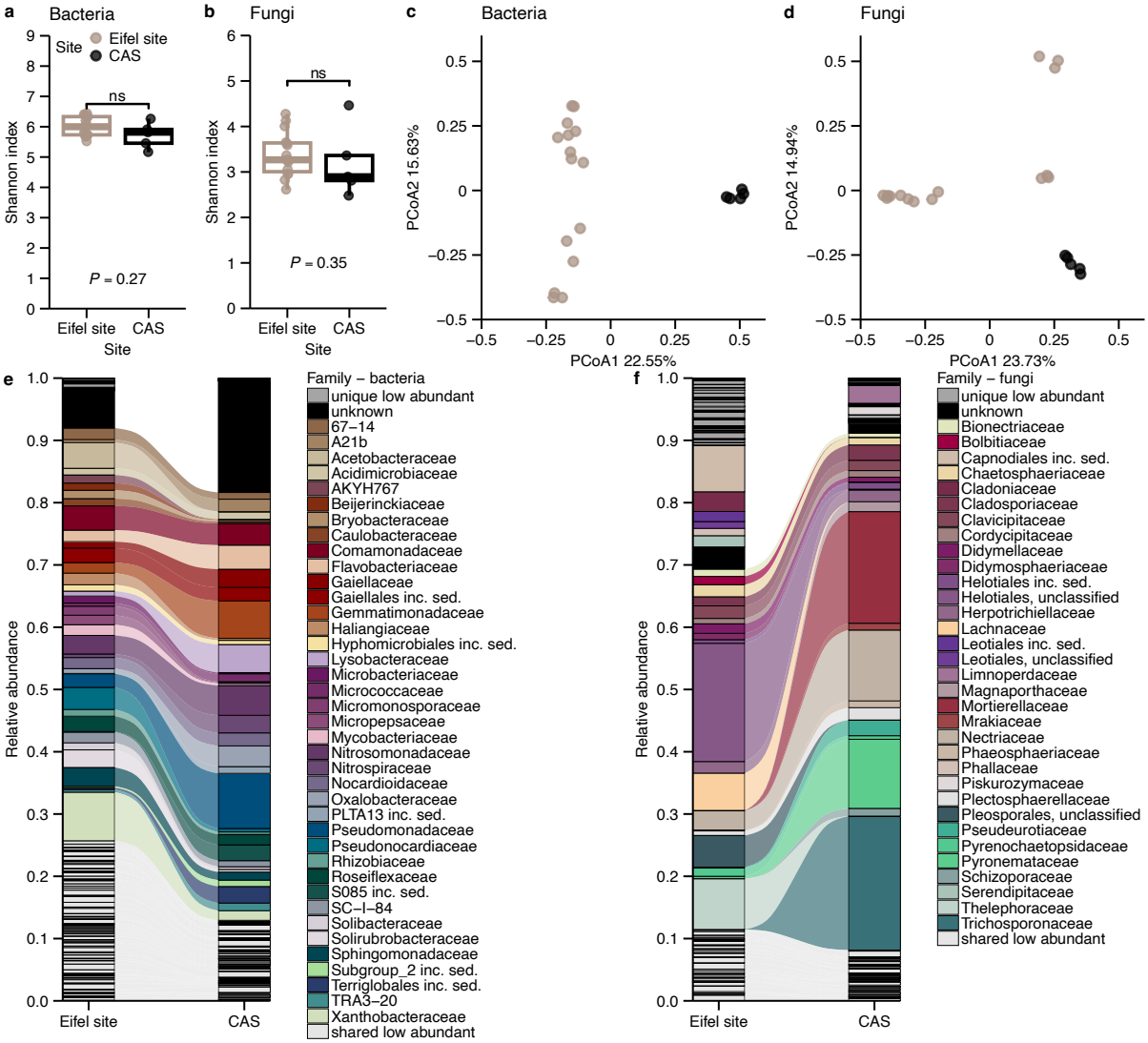


Figure 2.1 Comparison of bacterial and fungal community structures from the Eifel site and native CAS. a, b – Alpha diversity estimates of bacterial (a) and fungal (b) communities from Eifel site (n = 15) sample and CAS (n = 5). The two-sided Mann–Whitney test was used to assess significant differences among groups (P < 0.05). c, d – Principal Component Analysis of Bray–Curtis dissimilarities (ASV-level) of bacterial (c) and fungal (d) communities. with site explaining 15.0% and 12.4% of variance in community dissimilarity, respectively (constrained PCoA and permutation test, P < 0.001), The shared legend for panels a–d is presented in panel a. e, f – Alluvial plots showing the average relative abundances of bacterial (e) and fungal (f) families in Eifel site and CAS samples.

To explore which taxa drive the compositional differences between the Eifel site and CAS, we analyzed the taxonomic composition at the family level and identified families shared between the two sites. Bacterial families shared between the sites accounted for 92% of the aggregated relative abundance (RA) in the Eifel samples and 81.6% in CAS samples (Fig. 2.1e). Although most bacterial families were found in both sites, their aggregated RA differed (Fig. 2.1e, Supplementary Figure S2.1). Caulobacteraceae and Xanthobacteraceae were less abundant in CAS, whereas Lysobacteraceae and

Pseudomonadaceae were more abundant, which might reflect the past agricultural use of CAS[155]. The differences in the taxonomic composition were more pronounced in the fungal community (Fig. 2.1f, Supplementary Figure S2.2). Fungal families shared between the sites accounted for 69.3% aggregated RA in Eifel and 91.1% in CAS, and some of the highly abundant families were unique to one of the sites. Compared to bacteria, the shifts in RAs are more pronounced, and the dominant fungal families differed markedly between the sites, with the RAs of the three most abundant families in CAS (Mortierellaceae, Pyronemataceae, Trichosporonaceae) summing up to 50.5%, whereas in the Eifel samples, they sum up to only 0.4%. This observation indicates that fungal communities are more site-specific than bacterial ones, as previously demonstrated for fungal communities associated with land plants[91,156]. Analysis of the bacterial and fungal communities of the Eifel site and CAS showed that the communities of both sites are highly diverse. Although they differ at the low taxonomic level, they share dominant bacterial phyla and families, indicating that bacterial taxa broadly representative of the algae's native environment are present in CAS. In contrast, the fungal communities showed greater site-specificity with less compositional overlap, but some similarities between the two sites were observed

2.3.2 *C. reinhardtii* shapes bacterial and fungal communities in soil

We previously observed the ability of the model alga *C. reinhardtii* to associate with soil-derived bacteria, showing parallels to the root microbiota assembly in land plants[105]. To examine whether *C. reinhardtii* also influences soil fungal communities, we inoculated CAS-filled pots with axenic alga cultures and collected samples after the alga had grown substantially on the soil surface (see Materials and methods). We then characterized both bacterial and fungal microbiomes by 16S and ITS2 rRNA amplicon sequencing. Consistent with previous findings, the *C. reinhardtii* phycosphere bacterial community clustered separately from both input CAS samples and the no alga controls (Fig. 2.2a), recapitulating our previous findings[105].

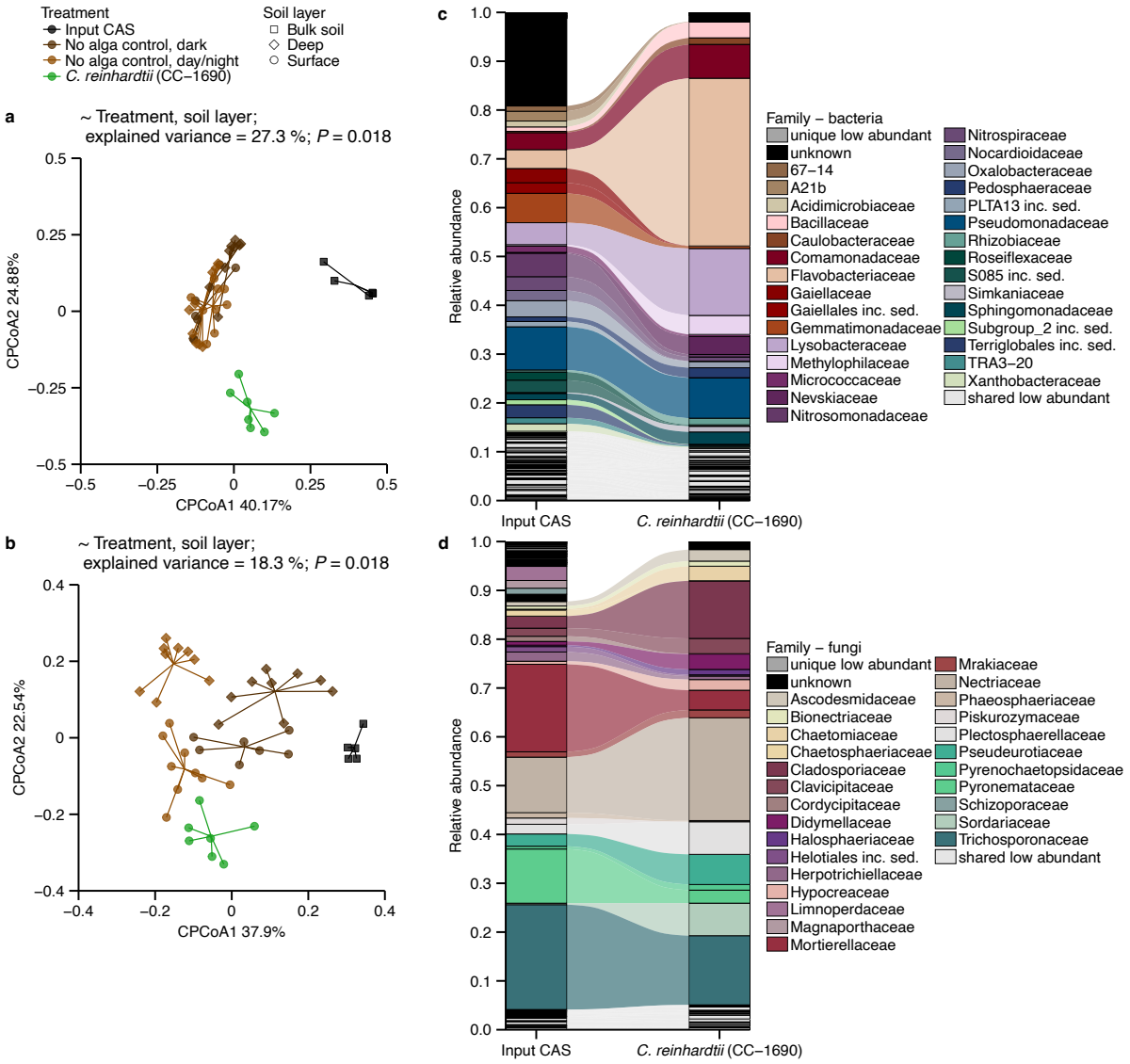


Figure 2.2 *C. reinhardtii* associates with bacteria and fungi derived from native CAS. a, b – ASV-level beta-diversity analysis (Principal Component Analysis of Bray–Curtis dissimilarities, constrained by treatment and soil layer combined as a single variable) of bacterial (a, 27.3% of the variance; $P < 0.001$) and fungal communities (b, 18.3% of the variance; $P < 0.05$) from the input CAS ($n = 5$), no alga dark control: deep soil layer ($n = 9$) and surface soil ($n = 7$), no alga day/night control: deep soil layer ($n = 9$) and surface soil ($n = 9$) and *C. reinhardtii* phycosphere ($n = 7$) samples. Points are color-coded by treatment, with point shapes corresponding to the soil layer. c, d – Alluvial plots showing the average relative abundances of bacterial (c) and fungal (d) families in the input CAS samples and *C. reinhardtii* phycosphere samples.

Similarly, fungal community profiles showed significant separation between the *C. reinhardtii* phycosphere, input CAS, and no alga controls (Fig. 2.2b). However, the separation was smaller for fungi than for bacteria: the combined "Treatment, soil layer" factor explained 18.3% of variance in the fungal composition versus 27.3% in bacterial composition (Fig. 2.2a, d), indicating a weaker host effect on the fungal community structure. Compared to the input CAS, Flavobacteriaceae and

Lysobacteraceae were more abundant in *C. reinhardtii* phycosphere (Fig. 2.2c, Supplementary Figure S2.3). Among fungi, three of the families most abundant in CAS, Pyronemataceae, Mortierellaceae, and Trichosporonaceae were strongly depleted in *C. reinhardtii* phycosphere (Fig. 2.2d, Supplementary Figure S2.4). Together, these results recapitulate previous observations for bacterial communities associated with *C. reinhardtii* and demonstrate that the model alga can also alter soil fungal communities.

2.3.3 Subaerial algae reshape soil bacterial and fungal communities

To test whether the ability to assemble phycosphere microbiota from soil-borne bacteria and fungi is conserved across algal lineages, we inoculated CAS pots with individual strains from our subaerial algae collection and characterized their bacterial and fungal phycosphere communities. Because *C. reinhardtii* associates with CAS-borne bacteria and fungi, this system provided a suitable framework to test whether other algae recruit soil microbes similarly to the model alga *C. reinhardtii*. During the experiment, substantial algal growth was visible on the soil surface, which was confirmed by measuring total chlorophyll content in surface soil samples (Supplementary Figure S2.5). Samples from the algae-inoculated pots were taken either when the algae grew similarly to *C. reinhardtii* on day 21, or the latest on day 35, when some algal growth was starting to be visible on the soil surface of the no alga control pots exposed to light.

Based on our observations in the experiment with *C. reinhardtii*, we hypothesized that the subaerial algae would recruit bacteria and fungi from soil, forming distinct phycosphere communities. Indeed, in beta diversity analysis, bacterial communities were separated according to algal presence along the first axis, and by host taxonomy along the second (Fig. 2.3a, Supplementary Figure S2.6a). In a single-factor Constrained Analysis of Principal Coordinates (CAP), the combined "Treatment, soil layer" factor explained the largest variance proportion (27.7%, $P < 0.001$), while light regime and total chlorophyll content explained smaller but detectable fractions of variance (4.9% and 4.5%, not statistically significant) (Fig. 2.3a), indicating the main driver of the clustering in the system was the host. Because each variable was evaluated independently, these estimates represent associations rather than unique variance

fractions. To disentangle host effects from environmental drivers, we used PERMANOVA analysis to compare algal phycosphere samples to surface soil samples from the day/night no alga controls. Bacterial community composition differed between the individual phycospheres and the no alga controls, with treatment explaining 33.9% variance. This effect was not statistically significant when permutations were restricted within pots (see Materials and methods), but its magnitude (33.9% variance explained) and consistency across analyses suggest a biologically meaningful host-associated signal.

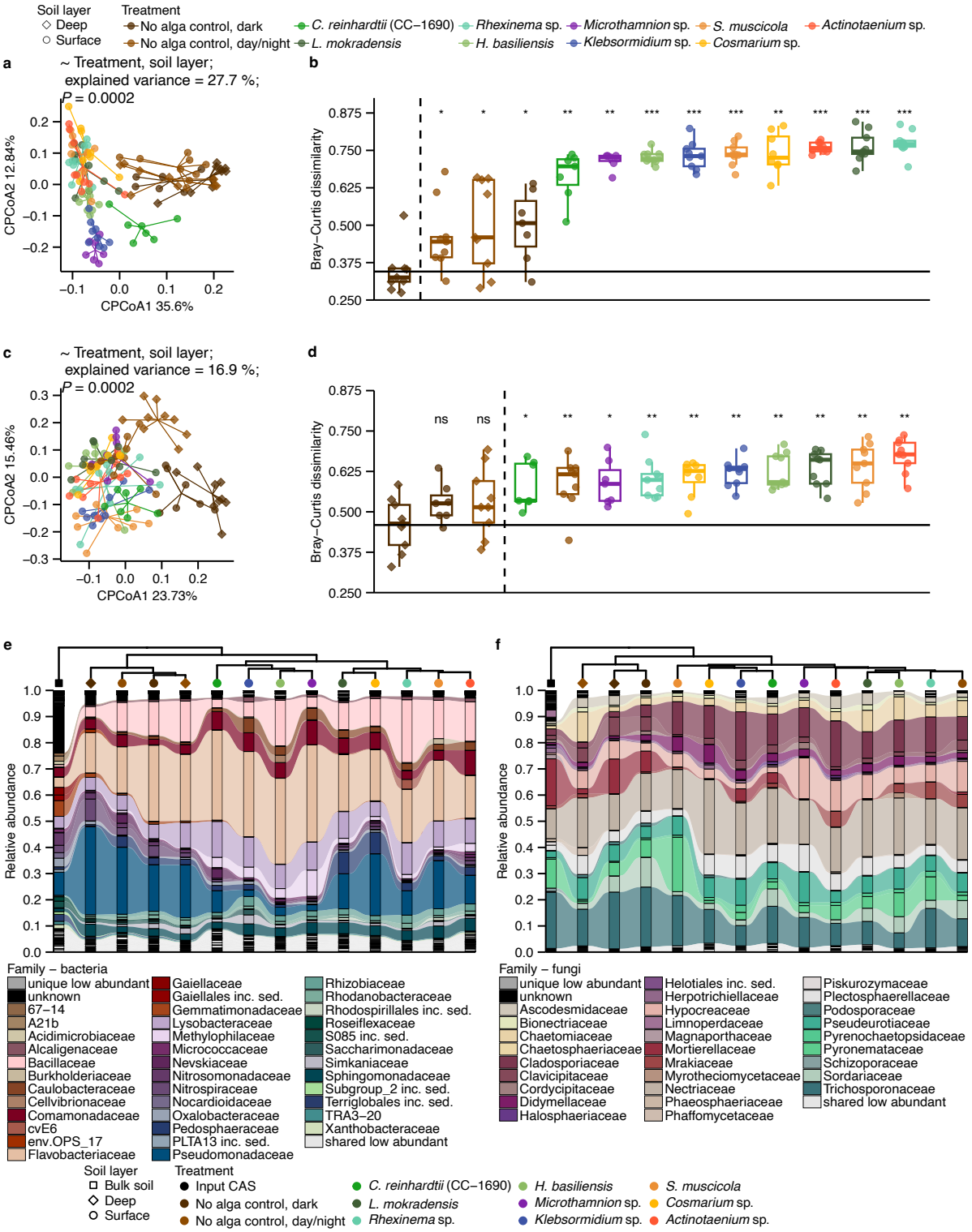


Figure 2.3 Host taxonomy influences bacterial community structure in soil, and only minimally the fungal community. a – ASV-level beta-diversity analysis (Principal Component Analysis of Bray–Curtis dissimilarities, constrained by treatment and soil layer combined as a single variable, 27.7% of the variance; $P < 0.001$) of bacterial communities from the no alga control samples and phycosphere samples. b – ASV-level analysis of Bray–Curtis dissimilarities of bacterial communities from no alga control and phycosphere samples, compared to the averaged no alga control, dark, deep soil sample. Significant differences are marked with stars (two-sided Mann–Whitney test, with Benjamini-

Hochberg correction; ns – not significant, $P > 0.05$, * – $0.05 \geq P > 0.01$, ** – $0.01 \geq P > 0.001$, *** – $0.001 \geq P > 0.0001$, **** – $P < 0.0001$), and samples are color-coded based on the treatment, with point shapes corresponding to the soil layer. The dashed vertical line separates the experimental conditions not significantly different from the averaged no alga control, dark, deep soil from those significantly different, and the horizontal line shows the average distance of the individual no alga control, dark, deep soil samples to the averaged one. $n = 7$ for no alga control surface soil and *C. reinhardtii* phycosphere, $n = 8$ for *Microthamnion* sp., otherwise $n = 9$. c – ASV-level beta-diversity analysis (Principal Component Analysis of Bray–Curtis dissimilarities, constrained by treatment and soil layer combined as a single variable, 16.9% of the variance; $P < 0.05$) of fungal communities from the no alga control samples and phycosphere samples. d – ASV-level analysis of Bray–Curtis dissimilarities of fungal communities from no alga control and phycosphere samples were compared to the averaged no alga control, dark, deep soil sample. Significant differences were analyzed and are marked as in panel b. e, f – Alluvial plots showing the overlap between bacterial (e) and fungal (f) communities of native CAS, no alga controls, and algal phycospheres. The dendrograms show the result of ASV-level analysis of Bray–Curtis dissimilarities of average community composition (per treatment and soil layer) compared by hierarchical clustering.

To quantify how much the phycosphere samples differed from deep soil not exposed to light, we computed amplicon sequence variant (ASV)-level Bray–Curtis dissimilarities relative to a reference no alga control (incubated in the dark, deep soil compartment), which showed moderate but significant separation between no alga control samples from different compartments (mean dissimilarities < 0.5). In contrast, all phycospheres diverged strongly from the control (mean dissimilarities of ca. 0.7 across algal hosts, Fig. 2.3b). Similar significant differences were observed when comparing phycosphere samples to a reference surface control (day/night incubation), further supporting a host-dependent effect on the soil bacterial community (Supplementary Figure S2.5b). In both comparisons, we did not observe grouping according to algal division. Although we detected a host-dependent effect on the soil bacterial community, the ASV-level Bray–Curtis dissimilarities did not correlate significantly with algal phylogenetic distances (Mantel Spearman $r = 0.13$, $P = 0.17$, Supplementary Table S2.2).

Fungal communities showed a similar, but less pronounced response (Fig. 2.3c). In contrast to bacterial communities, clustering along axis 1 in beta diversity was driven by light regime (axis 1, Fig. 2.3c). However, in a single factor CAP, the main factor explaining variance in the system was again “Treatment, soil layer”, and samples also separated according to host taxonomy (axis 2, Fig. 2.3c), indicating the host importance for the fungal community structure. The host-dependent effect was slightly stronger for Streptophyta than for Chlorophyta (Supplementary Figure S3.6b), and within Chlorophyta, the model *C. reinhardtii* phycosphere samples separated

from those of other Chlorophyta hosts. This was also true for bacterial communities, suggesting that host origin might influence the phycosphere community composition (Supplementary Figure S3.6). Similarly to the analysis of bacterial communities, we compared algal phycosphere samples to surface soil samples from the day/night no alga controls to disentangle host effects from environmental drivers. Fungal community composition differed between the individual phycospheres and the no alga controls, with treatment explaining 16.6% variance. Similar to the bacterial community analysis, even though the effect was not statistically significant when permutations were restricted within pots, the combined results from the analyses indicate a biologically meaningful host-associated effect on the fungal community composition. To further examine the host effect, we computed ASV-level Bray–Curtis dissimilarities to a reference no alga control (incubated in the dark, deep soil compartment), which revealed significant differences for both control and phycosphere samples. However, unlike for bacterial communities, algal presence did not universally increase dissimilarity compared to the no alga control surface soil exposed to light (Fig. 3d, Supplementary Figure S2.5b, c). Significant dissimilarities of the phycosphere samples to the surface control (day/night incubation) reference sample were detected only for two Streptophyta hosts, *Klebsormidium* sp. and *S. muscicola* (Supplementary Figure S3.5c). Consistent with these observations, the ASV-level Bray–Curtis dissimilarities did not correlate significantly with algal phylogenetic distances (Mantel Spearman r = 0.12, P = 0.23, Supplementary Table S2.2).

These observations of the host effect being smaller for fungal than for bacterial communities mirror what was observed in past research on root microbiota[90,157]. However, past comparative studies did not consistently show bacterial or fungal communities to be more affected by host identity, with some studies reporting fungi to be more affected by plant species[158], while others observed host species having a larger effect on bacterial than on fungal communities, which were more affected by soil type[90,157].

After observing the separation of bacterial and fungal communities based on host presence and taxonomy, we sought to identify taxa driving the compositional changes. We first examined RA shifts in bacterial and fungal families. Bacterial communities in

the no alga control samples clustered together, separate from the phycosphere samples (Fig. 2.3e). However, within the phycospheres, Chlorophyta and Streptophyta phycospheres were not strongly differentiated. In comparison to the input CAS, bacterial taxa commonly associating with plants and green algae increased in abundance[84,87,105]. Bacterial families that increased the most in the algal phycospheres samples compared to the no alga controls included Caulobacteraceae and Rhizobiaceae (Supplementary Figure S2.7). Conversely, Pseudomonadaceae were abundant in the no alga soil controls, but in the phycospheres, their abundance only moderately increased or decreased in comparison with the input CAS (Supplementary Figure S2.7), with several Pseudomonadaceae ASVs being excluded from the phycospheres (Supplementary Figure S2.8). The tested algae showed some shared patterns in their phycosphere community structures. While several Pseudomonadaceae ASVs, which were highly abundant in the no alga controls, were fully excluded from the phycospheres, Methylophilaceae increased in abundance more strongly in the algal phycospheres than in the controls, and some ASVs were commonly enriched for multiple algal hosts (Fig. 2.3e, Supplementary Figure S2.7–S2.9).

While the bacterial communities from the no alga controls clearly separated from the phycosphere communities, fungal communities were grouped mainly by the light regime. Input CAS clustered with dark-incubated controls, while day/night-incubated controls clustered together with the phycosphere samples (Fig. 3f). Taxonomic shifts included pronounced changes in the fungal families dominant in CAS: Mortierellaceae, Trichosporonaceae and Pyronemataceae (Fig. 2.3f, Supplementary Figure S3.10, S3.11). Mortierellaceae RA declined from 17.9% in the input CAS to 0.04–4.8% in the phycospheres, while Trichosporonaceae RA was at a level similar to the initial only for *S. muscicola*. Cladosporiaceae RA increased in the phycospheres 3.8–7.5 times.

Together, these results show that patterns observed for *C. reinhardtii* are found in other subaerial algae. Algal hosts strongly modify soil bacterial communities, and, to a lesser extent, fungal communities.

2.3.4 Temporal dynamics of the *C. reinhardtii* phycosphere community

To investigate temporal dynamics of bacterial communities, we used a previously established liquid-based system[105], that allows controlled observation of microbial community assembly over time. As in the soil-based system, we first analyzed the phycosphere community of *C. reinhardtii*. *C. reinhardtii* established a distinct phycosphere community, clearly separating from the input CAS-derived bacteria and the no alga controls (Fig. 2.4a). The taxonomic composition of the community progressively changed over time (Fig. 2.4b).

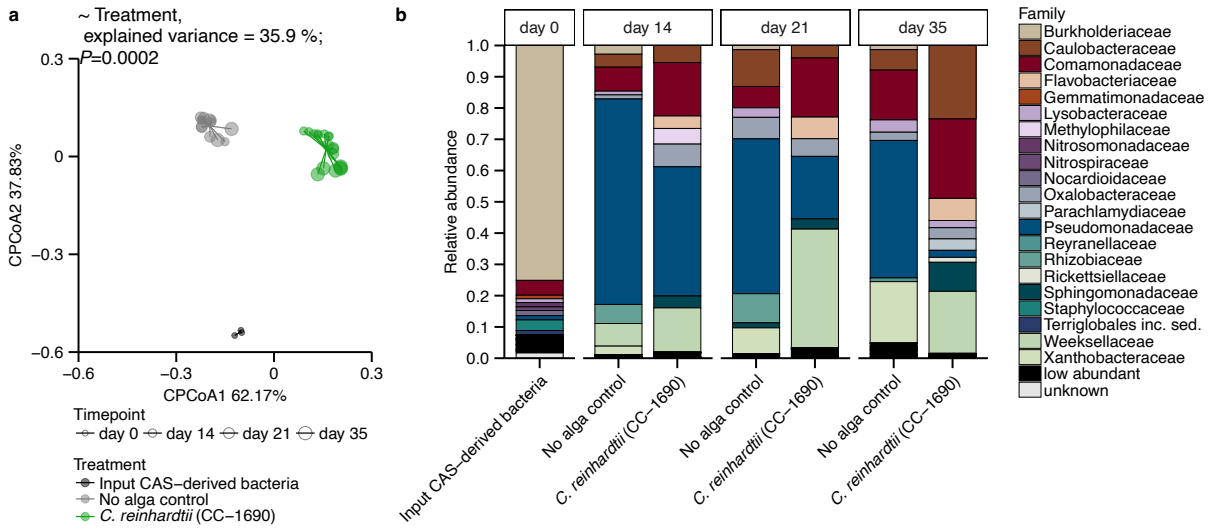


Figure 2.4 *C. reinhardtii* establishes a unique phycosphere bacterial community from CAS-derived bacteria in a mesocosm system. a – ASV-level beta-diversity analysis (Principal Component Analysis of Bray–Curtis dissimilarities, constrained by treatment, 35.9% of the variance; $P < 0.001$) of bacterial communities from samples obtained from a liquid-based system. The samples are colored-coded based on the treatment: input CAS-derived bacteria (black, $n = 3$), no alga control (gray; $n = 5$ for day 14 and 21 and $n = 3$ for day 35) and *C. reinhardtii* (green; $n = 8$ for day 14, $n = 4$ for day 21 and $n = 6$ for day 35) and the point size corresponds to the timepoint. b – Barplots of the average relative abundances of bacterial families in the input CAS-derived bacteria samples, no alga control samples, and *C. reinhardtii* phycosphere samples.

When comparing relative abundances between treatments, the ratio of Pseudomonadaceae RA ($RA_{phycosphere}/RA_{control}$) decreased over time — from 0.63 on day 14 to 0.05 on day 35. Xanthobacteraceae RA showed a sharp decline, almost disappearing by day 14 (Fig. 2.4b, Supplementary Figure S2.12). Among taxa whose RA increased over time, Sphingomonadaceae showed a moderate rise, while Flavobacteriaceae and Weeksellaceae were detected only in the phycosphere samples at the end of the experiment and were absent from the controls (Fig. 2.4b,

Supplementary Figure S2.11). Together, these results confirm that *C. reinhardtii* assembles a distinct phycosphere community in the liquid system that undergoes compositional changes over time. This progressive differentiation of the bacterial communities, combining selective elimination of some taxa and host-specific enrichment of others, suggests active filtering by *C. reinhardtii* rather than stochastic drift.

2.3.5 Host taxonomy is associated with phycosphere composition in the liquid system

Knowing that subaerial algae can modify microbial communities in soil, we next expanded the analysis to the subaerial algae collection to investigate how the different algae shape bacterial communities in the liquid system over time. Consistent with observations for *C. reinhardtii* and the soil system experiment, all algae assembled distinct phycosphere communities (Fig. 2.5a). Interestingly, in the comparison of phycosphere amplicon profiles, samples from Chlorophyta and Streptophyta algae generally separated along the first ordination axis and clustered by host identity along the second, indicating a strong influence of the host on the bacterial community composition (Fig. 2.5a). In contrast, we observed little separation by timepoint, suggesting that phycosphere community composition already stabilized by day 14 (Fig. 2.5a). At the ASV level, Bray–Curtis dissimilarities relative to a reference no alga control revealed strong and significant differences between the reference sample and the phycosphere samples (Fig. 2.5b), with mean dissimilarities of the phycospheres to the control per algal host ranging between 60% and 95%, further supporting the importance of the host for the bacterial community structure. ASV-level Bray–Curtis dissimilarities showed a strong correlation with algal phylogenetic distances (Mantel Spearman $r = 0.66$, $P = 0.008$, Supplementary Table 2.6), indicating a phylosymbiotic signal under simplified, controlled conditions. In soil, the algal phylogenetic distances were not significantly correlated with either bacterial or fungal community dissimilarities, suggesting the observed phylosymbiosis patterns are context-dependent and can be observed under highly controlled conditions.

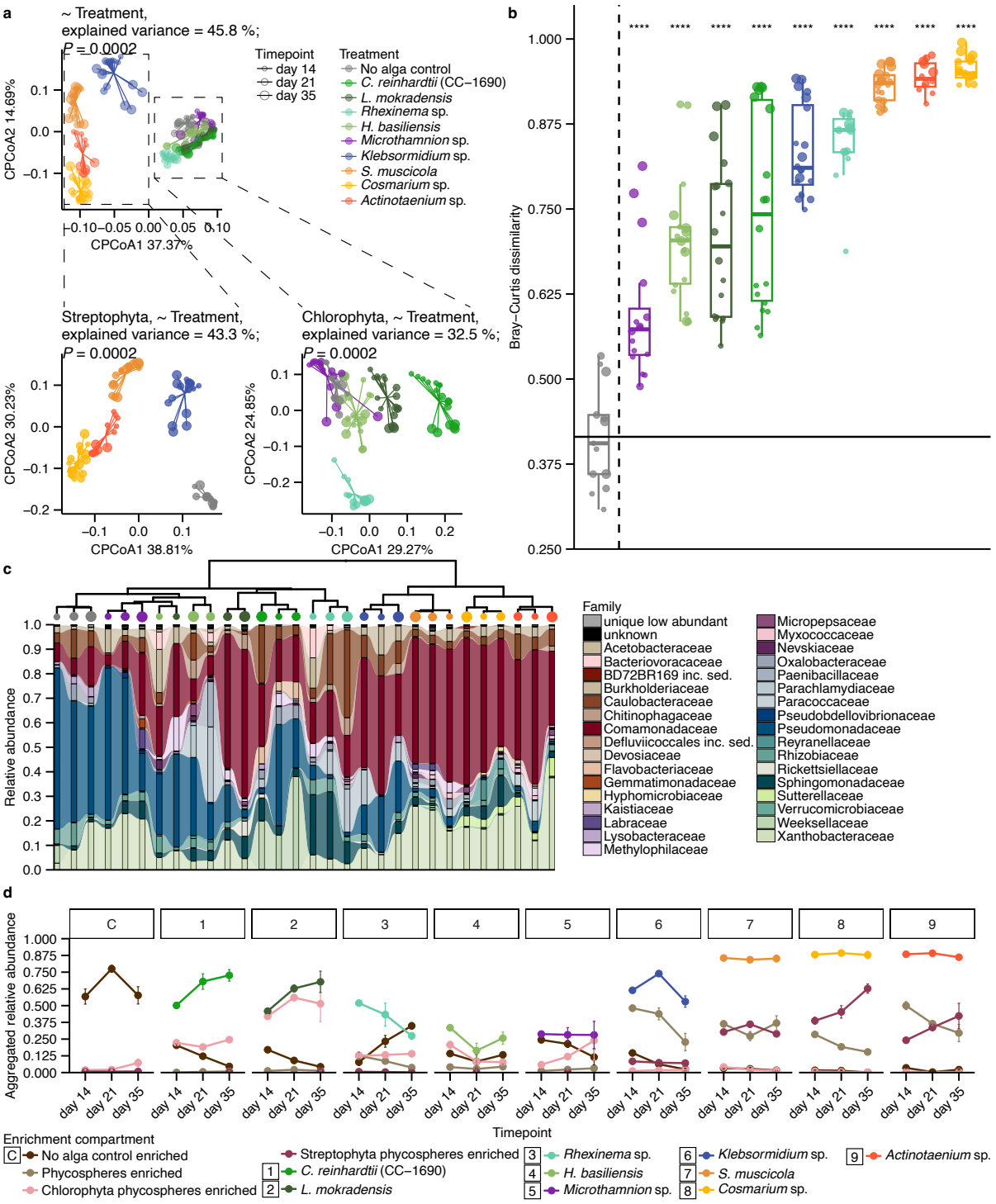


Figure 2.5 The phycosphere bacterial community depends on the algal host taxonomy. a – ASV-level beta-diversity analysis of bacterial communities from samples obtained from a liquid-based system, (per treatment and timepoint, *n* between 3 and 9, Principal Component Analysis of Bray–Curtis dissimilarities constrained by treatment, 45.8% of the variance; $P < 0.001$). The dashed lines show separate analyses for no alga control and Streptophyta (left, 43.3% of the variance; $P < 0.001$) or Chlorophyta (right, 32.5% of the variance; $P < 0.001$). b – ASV-level analysis of Bray–Curtis dissimilarities of bacterial communities from no alga control and phycosphere samples, compared to the averaged no alga control sample. Significant differences are marked with stars (two-sided Mann–Whitney test, with Benjamini-Hochberg correction; ns – not significant, $P > 0.05$, * – $0.05 \geqslant P > 0.01$,

** – 0.01 ⩾ P > 0.001, *** – 0.001 ⩾ P > 0.0001, **** – P < 0.0001), and samples are color-coded based on the Treatment, with point sizes corresponding to the timepoint. The dashed vertical line separates the Treatments not significantly different from the averaged no alga control from those significantly different, and the horizontal line shows the average distance of the individual no alga control soil samples to the averaged one. c – Alluvial plot showing the overlap between bacterial communities of no alga controls and algal phycospheres. The dendrograms show the result of ASV-level analysis of Bray–Curtis dissimilarities of average community composition (averaged per treatment and timepoint) grouped by hierarchical clustering. d – Dynamics of relative abundances of ASVs enriched over time in no alga controls, all phycospheres, Streptophyta or Chlorophyta phycospheres, and single alga phycospheres, color-coded by the enriched compartment. The plot is faceted by treatment: C – no alga control, 1 – *C. reinhardtii* (CC-1690), 2 – *L. mokradensis*., 3 – *Rhexinema* sp., 4 – *H. basiliensis*, 5 –*Microthamnion* sp., 7 – *Klebsormidium* sp., 8 – *S. muscicola*, 9 – *Cosmarium* sp., 10 – *Actinotaenium* sp.

Temporal dynamics of phycosphere assembly varied between hosts: Zygnematophyceae established distinct communities early on, whereas phycosphere communities of the Chlorophyceae (*C. reinhardtii*, *L. mokradensis*) continued to change over time (Fig. 2.5b). For some hosts, the Bray–Curtis dissimilarities, which show the extent the phycosphere community differs from the control, the communities were very different already on day 14 (*S. muscicola* and the two Desmidiales algae, *Actinotaenium* sp. and *Cosmarium* sp.), indicating these algae select for specific microbes early on. However, at the end of the experiment (day 35) we observed Bray–Curtis dissimilarities close to 0.9, which is similar to the dissimilarities of *S. muscicola*, *Actinotaenium* sp. and *Cosmarium* sp. phycospheres, also for other algae, including another Streptophyta alga, *Klebsormidium* sp., and two Chlorophyceae algae: *L. mokradensis* and the model alga *C. reinhardtii*. Slightly lower dissimilarities were observed for *Rhexinema* sp., and for the other algae, the mean dissimilarities were lower. These observations indicate that the extent to which algae modify the bacterial community, as well as the temporal dynamics of the phycosphere community assembly, differ between hosts and are not more similar within the taxonomic division.

We then sought to identify the bacterial taxa driving the compositional changes in the communities. We first analyzed RA changes and observed that Pseudomonadaceae dominated no alga control samples, but were depleted in the phycospheres of *S. muscicola*, *Klebsormidium* sp., *Cosmarium* sp. and *Actinotaenium* sp., where Comamonadaceae were up to 169-fold more abundant than Pseudomonadaceae (Fig. 5c, Supplementary Figure S2.15). The depletion of Pseudomonadaceae in the phycospheres was consistent with the observations in the soil system, suggesting that despite differences between the experimental systems, some patterns in the

phycosphere community assembly are universal (Fig. 2.3, Fig. 2.5, Supplementary Figure S2.15).

To further characterize the bacterial taxa driving the differences between the no alga control and the phycospheres, we performed an enrichment analysis (one-sided Wilcoxon test, BH-adjusted $P < 0.05$). Enrichment analysis identified ASVs enriched in the no alga control, all phycospheres, phycospheres of Chlorophyta or Streptophyta hosts, and individual host-specific ASVs (Fig. 2.5d, Supplementary Figure S2.16). ASVs enriched in the no alga control accounted for 57–76% RA in the no alga control samples, but below 35% in the phycospheres (Fig. 2.5d). Only three ASVs were enriched for all algal hosts, suggesting only a minimal shared core of bacteria associating with all the investigated algae. Some ASVs were consistently enriched per algal division, indicating algal-division similarities rather than a large universal phycosphere microbiota. However, as algal hosts differed in growth rates and biomass accumulation, we cannot fully disentangle phylogenetic effects from differences in resource availability.

Generally, Chlorophyta-enriched ASVs represented 6–56% RA in Chlorophyta phycospheres but less than 5% in Streptophyta phycospheres. Conversely, Streptophyta-enriched ASVs accounted for 7–63% RA in Streptophyta phycospheres but were almost absent from the Chlorophyta phycospheres. The observed patterns indicate a strong phylogenetic signal in the community composition. Similarly to the differences observed in the temporal dynamics of Bray–Curtis dissimilarities from the no alga control (Fig. 5b), the temporal trajectories of the host-enriched ASVs also varied among hosts. Some algae (*S. muscicola*, *Cosmarium* sp. and *Actinotaenium* sp.) established stable phycosphere communities early on, suggesting strong host influence on the microbial community. In contrast, Chlorophyceae (*C. reinhardtii*, *L. mokradensis*) showed a gradual increase in host-enriched ASVs over time, reaching up to 73% RA. Other algal hosts displayed a more variable or declining pattern.

Together, these results show that in the liquid system, host phylogeny was the major determinant of bacterial phycosphere composition. While the variance explained by CPCoA is not directly comparable between the liquid and soil systems due to differing

constraints, host-driven differences in phycosphere community composition appeared stronger in the liquid system than in the more complex soil system. Additionally, we observed varying temporal dynamics among hosts.

2.4 Discussion

Our results demonstrate that diverse subaerial algae, spanning Chlorophyta and Streptophyta, assemble distinct bacterial phycosphere communities in both soil and liquid medium, with host identity as the primary driver of community composition. This extends and generalizes our previous findings[105] to a broader algal panel and reveals that subaerial algae can assemble phycosphere communities not only from soil-borne bacteria but also from soil-borne fungi. Although the fungal communities were host-associated, host identity explained less variance in the fungal than the bacterial community composition, and algal presence did not consistently increase the fungal community dissimilarity from no alga controls – unlike the strong, consistent effect observed for bacteria.

In soil, the bacterial phycosphere communities diverged strongly from the no alga control when algae were cultivated on soil in day/night conditions (Fig. 2.3). The presence of all algal hosts increased dissimilarity compared to the no alga control surface soil exposed to light, and the host identity was the main driver of the differences between the bacterial community compositions. This indicates that in the complex soil environment, subaerial algae exert selective pressure on bacterial community assembly. While the soil bacterial communities were strongly reshaped by the presence of algae, the host effect on the fungal community was smaller. Although host identity was the main factor explaining variance in the system, algal presence did not universally increase dissimilarity relative to the no alga control surface soil exposed to light. Notably, the effect sizes observed for the bacterial and the fungal communities were consistent across single-factor constrained ordination and PERMANOVA analyses, indicating a robust host-associated signal despite reduced inferential power when permutations were restricted within individual pots. These observations are consistent with previous studies in algae, demonstrating that host identity is the main driver of phycosphere and root microbiota composition[89,145,159,160],

and our data align with studies showing stronger host effects on bacterial than on fungal communities[90,157].

The weaker fungal response to the host may reflect several factors. Fungal communities may be primarily determined by the soil properties rather than host identity. Additionally, fungi typically exhibit slower colonization dynamics than bacteria[161], and our sampling timepoints may have been too early to capture full fungal community differentiation.

Among the bacterial and fungal lineages more abundant in algal phycospheres than in no alga controls, we found taxa known to associate with land plants, such as members of Rhizobiales (Hyphomicrobiales), and Hypocreales[87,105], respectively. This supports the finding that some bacterial taxa can associate with a wide range of photosynthetic hosts, and some of these patterns can also be observed in fungi, consistent with the hypothesis that preadaptation for symbiosis was present in the common ancestor of algae and land plants[50].

The host effect on the bacterial community was not limited to the terrestrial environment and was stronger in a more controlled, liquid-based system, where the bacterial community composition was highly dependent on host identity (Fig. 2.5). While we did not detect a significant phylosymbiotic signal in soil, the correlation between the bacterial community composition and algal phylogenetic distance was strong and significant in the liquid system. This suggests that phylosymbiosis patterns may emerge under controlled conditions, where environmental heterogeneity does not obscure host-driven assembly. The liquid-based system can be more tightly controlled, and algal growth in liquid medium is more uniform than in soil, both of which may facilitate detection of host-driven phycosphere assembly. In previous research, the *C. reinhardtii* phycosphere community composition was established within the first 4 days from inoculation and remained stable until the end of the experiment (day 11)[105]. Although the algae we included in our experiments grew more slowly than *C. reinhardtii*, we also observed phycosphere communities remaining relatively stable between days 14 and 35, indicating that the subaerial algae can assemble stable phycosphere communities relatively early (Fig. 2.5). Together, these results show that in the liquid system, host phylogeny was the major determinant of the bacterial

phycosphere composition. The host-driven differences in the bacterial community composition appeared more pronounced in the liquid system than in the soil system, possibly because environmental heterogeneity in soil likely reduces the detectability of host effects relative to the controlled liquid system, and differences in algal growth rates among host species may influence the magnitude and timing of host effects.

Taken together, the results from both experimental systems indicate that subaerial green algae can assemble distinct phycosphere communities across different environments, and in both cases, the community composition is mainly influenced by host identity. These findings suggest that the capacity to assemble distinct phycosphere communities is a feature shared between subaerial green algae, possibly predating the divergence of Chlorophyta and Streptophyta. The weaker and more variable fungal response to the algal hosts aligns with patterns also observed in land plant microbiota[90,157], suggesting that the way bacteria and fungi interact with algae and land plants is similar across photosynthetic hosts.

2.5 Materials and methods

2.5.1 Algal phylogeny

Subaerial algae were initially identified based on their morphology and their identity was confirmed by Sanger sequencing of 18S or rbcL sequences. DNA was extracted from the individual algal cultures either with the FastDNA™ SPIN Kit for Soil following the manufacturer's instructions (MP Biomedicals, Solon, USA) or using the SPRI beads method. All fragments were amplified in reactions with 2 U SuperHotTaq DNA Polymerase, 1× incomplete buffer, 2 mM MgCl2 (Bioron GmbH, Ludwigshafen, Germany), 0.3% BSA, 0.2 mM dNTPs (ThermoFisher Scientific), and 0.3 μM primers. 18S fragment was amplified with primers EukA and EukB[162] or CrN1F[163] and EukB. rbcL fragments were amplified with BP22-1F, BP23-724R[164], RH1[165] and 1385R[166], or in a 2-step PCR with MaGo1F-MaGo-3R and MaGo2F-MaGo-3R[167]. All the primers used for amplification and sequencing are in Supplementary Table S2.3 and PCR programs are in Supplementary Table S2.4. The sequencing was performed by Eurofins Genomics, Germany. Low-quality bases were identified by visual inspection and the sequences were trimmed in 4Peaks software. The trimmed high-quality sequences were assembled in UGENE[168] using the function "Reads de novo assembly

(with CAP3)" and blastn[169,170] (2.14.1) was used to find best hits to confirm algal identity. Molecular identification results were compared with morphological observations and, in some cases, an alternative taxonomic assignment was made based on the combined evidence (Supplementary Table S2.5). Preliminary Sanger sequencing (18S rRNA) and BLAST searches of strain MEL1034B did not yield a confident identification. Although phylogenetic analysis of the fragment suggested a closer relationship to *Mesotaenium* sp., the lack of corroborating molecular markers and repeated difficulties in DNA extraction and amplification, including unsuccessful amplification of rbcL sequences using multiple primer sets, prevented definitive taxonomic resolution. To avoid potential misclassification, the isolate was excluded from the present study and will be re-evaluated after additional genomic data become available. All subaerial algae used in the experiments are summarized in Supplementary Table S2.1.

2.5.2 Algal culture conditions

C. reinhardtii was cultured in TP medium, pH 7.0 either at 25 °C, and the illumination of 125 µmol $m^{-2}s^{-1}$ under continuous light conditions with shaking on a rotatory shaker at 70 RPM, or, for the soil experiment, under 16 h/8 h light/dark cycle without shaking[105]. Subaerial microalgae (isolated as described previously[105] were cultured in modified Synthetic Freshwater Medium (SFM, Supplemental Table S2.6)[105] under 16 h/8 h light/dark cycle without shaking. Cell growth was determined either by measuring samples in a Multisizer 4e Coulter counter (Beckman Coulter Inc., California, USA) particle counter with the Beckman Coulter Multisizer software (v4.03) or using an Infinite M200Pro (TECAN Austria GmbH, Grödig, Austria) plate reader with the TECAN i-control software (v2.0.10.0), to determine either absorbance at 750 nm or chlorophyll fluorescence (excitation 440/9 nm, emission 680/20 nm).

2.5.3 Eifel site samples collection

Multiple samples were collected in June 2022 from a wet rock surface from which the subaerial algae were isolated (the Eifel site)[43,105]. Multiple samples were taken from different spots on the rock. Samples were collected using a metal spatula, wiped thoroughly with 70% ethanol, and transported to the laboratory in a cooled box and stored at −20 °C until further processing.

2.5.4 Soil experiment setup

6 cm × 6 cm plastic pots were filled to fullness with Cologne Agricultural Soil (CAS). Each CAS pot was placed on the lower part of a sterile Petri dish (90 mm diameter). Before inoculation, each Petri dish was filled with 25 mL sterile MiliQ water to pre-moisturize the soil. *C. reinhardtii* cultures were centrifuged (1,900 × g, 5 min, room temperature (RT)), washed three times with sterile 10 mM $MgCl_2$ (1900 × g, 5 min, RT), and resuspended in 10 mL sterile 10 mM $MgCl_2$. Cell density of the algal suspension was measured in the Multisizer 4e Coulter counter, diluted with sterile 10 mM $MgCl_2$ to a cell density of 2.25×10^6 cells mL^{-1} , and the chlorophyll fluorescence of the suspension was determined. Subaerial algae were centrifuged (800 × g, braking and acceleration speed 2, 15 min, RT), washed once with sterile 10 mM $MgCl_2$ (same centrifugation settings), and resuspended in 30 mL sterile 10 mM $MgCl_2$. Chlorophyll fluorescence of the algal resuspensions was determined, and if needed, the suspensions were diluted to a chlorophyll fluorescence of *C. reinhardtii* 2.25×10^6 cells mL^{-1} suspension. Chlorophyll fluorescence was used as a proxy of algal culture cell densities[105]. If the fluorescence was lower than that of of *C. reinhardtii* 2.25×10^6 cells mL^{-1} suspension, the suspensions were used directly to inoculate the pots. Directly before the inoculation, soil in each pot was moistened with 10 mL sterile MiliQ water, carefully applied with a pipette to ensure uniform moisture distribution. To inoculate, algal suspensions or sterile 10 mM $MgCl_2$ were carefully applied to the soil surface using a pipette to ensure uniform distribution. The no alga control, dark pots were covered with aluminum foil. The pots were placed on a tray, covered with a plastic cover with holes covered with parafilm, and transported to a climate chamber, where 25 mL sterile MiliQ water was added to each Petri dish. The pots were incubated for up to 35 days under 16 h/8 h light/dark cycle. Soil moisture, algal growth, and general soil condition were assessed periodically. The position of the pots in the trays was shuffled weekly to minimize edge and location effects. Water was added to the Petri dishes under each pot or gently applied to the soil surface. Samples from the algae-inoculated and control pots were taken either when the algae grew similarly to *C. reinhardtii* on day 21 (predominantly green soil surface) or at the latest on day 35, when some algal growth was observed in the control pots. Soil samples were collected using a sterile metal spatula. For each pot, a single spatula was used, and between taking replicate samples from each pot, it was washed with ethanol. In addition to soil

samples for microbial profiling, surface soil samples were obtained for chlorophyll extraction as a proxy for the algal growth.

2.5.5 Liquid-based experiment setup

Microbial soil wash was prepared as described previously[105]. Briefly, 5 g CAS were manually resuspended in 30 mL of sterile 1× Tris-EDTA (TE) supplemented with 0.1% Triton X-100 (SERVA Electrophoresis GmbH, Heidelberg, Germany), homogenized by inversion in a rotary mixer (40 RPM, 30 min, RT). Bigger soil particles were removed by centrifugation for 1 min at 450 × g and the supernatant was centrifuged again at 3000 × g for 20 min. Afterwards, the final microbial soil wash was prepared by resuspending the pellet in 50 mL of SFM. Subaerial algae pre-cultured for two weeks in SFM were inoculated to a similar chlorophyll fluorescence to that equivalent to *C. reinhardtii* CC-1690 culture with a cell density of 1×10^5 cells mL^{-1} in SFM. To each flask, 1 mL of the microbial soil wash was added, which equals to ca. 1×10^9 cells (estimate based on Multisizer 4e Coulter counter measurements). As a "no alga control", SFM inoculated with an estimate of 1×10^9 cells from the microbial soil wash was used. The cultures were incubated for 35 days under 12/12 h light/dark cycle without shaking and samples for DNA extraction were collected at days 0, 7 (excluded during bioinformatical analysis because of the failure to pass quality control), 14, 21 and 35.

2.5.6 DNA extraction from soil and algal samples

Total DNA was extracted from the Eifel site and soil samples using the FastDNA™ SPIN Kit for Soil following the manufacturer's instructions (MP Biomedicals, Solon, USA). DNA samples were eluted in 50 µL nuclease-free water and used for microbial community profiling after adjusting DNA concentration to 5 ng/µL.

2.5.7 DNA extraction from liquid samples by alkaline lysis

Bacterial DNA from liquid samples was extracted using alkaline lysis, which we previously determined to be efficient for extracting DNA from liquid samples[105,147]. Briefly, 12 µL of the sample were diluted in 20 µL of Buffer I (NaOH 25 mM, EDTA(Na) 0.2 mM, pH 12), mixed by pipetting, and incubated at 94 °C for 30 min. Next, 20 µL of Buffer II (Tris-HCl 40 mM, pH 7.46) was added to the mixture and stored at –20 °C.

2.5.8 SPRI beads preparation

To prepare SPRI beads for DNA extraction, SpeedBeads magnetic carboxylate-modified beads (Cytiva 65152105050250) were added to a tube and placed on a magnet. After 5 min binding, the supernatant was discarded, and the beads were washed twice with 1 volume 1× E (10 mM Tris, 1 mM EDTA). The tube was removed from the magnet, the beads were mixed with 1 volume 1× TE, added to 50 volumes NaCl-PEG solution, and carefully mixed. After preparation, their performance was tested against commercial AMPure XP beads by mixing them with 0.5× GeneRuler 1 kb DNA Ladder (ThermoScientific™) in different ratios, performing two washes with 80% ethanol, eluting, loading on 1% agarose gel, and running for 20 min at 150 V. SPRI beads were stored in the dark at 4 °C until use.

2.5.9 SPRI beads DNA extraction

Samples were ground in a Bertin Precellys Evolution HP Homogenizator (7,600 rpm, 20 s, RT) three times, with snap-freezing in liquid nitrogen between rounds. After grinding, the samples were shortly centrifuged at 4 °C and stored at –80 °C until grinding of all the samples was finished. To extract DNA, 800 µL of preheated lysis buffer (Supplementary Table S2.7) (water bath, 55 °C, 20 min) was added to each sample, mixed thoroughly by hand, and frozen in liquid nitrogen. Samples were ground again in the Bertin Precellys Evolution HP Homogenizator (7,600 rpm, 1 min, RT) and incubated at 65 °C for 1 h. After incubation, they were mixed by inversion ~20x and spun down (10,000 × g, 5 min, RT). 500 µL of each sample was transferred to a deep well plate already filled with 187.5 µL 5 M potassium acetate. The plates were centrifuged at 3214 × g for 10 min, and 480 µL of each supernatant were transferred into clean plates. 288 µL SPRI beads (well mixed before, RT) were added to each well, mixed by inversion ~20×, briefly spun down, and incubated at RT for 5 min. The plate was placed on the magnet until the solution was clear. Supernatants were removed, and each sample was washed twice with 80 % ethanol. The beads were air dried for 30 s, the plate was removed from the magnet, and 20 µL nuclease-free water were added to each well. After mixing the solution well, the plate was placed back on the magnet, and the clear solution was transferred to a new plate. The eluted DNA was used for microbial profiling.

2.5.10 Chlorophyll extraction from soil samples

Soil samples were oven-dried (60 °C) in 15 mL-falcon tubes for several days until the soil weight in a given sample was constant, indicating the soil was fully dry. Chlorophyll extraction was performed by solvent extraction and shaking. Briefly, a spatula tip of $CaCO_3$ (Carl Roth, Karlsruhe, Germany) was added to each soil sample, as well as 6 mL of DMSO (Sigma-Aldrich, Darmstadt, Germany). Samples were boiled for 90 min in a water bath (65 °C) and then shaken for 20 min on a horizontal shaker (RT). After placing the tubes in an upright position and allowing the soil to settle, the supernatant was transferred to a new tube, and the procedure was repeated. Merged supernatants were centrifuged (10 min, 3000 × g, 15°C). Absorption (700 nm, 665 nm, 648 nm) was measured using the plate reader, and total chlorophyll content was calculated according to the adapted formula Chl *a* + *b* [µg/g soil]=([(A665-A700) × 8.02 + (A648 – A700) × 20.2] × DF × S)/DSW, where DF – dilution factor, S – amount of solvent [ml], DSW – dry soil weight [g][171].

2.5.11 16S rRNA amplicon sequencing

16S rRNA amplicon sequencing from the liquid system experiment was performed as described previously[105]. 16S rRNA from the Eifel site and soil samples was amplified using a two-step PCR amplification protocol. In the first step, V5–V7 region of the bacterial 16S rRNA was amplified in a 25 µL reaction volume containing 2 U SuperHotTaq DNA Polymerase, 1× incomplete buffer, 2 mM $MgCl_2$ (Bioron GmbH, Ludwigshafen, Germany), 0.3% BSA, 0.2 mM dNTPs (ThermoFisher Scientific), and 0.3 µM primers 799F[172] and 1192R[173] (sequences in Supplementary Table S2.3). PCR was performed (initial denaturation at 94 °C/2 min, followed by 25 cycles of 94 °C/30 s, 55 °C/30 s, 72 °C/30 s and final elongation at 72 °C/5 min). Afterwards, single-stranded DNA and proteins were digested by adding 1 µL Antarctic Phosphatase, 1 µL Exonuclease I and 3 µL Antarctic Phosphatase buffer (New England BioLabs GmbH, Frankfurt, Germany) to each sample. Reactions were incubated at 37 °C for 30 min, followed by enzyme deactivation (85 °C, 15 min) and centrifugation (3,000 rpm, 10 min, 4 °C). 4.8 µL of the supernatant was used as the DNA template for a second PCR, prepared in the same way as described but reaction volume of 40 µL, primers including barcodes and Illumina adapters, and cycles reduced to 10. PCR

product concentration was assessed using PicoGreen™ assay (Quant-iTTM PicoGreenTM dsDNA reagent, Life Technologies Corporation, ThermoFisher Scientific): 2 µL PCR product was mixed with 0.04 µL Pico 200x and 72.96 µL 0.5x TE, incubated for 2-5 min covered from light, and measured in Infinite M200Pro (TECAN Austria GmbH, Grödig, Austria) plate reader with the TECAN i-control software (v2.0.10.0) (excitation 485/9 nm, emission 535/20 nm). Samples were pooled based on their fluorescence intensities. The pooled sample was loaded on 1.5% agarose gel and run for 40 min, either at 130 V (medium-sized gel) or 90 V (small gel). Afterwards, the bands were excised and DNA was purified using the Gel and PCR-Clean-up kit (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany). DNA concentration was determined from fluorescence measurements with the QuantiFluor® dsDNA System (Promega GmbH) according to the manufacturer's protocol. The quality assessment of the library and paired-end Illumina sequencing on NovaSeq PE250 platform was performed by Novogene GmbH.

2.5.12 ITS2 rRNA sequencing

ITS2 rRNA from the Eifel site and soil samples was amplified according to a two-step PCR amplification protocol. In the first step, the ITS2 rRNA region was amplified in a 10 µL reaction volume containing 2 U SuperHotTaq DNA Polymerase, 1× incomplete buffer, 2 mM $MgCl_2$ (Bioron GmbH, Ludwigshafen, Germany), 0.3% BSA, 0.2 mM dNTPs (ThermoFisher Scientific) and 0.3 µM tagged primers ITS2_F1 and ITS2_R1[91] (sequences in Supplemental Table S2.3). PCR was performed (initial denaturation at 94 °C/2 min, followed by 30 cycles of 94 °C/30 s, 59 °C/30 s, 72 °C/30 s and final elongation at 72 °C/5 min). Afterwards, single-stranded DNA and proteins were digested by adding 0.1 µL Antarctic Phosphatase, 0.05 µL Exonuclease I and 3.85 µL nuclease-free water to each sample. Reactions were incubated at 37 °C for 30 min, followed by enzyme deactivation (85 °C, 15 min) and centrifugation (3000 rpm, 10 min, 4 °C). 5 µL of the supernatant was used as the DNA template for a second PCR, prepared in the same way as described but reaction volume of 15 µL and primers including barcodes and Illumina adapters. PCR was performed (initial denaturation at 94 °C/2 min, followed by 10 cycles of 94 °C/30 s, 65 °C/30 s, 72 °C/30 s and final elongation at 72 °C/5 min). PCR product concentration was assessed using PicoGreen™ assay (Quant-iTTM PicoGreenTM dsDNA reagent, Life Technologies

Corporation, ThermoFisher Scientific): 2 µL PCR product was mixed with 0.04 µL Pico 200x and 72.96 µL 0.5x TE, incubated for 2-5 min covered from light and measured in Infinite M200Pro (TECAN Austria GmbH, Grödig, Austria) plate reader with the TECAN i-control software (v2.0.10.0) (excitation 485/9 nm, emission 535/20 nm). Samples were pooled based on their fluorescence intensities. The pooled library was re-concentrated twice with AMPure XP beads (Beckman Coulter Inc.) and loaded on 1.5% agarose gel and run for 40 min, either at 130 V (medium-sized gel) or 90 V (small gel). Afterwards, the bands were excised and DNA was purified using the Gel and PCR-Clean-up kit (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany). DNA concentration was determined from fluorescence measurements with the Quantus™ Fluorometer and QuantiFluor® dsDNA System (Promega GmbH) according to the manufacturer's protocol. The quality assessment of the library and paired-end Illumina sequencing on NovaSeq PE250 platform was performed by Novogene GmbH.

2.5.13 Analysis of 16S rRNA amplicon profiling

Amplicon sequencing data were processed using the QIIME2 2024.10 pipeline[174]. Raw reads were imported into QIIME2, demultiplexed according to their barcode sequences, trimmed, filtered with seqkit[175], and orphan reads were removed before merging. Merged reads were denoised with DADA2 within QIIME2 (qiime dada2 denoise-single with --p-trunc-len 0 to retain full-length reads) to infer unique amplicon sequence variants (ASVs) after error correction and chimera removal[176].

For taxonomic classification, a V5–V7 region-specific 16S rRNA gene classifier was trained using the SILVA 138.2 reference database[177,178] and the RESCRIPt plugin for QIIME2[174]. Briefly, full-length SSURef sequences were downloaded, RNA sequences were reverse-transcribed to DNA, and low-quality sequences containing ≥5 ambiguous bases or homopolymers of length ≥8 bp were excluded. Sequences were length-filtered by taxonomic domain with minimum lengths of 900 bp (Archaea), 1200 bp (Bacteria), and 1400 bp (Eukaryota) and dereplicated by identical sequence and taxonomy combinations (qiime rescript dereplicate, uniq mode). The V5–V7 region was extracted with the primer pair AACMGGATTAGATACCCKG (forward) and ACGTCATCCCCACCTTCC (reverse), restricting fragment lengths to 250–450 bp.

After additional dereplication to minimize redundancy, the naïve Bayes classifier was trained and evaluated on the dereplicated V5–V7 fragments (qiime rescript evaluate-fit-classifier).

The trained classifier was used to assign taxonomy to the ASVs to generate the abundance tables. The abundance tables were then imported into R (v4.4.0)[179], which was used for all the downstream analyses. Non-bacterial reads and reads assigned to bacterial ASVs present in the initial algal cultures were excluded from the tables.

2.5.14 Analysis of ITS2 rRNA amplicon profiling

The first round of amplicon sequencing data demultiplexing was performed by Novogene GmbH. Next, data were demultiplexed according to their barcode sequence using Cutadapt[180], and reads were merged and denoised with DADA2 within QIIME2 (qiime dada2 denoise-single with --p-trunc-len 0 to retain full-length reads[176]). Unique ASVs were inferred from reads after error correction and chimera filtering. Taxonomic annotation of ASVs was performed using the naïve Bayesian classifier trained on UNITE data[181]. To train the classifier, reference sequences and taxonomy files were retrieved using the qiime rescript get-unite-data command with the parameters --p-taxon-group eukaryotes, --p-cluster-id dynamic, and --p-singletons[182]. To improve classifier efficiency, SH accession identifiers were removed from the taxonomy strings. The full-length ITS naïve Bayes classifier was then trained and validated using qiime rescript evaluate-fit-classifier. To generate an abundance table, raw reads were mapped to the inferred ASVs. The resulting tables after excluding non-fungal reads were subsequently used for analyses of diversity and community composition using R (v4.4.0)[179].

2.5.15 Statistics and reproducibility

For each experiment, the number of samples (*n*) is given in the figure legends.

For analyses integrating bacterial (16S) and fungal (ITS2) data, only samples with successful sequencing of both 16S and ITS2 were kept for further analyses. In the soil experiment, a small subset of soil samples exhibited extreme dominance of the fungal order *Pezizales*, with relative abundances up to 86%. Although *Pezizales* were observed at moderate abundances in other samples across treatments, these extreme

profiles formed a distinct subset and deviated from the overall community structure. Such dominance patterns might disproportionately influence multivariate analyses and obscure treatment-level effects. Because the pot constituted the experimental unit, all samples originating from pots exhibiting such extreme dominance were excluded collectively from downstream analyses. The exclusion criterion was based on the magnitude of *Pezizales* dominance rather than a fixed abundance threshold.

For comparisons, we used a two-sided, non-parametric Mann-Whitney test and when appropriate, *P*-values were adjusted for multiple comparisons using the Benjamini-Hochberg method.

For analysis of alpha and beta diversities, rarefied ASV tables were used, whereas relative abundances were computed from non-rarefied data. Community structure was explored using principal coordinates analysis (PCoA) based on Bray–Curtis dissimilarities. For the soil experiment data, to quantify the association between environmental and experimental variables and microbial community composition, constrained analysis of principal coordinates (CAP; vegan::capscale) was applied. Each explanatory variable (Treatment × soil layer, algal class, algal presence/absence, light regime, and chlorophyll content) was evaluated in a separate single-factor model of the form: Bray–Curtis ~ X. These models were interpreted descriptively to estimate associations between individual variables and community composition. Statistical significance tests on beta-diversity analyses were performed using anova.cca function from the *vegan*[183] package with 5000 permutations. The variance explained by each constraint was calculated as the ratio of the constrained inertia to the total inertia. To visualize treatment-associated community structure, CAP was performed with "Treatment,soil layer" as the constraining variable. Because algal treatments were applied at the pot level and multiple subsamples were collected per pot, permutation tests were restricted within pots using block permutations (vegan::anova.cca with permute, n = 5000). CAP ordinations were used for visualization and effect-size estimation. To isolate host-specific bacterial and fungal responses independent of environmental variation, analyses were restricted to phycosphere samples collected under identical conditions (surface soil, day/night light regime). Bray–Curtis dissimilarities were subset to the corresponding samples without recomputation. Host effects on fungal community composition were evaluated using PERMANOVA

(vegan::adonis2) with algal treatment (individual alga/no alga control) as the explanatory variable. Two permutation schemes were applied: (i) permutations constrained within pots to account for non-independence of subsamples, yielding effect size estimates, and (ii) unrestricted permutations to formally test for host-associated differences in community composition. In both cases, marginal effects were evaluated using 9,999 permutations.

To assess whether microbial community similarity reflected algal evolutionary relatedness, we tested for phylosymbiosis using Mantel correlations between microbial Bray–Curtis dissimilarities and algal phylogenetic distances in R (v4.4.0)[179] using package *ape* (5.8-1)[184]. Because marker-gene phylogenies (18S rRNA and rbcL) did not consistently resolve relationships among all algal strains and produced unstable topologies, we constructed a conservative, taxonomy-informed phylogenetic tree reflecting accepted higher-level relationships within Chlorophyta and Streptophyta. Branch lengths were assigned using Grafen's method[185] and converted to an ultrametric tree to ensure comparability of phylogenetic distances. Phylogenetic distances were extracted as a cophenetic distance matrix and used exclusively for correlation analyses, without evolutionary inference. Microbial community dissimilarities were calculated using Bray–Curtis distances on group-averaged, rarefied ASV tables. Mantel tests were performed with 9,999 permutations using both Pearson and Spearman correlations to account for linear and monotonic relationships, respectively. Analyses were conducted separately for soil bacterial (16S), soil fungal (ITS), and liquid-culture bacterial (16S) communities. This approach has several limitations: the phylogenetic framework is based on taxonomy rather than sequence-derived branch lengths, and therefore reflects relative relatedness rather than evolutionary divergence times, and Mantel tests have limited statistical power at small sample sizes. Nevertheless, this analysis provides a conservative assessment of whether broad algal phylogenetic relatedness is associated with systematic differences in bacterial and fungal community composition across experimental contexts.

Whenever boxplots were used in figures, data were represented as median values (horizontal line), interquartile ranges (boxes, spanning Q1–Q3), and whiskers extending to the most extreme values within 1.5 × the interquartile range.

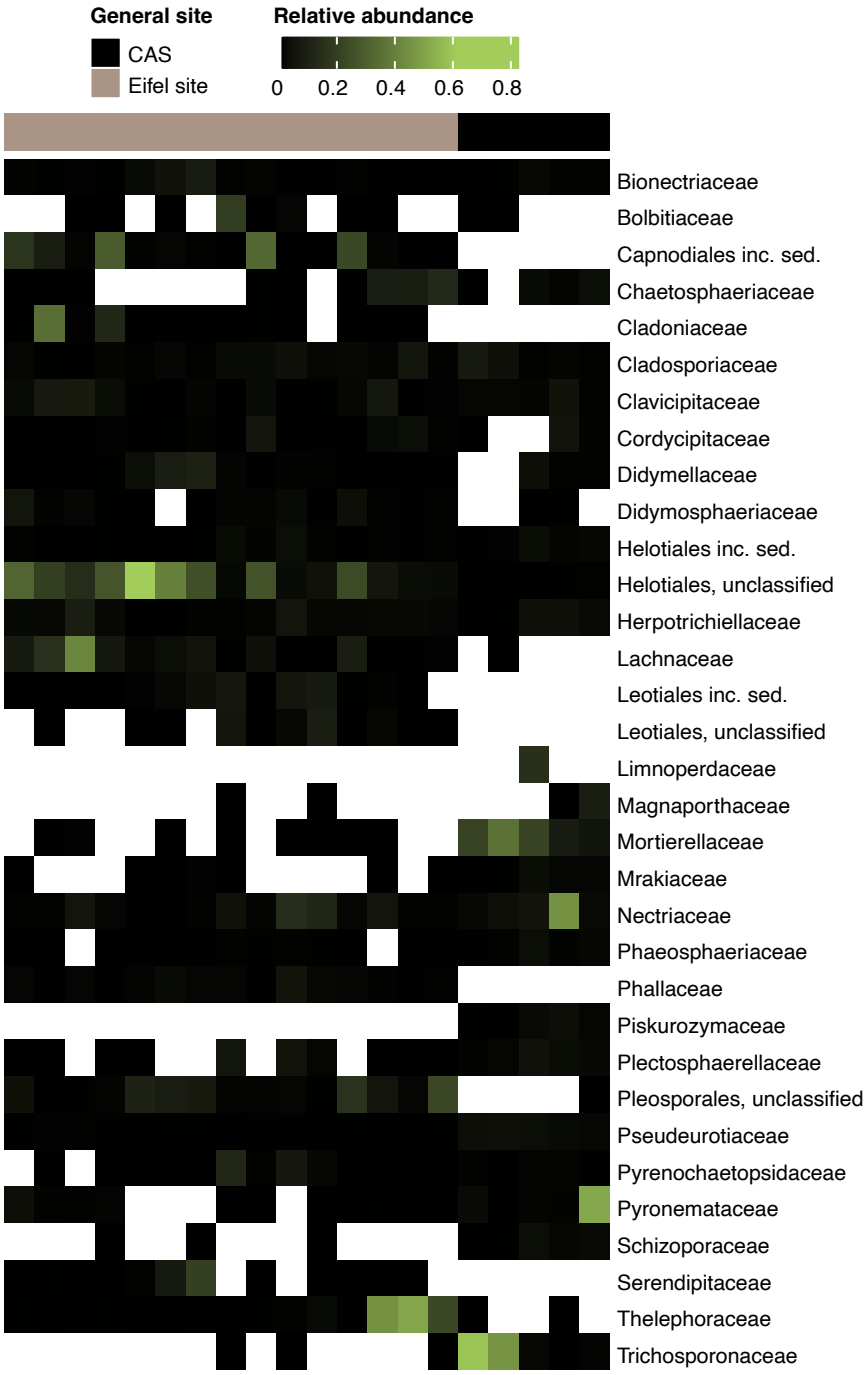
To identify bacterial ASVs significantly enriched in algal phycospheres relative to control treatments, or in control relative to phycospheres, we performed a series of Mann–Whitney tests on ASV relative abundances. ASVs were filtered to retain only those with a relative abundance ≥ 0.5 % in at least one sample, ensuring that low-abundance taxa did not inflate multiple-testing corrections. For each ASV, we tested whether its relative abundance was significantly more abundant than in the no alga control samples in (i) each individual algal phycosphere compartment, (ii) all phycosphere samples combined, and (iii) subsets corresponding to Streptophyta or Chlorophyta hosts; and if it was significantly more abundant in the no alga control than in all phycosphere samples combined (no alga control enriched). *P*-values were adjusted using the Benjamini-Hochberg method, and ASVs with adjusted *P*-values < 0.05 were classified as significantly enriched.

2.5.16 Transparency and AI use

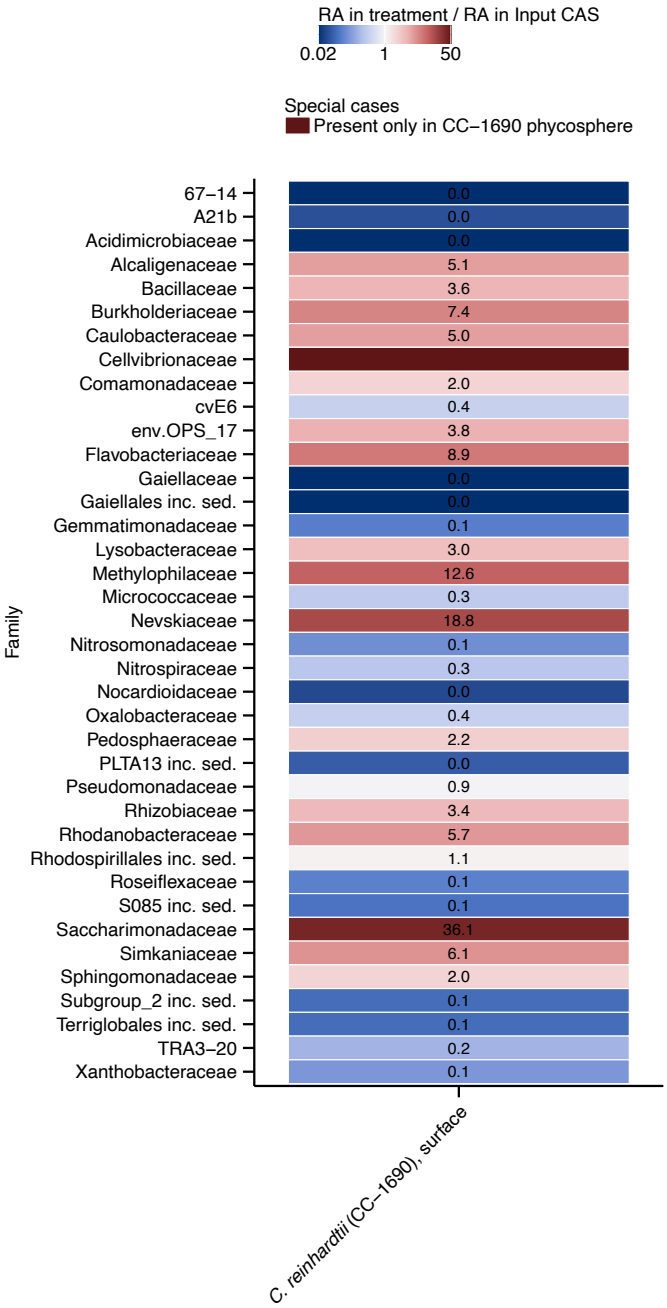
Generative AI models (ChatGPT by OpenAI and Claude by Anthropic) were used to assist in drafting code templates, refining analytical workflows, language refinement and improving textual explanations. The models did not independently generate results or draw conclusions. All AI-generated content was reviewed, verified, and edited by the author. All AI-generated content was reviewed, verified, and edited by the author. All analyses were designed, executed, and interpreted by the authors. The author retains full responsibility for the accuracy, originality, and integrity of the work.

2.6 Supplementary Figures

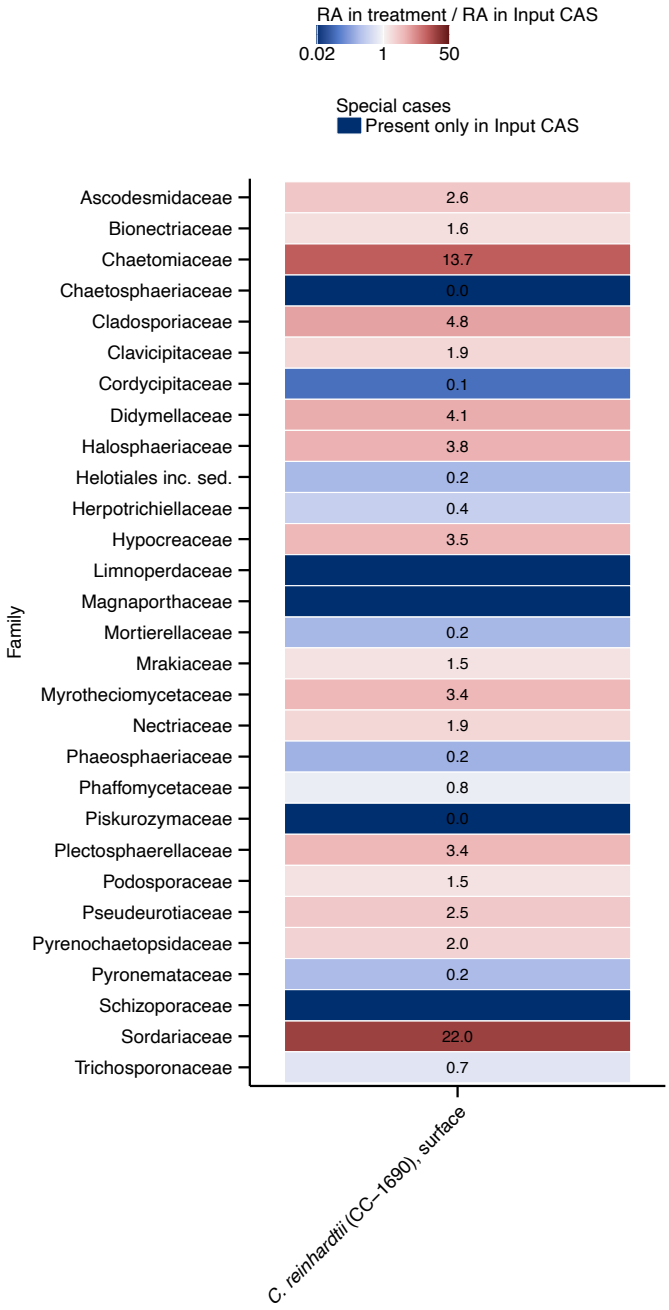
Supplementary Figure S2.1 Heatmap of bacterial family-level relative abundances in samples from the Eifel site (n = 15) and CAS (n = 5). The heatmap shows relative abundances of bacterial families (rows) across samples (columns). Only families with $\geqslant$ 1% mean relative abundance in at least one site are included. Columns are annotated by site (top). Color intensity indicates relative abundance, with white representing zero abundance (family not detected in the sample).



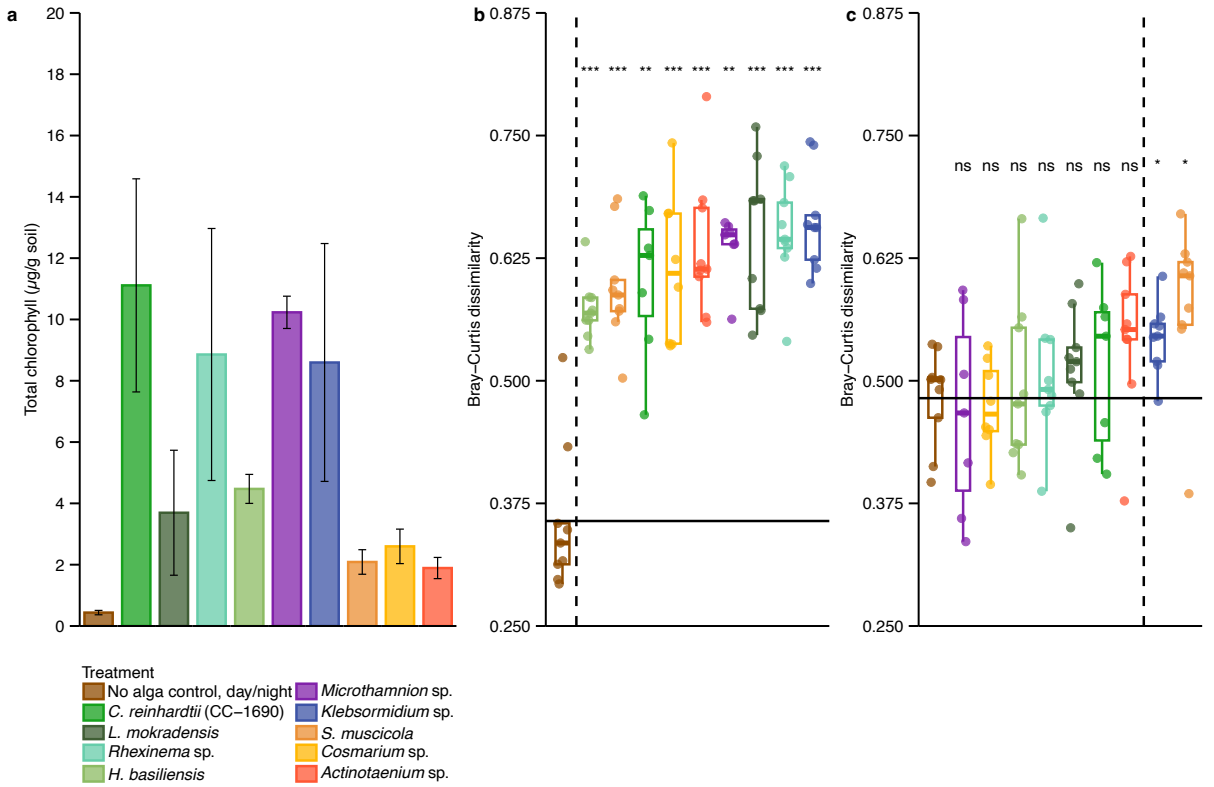
Supplementary Figure S2.2 Heatmap of fungal family-level relative abundances in samples from the Eifel site (n = 15) and CAS (n = 5). The heatmap shows relative abundances of fungal families (rows) across samples (columns). Only families with $\geqslant$ 1% mean relative abundance in at least one site are included. Columns are annotated by site (top). Color intensity indicates relative abundance, with white representing zero abundance (family not detected in the sample).



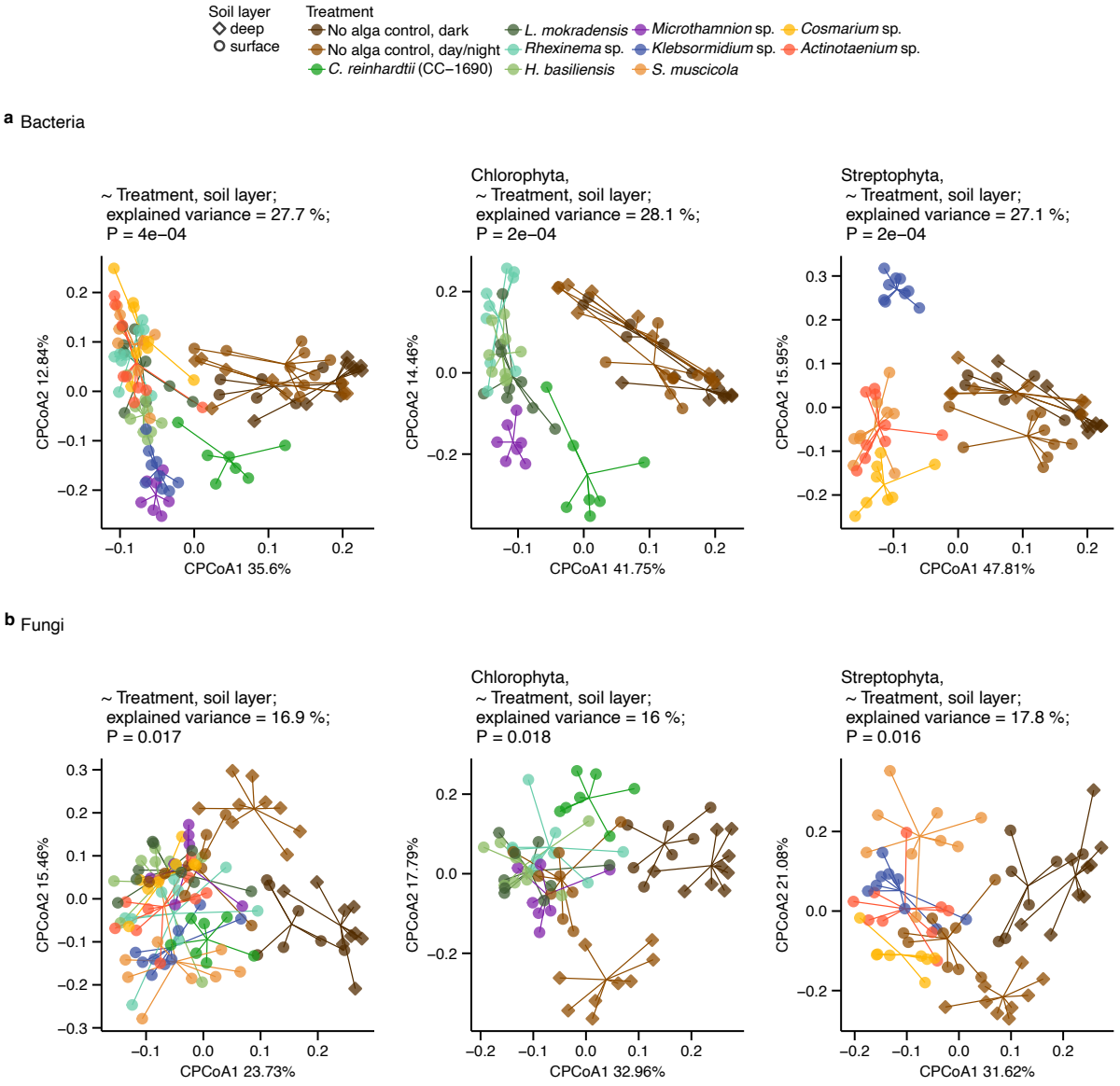
Supplementary Figure S2.3 Ratios of mean family-level bacterial RA between *C. reinhardtii* CC-1690 phycosphere and input CAS ($RA_{treatment}/RA_{input}$). Cells are colored according to the ratio: shades of blue indicate a decrease in family-level RA in the phycosphere, and shades of red indicate an increase. Numbers represent the ratio value; families detected only in the phycosphere are shown without a number. Only families with $\geqslant$ 1% mean relative abundance in either input CAS or *C. reinhardtii* CC-1690 phycosphere are included.



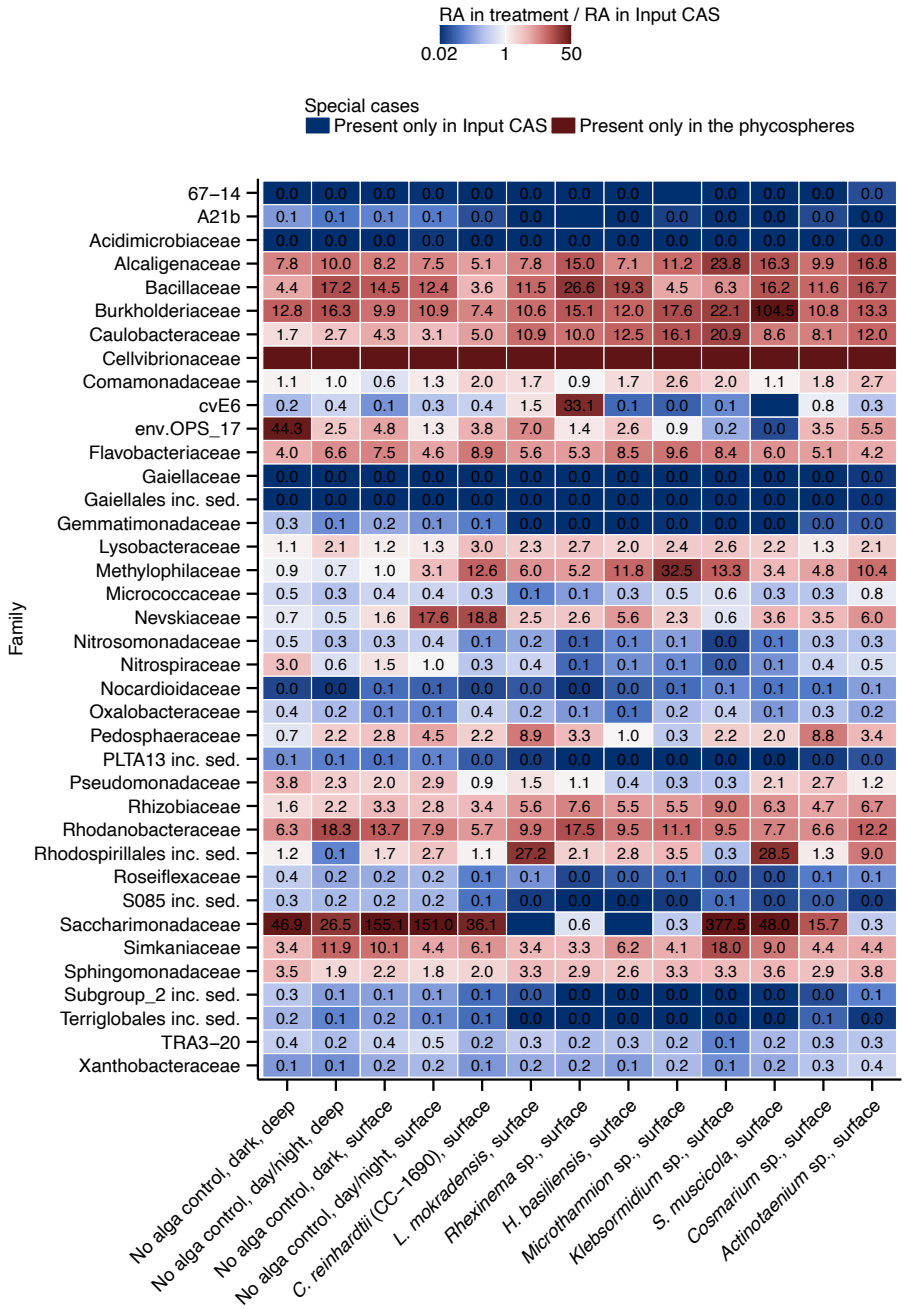
Supplementary Figure S2.4 Ratios of mean family-level fungal RA between *C. reinhardtii* CC-1690 phycosphere and input CAS ($RA_{treatment}/RA_{input}$). Cells are colored according to the ratio: shades of blue indicate a decrease in family-level RA in the phycosphere, and shades of red indicate an increase. Numbers represent the ratio value; families detected only in the input CAS are shown without a number. Only families with $\geqslant$ 1% mean relative abundance in either input CAS or *C. reinhardtii* CC-1690 phycosphere are included.



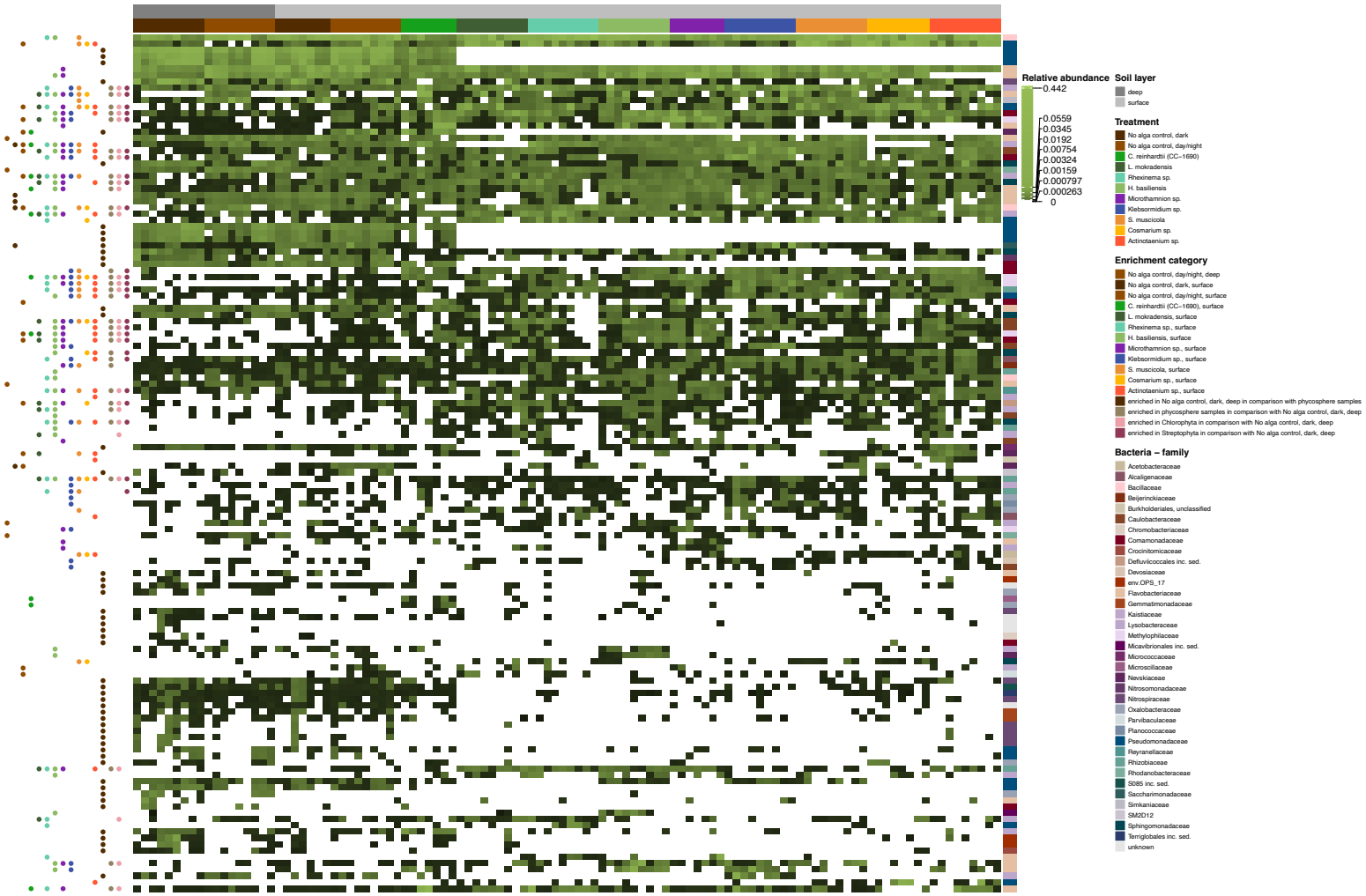
Supplementary Figure S2.5 All algae grew on soil and significantly changed soil bacterial community composition, whereas only some significantly influenced the fungal community composition. a – Total chlorophyll content in soil surface samples from pots inoculated with different algae and no alga control. Bars represent mean total chlorophyll per gram of soil (± standard error, n = 3 biological replicates, each with 3 technical replicates) for each treatment. Colors correspond to individual algal treatments. The no alga control incubated in the dark is excluded. b – ASV-level analysis of Bray–Curtis dissimilarities of bacterial communities from no alga control and phycosphere samples, compared to the averaged no alga control, day/night, surface soil sample. Significant differences are marked with stars (two-sided Wilcoxon test, with Benjamini-Hochberg correction; ns – not significant, $P > 0.05$, * – $0.05 \geq P > 0.01$, ** – $0.01 \geq P > 0.001$, *** – $0.001 \geq P > 0.0001$, **** – $P < 0.0001$), and samples are color-coded based on the treatment, with point shapes corresponding to the soil layer. The dashed vertical line separates the experimental conditions not significantly different from the averaged no alga control, day/night, surface soil from those significantly different, and the horizontal line shows the average distance of the individual no alga control, day/night, surface soil samples to the averaged one. n = 7 for no alga control surface soil and *C. reinhardtii* phycosphere, n = 8 for *Microthamnion* sp., otherwise n = 9. c – ASV-level analysis of Bray–Curtis dissimilarities of fungal communities from no alga control and phycosphere samples, compared to the averaged no alga control, day/night, surface soil sample. Significant differences are marked as in panel b.



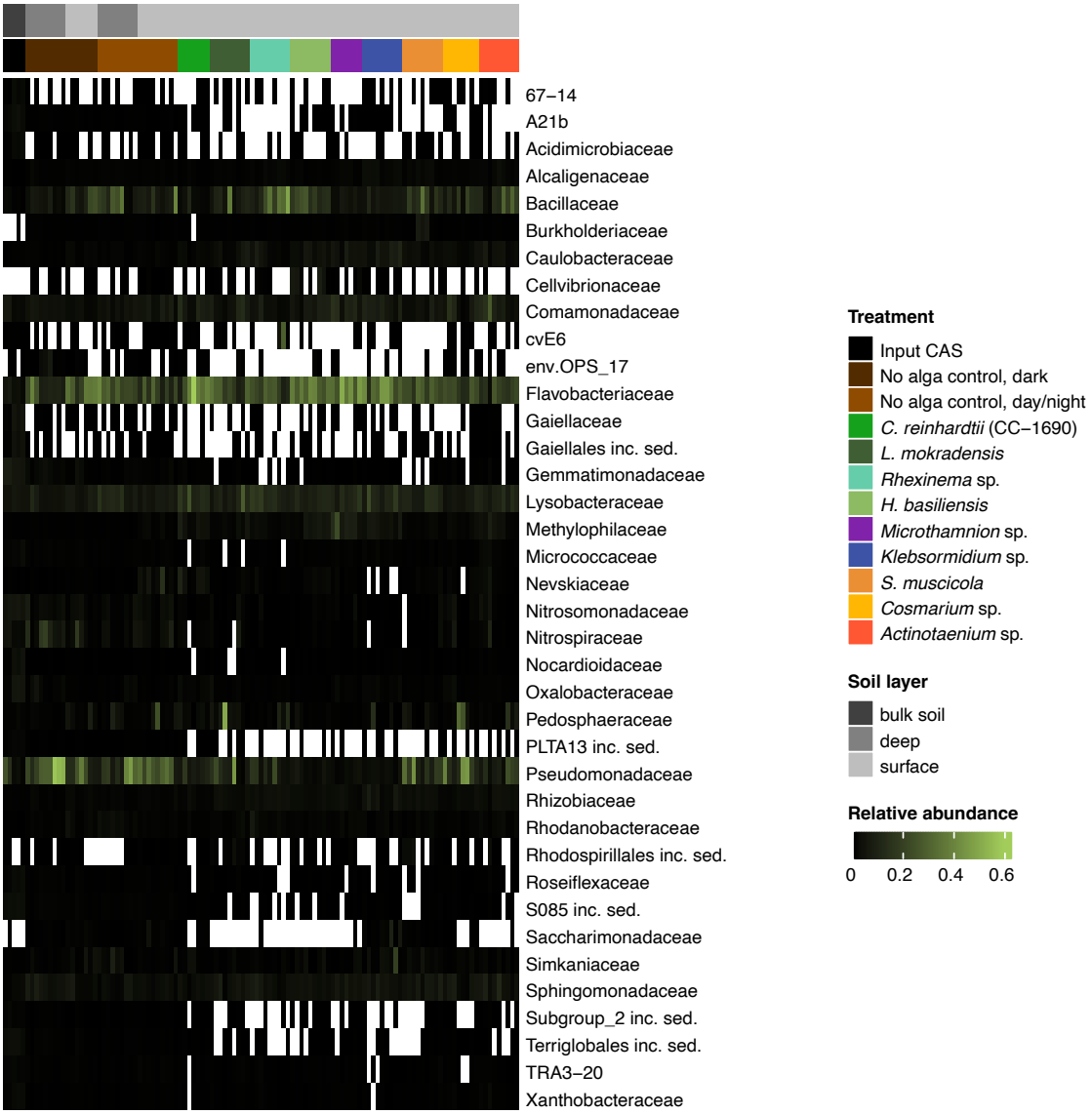
Supplementary Figure S2. 6 Phycosphere community structure depends on algal host taxonomy. a – ASV-level beta-diversity analysis (Principal Component Analysis of Bray–Curtis dissimilarities, constrained by treatment and soil layer combined as a single variable) of bacterial communities from the no alga control samples and phycosphere samples, including an analysis of all treatments (27.7% of the variance; $P < 0.001$; equivalent to Fig. S2.3) and separate analyses for no alga control and Chlorophyta (28.1% of the variance; $P < 0.001$) or Streptophyta (27.1% of the variance; $P < 0.001$). b – ASV-level beta-diversity analysis (Principal Component Analysis of Bray–Curtis dissimilarities, constrained by Treatment and soil layer combined as a single variable) of fungal communities from the no alga control samples and phycosphere samples, including an analysis of all treatments (16.9% of the variance; $P < 0.05$) and separate analyses for no alga control and Chlorophyta (16% of the variance; $P < 0.05$) or Streptophyta (17.8% of the variance; $P < 0.05$).



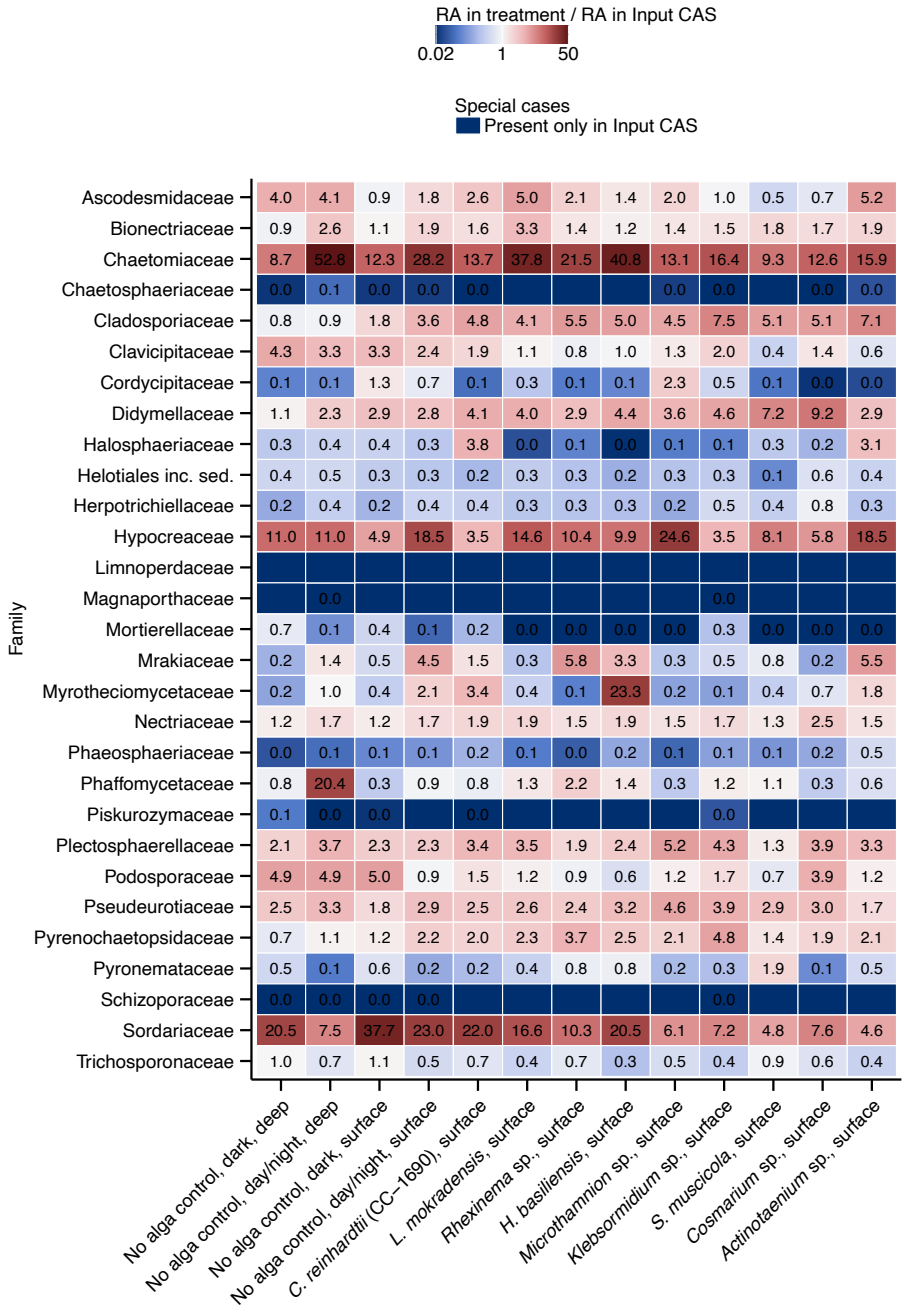
Supplementary Figure S2.7 Ratios of mean family-level bacterial RA between phycospheres or no alga controls and input CAS ($RA_{treatment}/RA_{input}$). Cells are colored according to the ratio: shades of blue indicate a decrease in family-level RA in the phycosphere, and shades of red indicate an increase. Numbers represent the ratio value; families detected only in the phycospheres or input CAS are shown without a number. Only families with $\geqslant$ 1% mean relative abundance in either input CAS or one of the treatments are included.



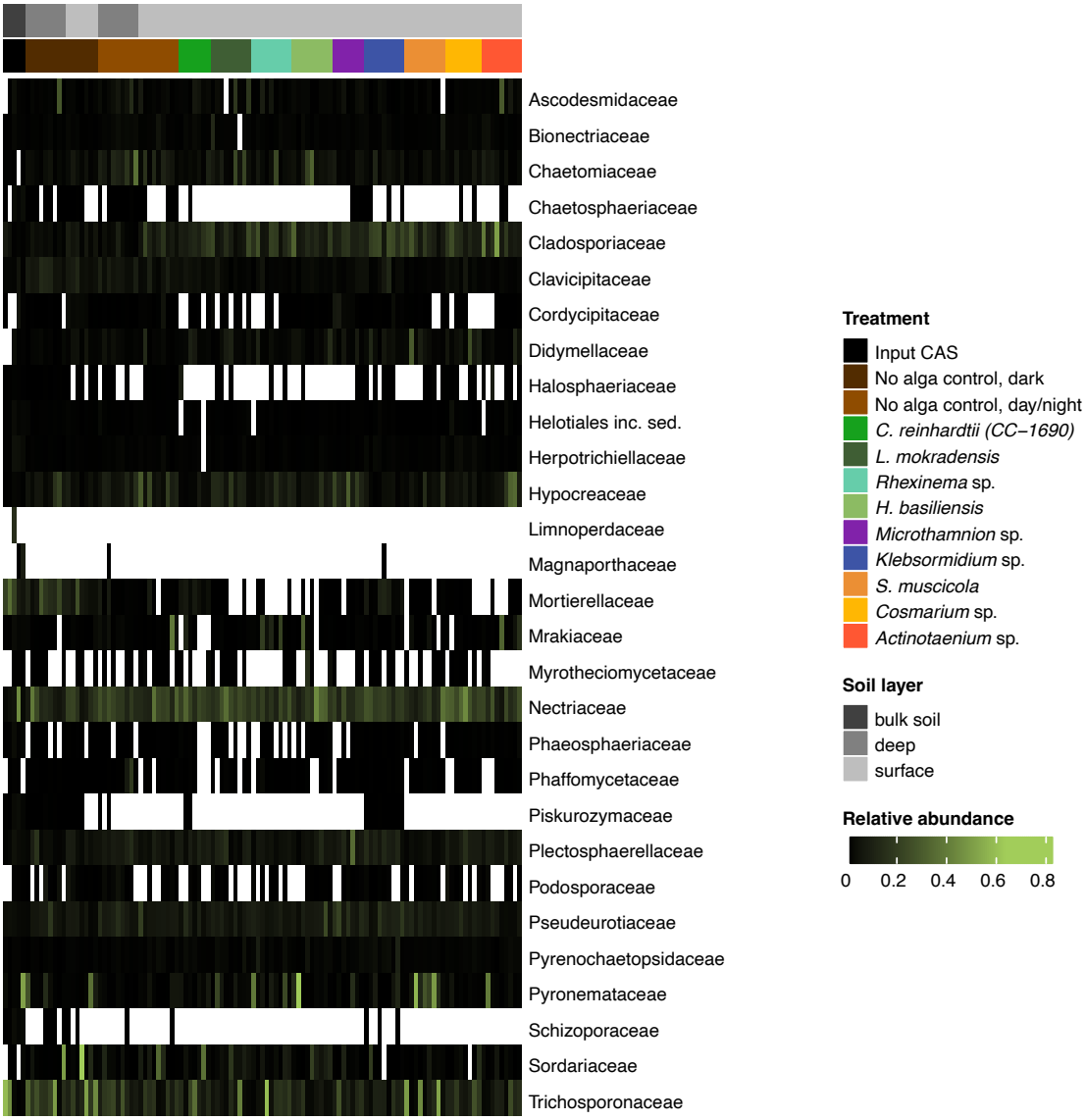
Supplementary Figure S2.8 Heatmap of relative abundances of ASVs enriched in no alga controls, all phycospheres, Streptophyta or Chlorophyta phycospheres, and single alga phycospheres. Left annotation indicates the enriched compartments, and right annotation indicates the bacterial family the ASV belongs to.



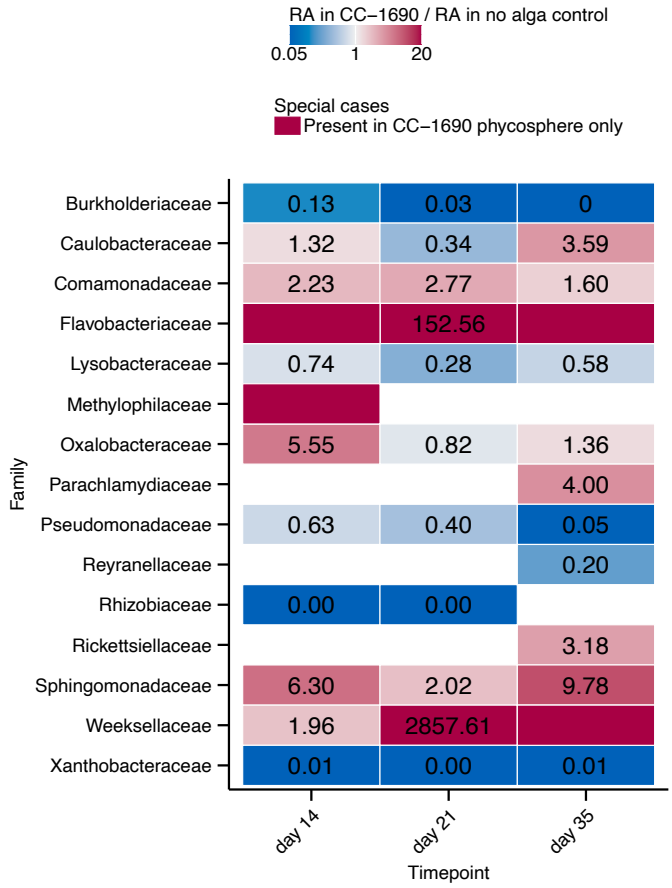
Supplementary Figure S2.9 Heatmap of bacterial family-level relative abundances in input CAS, no alga controls, and phycosphere samples. The heatmap shows relative abundances of bacterial families (rows) across samples (columns). Only families with $\geqslant$ 1% mean relative abundance in at least one condition are included. Columns are annotated by treatment (top). Color intensity indicates relative abundance, with white representing zero abundance (family not detected in the sample).



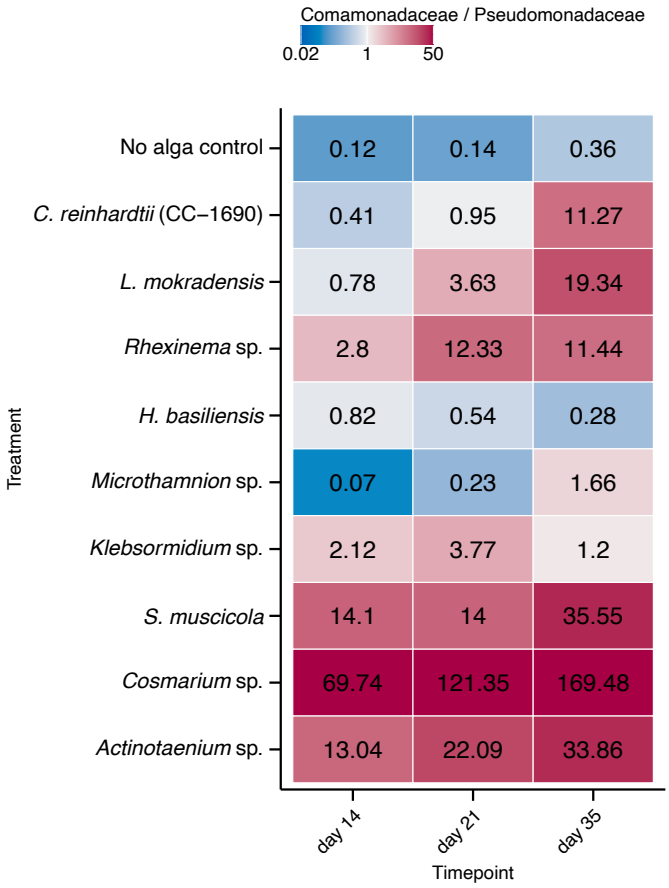
Supplementary Figure S2.10 Ratios of mean family-level fungal RA between phycospheres or no alga controls and input CAS ($RA_{treatment}/RA_{input}$). Cells are colored according to the ratio: shades of blue indicate a decrease in family-level RA in the phycosphere, and shades of red indicate an increase. Numbers represent the ratio value; families detected only in the phycospheres or input CAS are shown without a number. Only families with $\geqslant$ 1% mean relative abundance in either input CAS or one of the treatments are included.



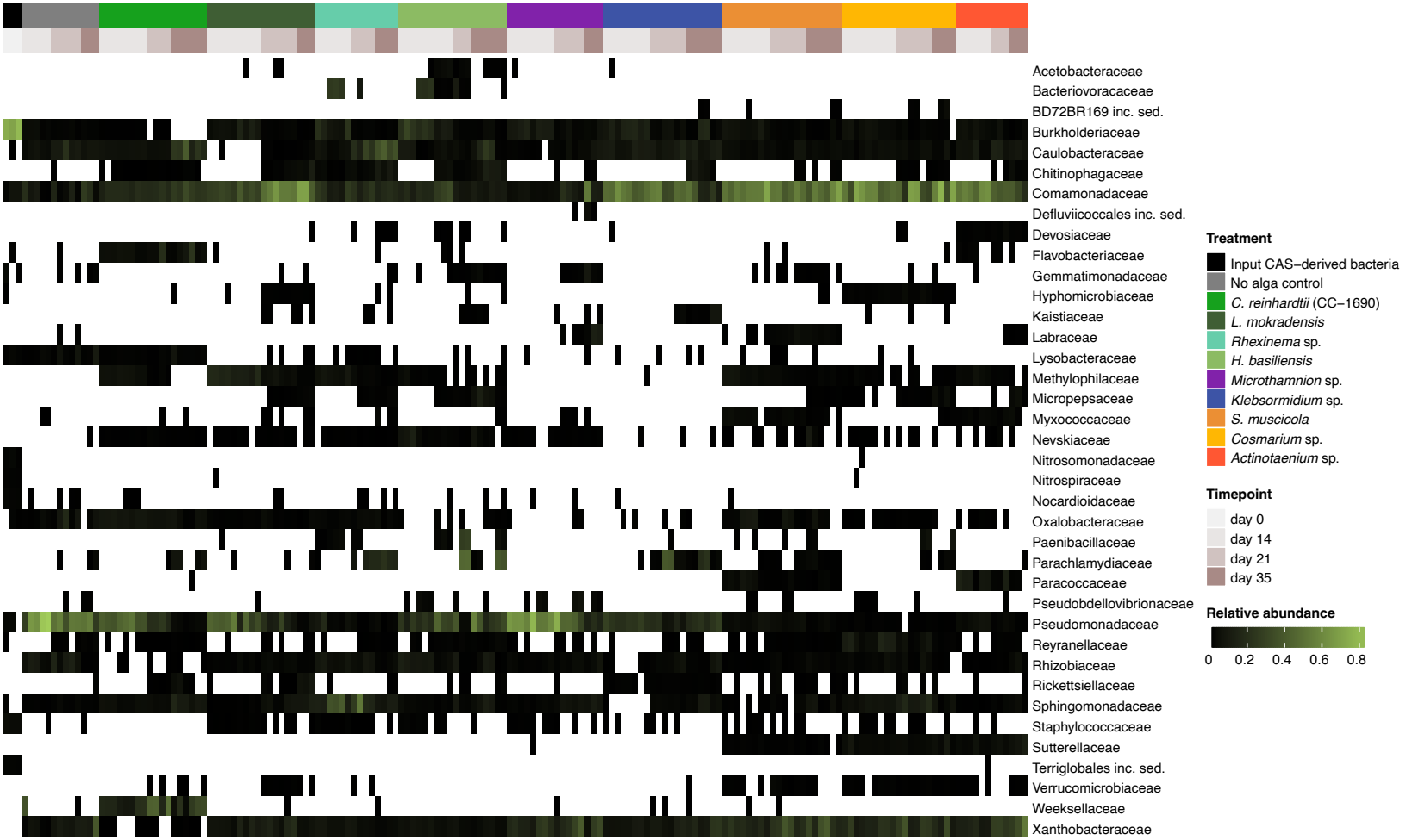
Supplementary Figure S2.11 Heatmap of fungal family-level relative abundances in input CAS, no alga controls, and phycosphere samples. The heatmap shows relative abundances of fungal families (rows) across samples (columns). Only families with $\geq$ 1% mean relative abundance in at least one condition are included. Columns are annotated by treatment (top). Color intensity indicates relative abundance, with white representing zero abundance (family not detected in the sample).



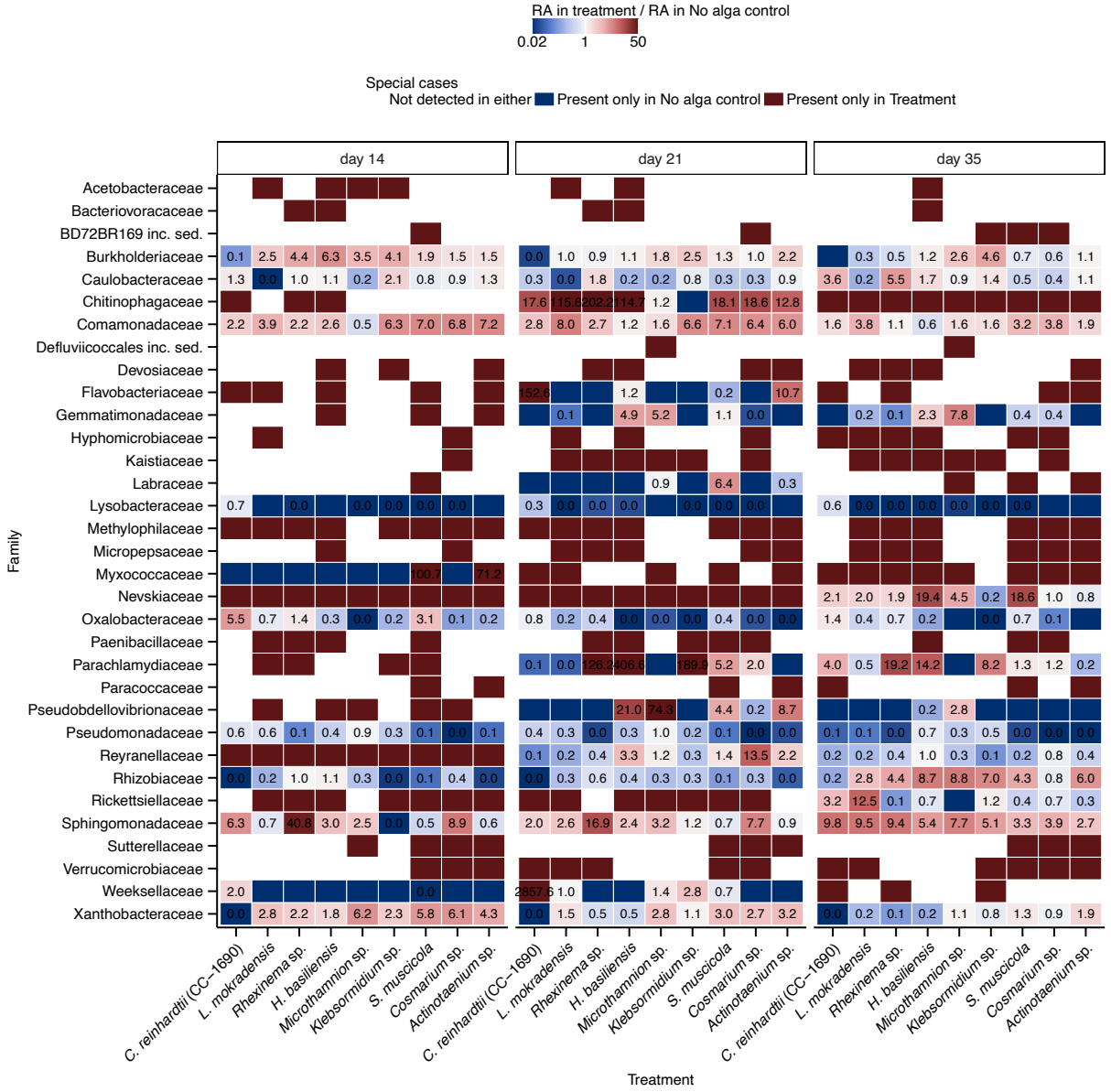
Supplementary Figure S2.12 Ratios of mean family-level bacterial RA between *C. reinhardtii* CC-1690 phycosphere and no alga control at different timepoints ($RA_{treatment}/RA_{control}$). Cells are colored according to the ratio: shades of blue indicate a decrease in family-level RA in the phycosphere, and shades of red indicate an increase. Numbers represent the ratio value; families detected only in the phycosphere are shown without a number. White cells indicate families absent from both the control and phycosphere at a given time point. Only families with $\geqslant$ 1% mean relative abundance in either control or *C. reinhardtii* CC-1690 phycosphere are included.



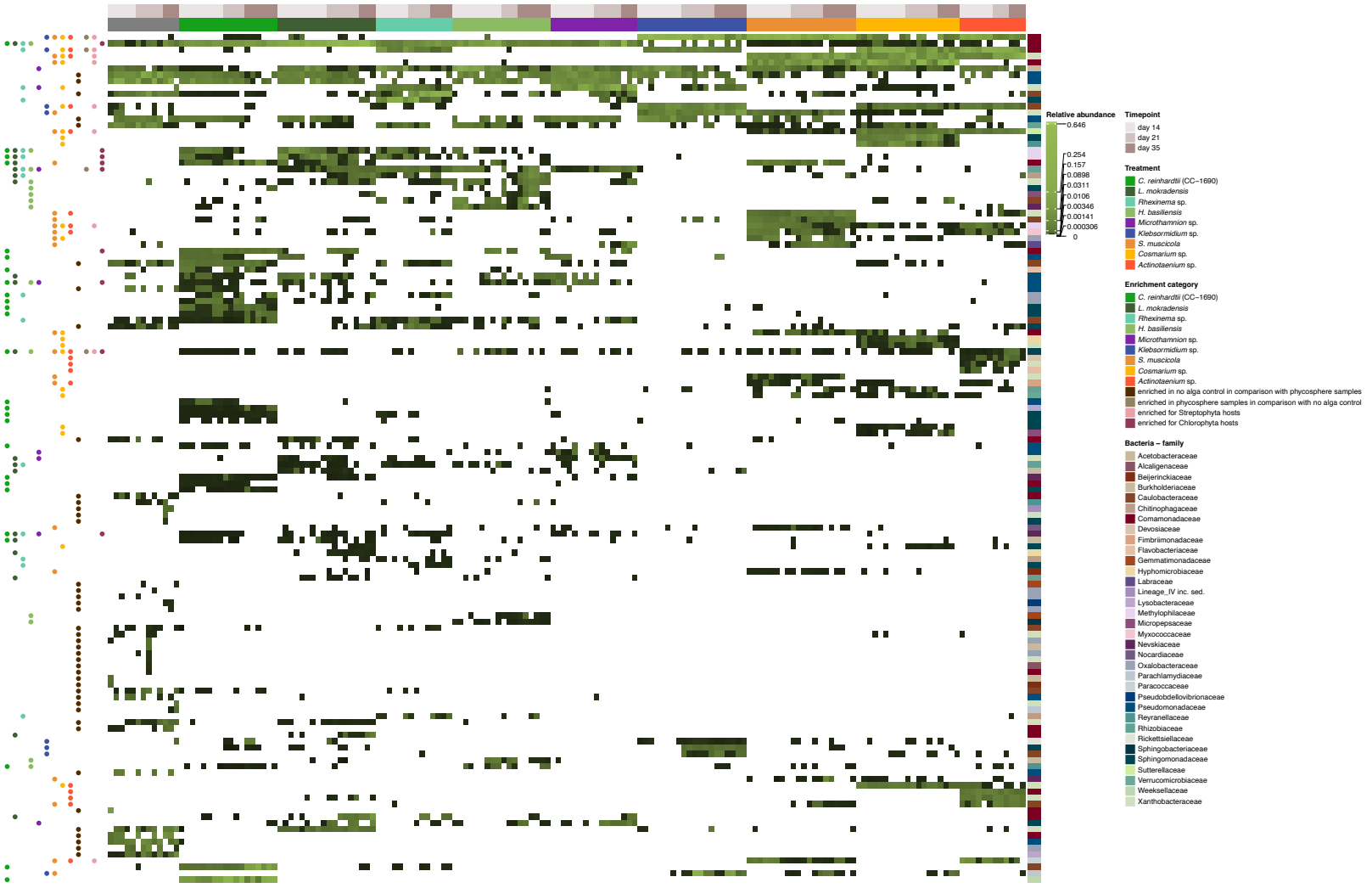
Supplementary Figure S2.13 Ratios of mean Comamonadaceae RA to Pseudomonadaceae RA for different treatments ($RA_{Comamonadaceae}/RA_{Pseudomonadaceae}$). Cells are colored according to the ratio: shades of blue indicate Pseudomonadaceae domination, and shades of red indicate Comamonadaceae domination. Numbers represent the ratio value.



Supplementary Figure S2.14 Heatmap of bacterial family-level relative abundances in input CAS-derived bacteria, no alga controls, and phycosphere samples. The heatmap shows relative abundances of bacterial families (rows) across samples (columns). Only families with $\geqslant$ 1% mean relative abundance in at least one condition are included. Columns are annotated by treatment (top). Color intensity indicates relative abundance, with white representing zero abundance (family not detected in the sample).



Supplementary Figure S2.15 Ratios of mean family-level bacterial RA between phycospheres and no alga control at a given timepoint ($RA_{treatment}/RA_{control}$). Cells are colored according to the ratio for a given family: shades of blue indicate a lower RA in the phycosphere, and shades of red indicate a higher RA in the phycosphere. Numbers represent the ratio value; families detected only in the phycospheres or controls are shown without a number. White cells indicate families absent from both the control and phycosphere at a given time point. Only families with $\geqslant$ 1% mean relative abundance in either control or one of the treatments are included.



Supplementary Figure S2.16 Heatmap showing relative abundances of ASVs enriched in no alga controls, all phycospheres, Streptophyta or Chlorophyta phycospheres, and single alga phycospheres. Left annotation indicates the enriched compartments, and right annotation – the bacterial family the ASV belongs to.

2.7 Supplementary Tables

Supplementary Table S2.1 Taxonomy of the subaerial algae used in the study.

AlgaID	Division	Class	Order	Family	Designation
CC-1690	Chlorophyta	Chlorophyceae	Chlamydomonadales	Chlamymonadaceae	*Chlamydomonas reinhardtii* CC-1690
MEL1030B	Chlorophyta	Chlorophyceae	Chlamydomonadales	Chlamymonadaceae	*Lercomonas mokradensis*
MEL1032B	Chlorophyta	Ulvophyceae	Ulotrichales	Helicodictyaceae	*Rhexinema* sp.
MEL1033B	Chlorophyta	Ulvophyceae	Ulotrichales	Hazeniaceae	*Hazenia basiliensis*
MEL1108B	Chlorophyta	Trebouxiophyceae	Microthamniales	Microthamniaceae	*Microthamnion* sp.
MEL1122B	Streptophyta	Klebsormidiophyceae	Klebsormidiales	Klebsormidiaceae	*Klebsormidium* sp.
MEL1126B	Streptophyta	Zygnematophyceae	Spirogloeales	Spirogloeaceae	*Spirogloea muscicola*
MEL1127B	Streptophyta	Zygnematophyceae	Desmidiales	Desmidiaceae	*Cosmarium* sp.
MEL1134B	Streptophyta	Zygnematophyceae	Desmidiales	Desmidiaceae	*Actinotaenium* sp.

Supplementary Table S2.2 Results of the Mantel tests.

Dataset	Community	Correlation	Mantel *r*	*P*	Permutations
Soil	Bacterial (16S)	Pearson	0.16	0.092	9999
Soil	Bacterial (16S)	Spearman	0.13	0.167	9999
Soil	Fungal (ITS)	Pearson	0.11	0.227	9999
Soil	Fungal (ITS)	Spearman	0.12	0.228	9999
Liquid medium (flask)	Bacterial (16S)	Pearson	0.84	0.0003	9999
Liquid medium (flask)	Bacterial (16S)	Spearman	0.66	0.008	9999

Supplementary Table S2.3 Primers used in the study.

Primer name	Sequence 5'-3'	Fragment	Purpose	Reference	PCR program
F-566	CAGCAGCCGCGGTAATTCC	18S_EukAB	sequencing	Hadziavdic *et al.*, 2014[186]	NA
R-1200	CCCGTGTTGAGTCAAATTAAGC	18S_EukAB	sequencing	Hadziavdic *et al.*, 2014[186]	NA
BP23-724R	TCGCATGTACCTGCAGTAGC	rbcL_BP	amplification, sequencing	Fay *et al.*, 1997[164]	phylo50
BP22-1F	ATGTCACCACAAACAGAAAC	rbcL_BP	amplification, sequencing	Fay *et al.*, 1997[164]	phylo50
EukA	AACCTGGTTGATCCTGCCAGT	18S_EukAB	amplification, sequencing	modified from Medlin *et al.*, 1988[162]	phylo58
EukB	TGATCCTTCTGCAGGTTCACCTAC	18S_EukAB, 18S_CrN1-F-EukB	amplification, sequencing	modified from Medlin *et al.*, 1988[162]	phylo58, phylo60
CrN1-F	CTGCCAGTAGTCATATGCTTGTCTC	18S_CrN1-F-EukB	amplification, sequencing	Marin *et al.*, 1998[163]	phylo60
RH1	ATGTCACCACAAACAGAAACTAAAGC	rbcL_RH1_1385R	amplification, sequencing	Manhart, 1994[165]	phylo50
1385R	GGAAAGAAATTAAATTTGAATT	rbcL_RH1_1385R	amplification, sequencing	McCourt et al., 2000[166]	phylo50
rbc571s	TGTTTACGAGGTGGTCTTGA	rbcL_RH1_1385R	sequencing	Hayden & Waaland, 2002[187]	NA
rbc590s	TCAAGACCACCTCGTAAACA	rbcL_RH1_1385R	sequencing	Hayden & Waaland, 2002[187]	NA

MaGo1F	ATGTCACCACAAACNGAAAC	rbcL_MaGo1F-3R	amplification	Gontcharov *et al.*, 2004[167]	phylo50
MaGo2F	ATGTCACCACAAACNGAAACTAAAGCWGG	rbcL_MaGo2F-3R	amplification	Gontcharov *et al.*, 2004[167]	phylo50nested2
MaGo3R	GTATCRATHGTWTCAAATTC	rbcL_MaGo1F-3R, rbcL_MaGo2F-3R	amplification	Gontcharov *et al.*, 2004[167]	phylo50, phylo50nested2
Seq-1F-rbcL	CACCACAAACNGAAACTAAAGCWGG	rbcL_MaGo2F-3R	sequencing	Gontcharov *et al.*, 2004[167]	NA
Seq-2F-rbcL	TRTTYACYTCYATTGTAGG	rbcL_MaGo2F-3R	sequencing	Gontcharov *et al.*, 2004[167]	NA
Seq-4R-rbcL	ATCRATHGTWTCAAATTC	rbcL_MaGo2F-3R	sequencing	Gontcharov *et al.*, 2004[167]	NA
Seq-5R-rbcL	TTWGGYTTAATDGTACAWCC	rbcL_MaGo2F-3R	sequencing	Gontcharov *et al.*, 2004[167]	NA
799F	AACMGGATTAGATACCCKG	16S rRNA, V5–V7	amplification	Chelius and Triplett, 2001[172]	16S_V5–V7_PCR1, 16S_V5–V7_PCR2
1192R	ACGTCATCCCCACCTTCC	16S rRNA, V5–V7	amplification	Beckers *et al.*, 2016[173]	16S_V5–V7_PCR1, 16S_V5–V7_PCR2
ITS_F1	GTGARTCATCGAATCTTTG	ITS2 rRNA	amplification	modified from Thiergart *et al.*, 2020[91]	ITS2_PCR1, ITS2_PCR2
ITS_R1	TCCTCCGCTTATTGATATGC	ITS2 rRNA	amplification	modified from Thiergart *et al.*, 2020[91]	ITS2_PCR1, ITS2_PCR2

Supplementary Table S2.4 PCR programs used in the study.

phylo58			
1	94 ˚C	2min	
2	94 ˚C	30s	x30
3	58˚C	30s	
4	72 ˚C	1.5 min	
5	72 ˚C	5min	
6	16 ˚C	forever hold	

phylo60			
1	94 ˚C	2min	
2	94 ˚C	30s	x30
3	60˚C	30s	
4	72 ˚C	1.5 min	
5	72 ˚C	5min	
6	16 ˚C	forever hold	

phylo50			
1	94 ˚C	2min	
2	94 ˚C	30s	x30
3	50˚C	30s	
4	72 ˚C	1.5 min	
5	72 ˚C	5min	
6	16 ˚C	forever hold	

phylo50nested2			
1	94 ˚C	2min	
2	94 ˚C	30s	x35
3	50˚C	30s	
4	72 ˚C	1.5 min	
5	72 ˚C	5min	
6	16 ˚C	forever hold	

16S_V5–V7_PCR1			
1	94 ˚C	2min	
2	94 ˚C	30s	x25
3	55˚C	30s	
4	72 ˚C	30 s	
5	72 ˚C	5min	
6	16 ˚C	forever hold	

16S_V5–V7_PCR2			
1	94 ˚C	2min	
2	94 ˚C	30s	x10
3	55 ˚C	30s	
4	72 ˚C	30 s	
5	72 ˚C	5min	
6	16 ˚C	forever hold	

ITS2_PCR1			
1	94 ˚C	2min	
2	94 ˚C	30s	x25
3	59 ˚C	30s	
4	72 ˚C	30 s	
5	72 ˚C	5min	
6	16 ˚C	forever hold	

ITS2_PCR2			
1	94 ˚C	2min	
2	94 ˚C	30s	x10
3	59 ˚C	30s	
4	72 ˚C	30 s	
5	72 ˚C	5min	
6	16 ˚C	forever hold	

Supplementary Table S2.5 Classification of subaerial algae strains used in this study based on initial morphological identification and Sanger sequencing of marker genes.

AlgaID	Initial classification (morphology based)	Molecular result	Current status
MEL1030B	*Chlamydomonas* sp.	18S_EukAB: *Lercomonas mokradensis*	*Lercomonas mokradensis*
MEL1032B	*Chlorosarcina* cf.	18S_EukAB: *Rhexinema* sp.	*Rhexinema* sp.
MEL1033B	*Pleurastrum* sp.	18S_EukAB: *Hazenia basiliensis*	*Hazenia basiliensis*
MEL1034B	*Stichococcus* sp.	18S_EukAB,: inconclusive, rbcL failed amplification; 18S_EukAB clustering with *Mesotaenium* sp.	Excluded, pending re-evaluation after additional data become available
MEL1108B	*Microthamnion* sp.	18S_EukAB: *Microthamnion* sp.	*Microthamnion* sp.
MEL1122B	*Klebsormidium* sp.	18S_EukAB: *Klebsormidium* sp.	*Klebsormidium* sp.
MEL1126B	*Spirogloea muscicola*	rbcL_RH1_1385R, rbcL_BP: *Spirogloea muscicola*	*Spirogloea muscicola*
MEL1127B	*Actinotaenium globosum*	18S_CrN1-F-EukB; rbcL, 2 step rbcL_MaGo1F-3R, rbcL_MaGo2F-3R: *Cosmarium* sp.	*Cosmarium* sp.
MEL1134B	*Cylindrocystis* sp.	18S_EukAB; rbcL, 2 step rbcL_MaGo1F-3R, rbcL_MaGo2F-3R: *Acinotaenium* sp.	*Actinotaenium* sp.

Supplementary Table S2.6 Composition of SFM culture medium[105] (pH 6.0 adjusted with HCl).

	Components	Final concentration	Stock solution	Addition per 1 L medium
1	HEPES	1 mM	238.10 g L^{-1} H_2O	1 mL
2	$Ca(NO_3)_2 \times 4\ H_2O$	0.21 mM	100.00 g L^{-1} H_2O	0.5 mL
3	$MgSO_4 \times 7\ H_2O$	0.203 mM	20.00 g L^{-1} H_2O	2.5 mL
4	$(NH_4)_2HPO_4$	87.8 µM	20.00 g L^{-1} H_2O	0.58 mL
5	$K_2HPO_4 \times 3\ H_2O$ +	13.2 µM	5.00 g L^{-1} H_2O	0.6 mL
	Na_2CO_3	0.19 mM	32.00 g L^{-1} H_2O	0.6 mL
6	$NaNO_3$	0.176 mM	50.00 g L^{-1} H_2O	0.3 mL
7	H_3BO_3	16 µM	1.00 g L^{-1} H_2O	1 mL
8	**Vitamin solution**			1 mL
	Vitamin B_{12}	0.15 nM	0.20 mg L^{-1} H_2O	
	Biotin	4.10 nM	1.00 mg L^{-1} H_2O	
	Thiamine-HCl	0.30 µM	100.00 mg L^{-1} H_2O	
	Niacinamide	0.80 µM	0.10 mg L^{-1} H_2O	
	pH 7.0			
9	**Trace metals**			1 mL
9	*Trace metal solution preparation*			
	$Na_2EDTA \times 2\ H_2O$		4.36 g L^{-1} H_2O	
	$FeCl_3 \times 6\ H_2O$		3.15 g L^{-1} H_2O	
	Dissolve in 1000 mL × dH_2O, then add 1 mL of Primary Trace Metals each (see below). Primary Trace Metals are stored frozen as 1 mL aliquots.			
9	**Primary Trace Metals**		1.00 g L^{-1} H_2O	
	CoCl2 × 6 H_2O		0.25 g L^{-1} H_2O	
	CuSO4 × 5 H_2O		18.00 g L^{-1} H_2O	
	MnCl2 × 4 H_2O		1.89 g L^{-1} H_2O	
	Na2MoO4 × 2 H_2O		0.27 g L^{-1} H_2O	
	NiSO4 × 6 H_2O		0.13 g L^{-1} H_2O	
	Na3VO4		0.184 g L^{-1} H_2O	
	ZnSO4 × 7 H_2O		2.20 g L^{-1} H_2O	

Supplementary Table S2.7 Chemical composition of selected buffers used in this study.

Buffer name	**Component**	**Concentration**	**Notes**
Lysis buffer	Tris, pH 8.0	100 mM	
	NaCl	100 mM	
	EDTA	10 mM	
	water	fill up to the final volume	
	SDS	1.50%	add just before use
	DTT	40 mM	
	Proteinase K	100 µg mL^{-1}	
NaCl-PEG solution	PEG 8000	18% w/v	
	NaCl	14.6% w/v	
	EDTA	1 mM	
	Tris, pH 8.0	10 mM	

3 Chapter 3: Transcriptomic Responses of *Chlamydomonas reinhardtii* to a Synthetic Phycosphere Community

3.1 Abstract

Like land plants, microscopic algae can recruit and associate with communities of heterotrophic, soil-borne bacteria. It is hypothesized that cues for phycosphere microbiota assembly include the provision of organic carbon compounds and the exchange of metabolites and/or molecular signals. However, the principles underlying the establishment of the phycosphere in soil algae remain poorly understood. Here, we provide the first characterization of the model chlorophyte *Chlamydomonas reinhardtii* transcriptional response to its phycosphere microbiota and identify processes potentially important for this interaction. We show that the phycosphere community triggers a coordinated reprogramming of *C. reinhardtii* gene expression, suggesting a shift from catabolism toward growth and biosynthesis-oriented metabolism. This change in gene expression is possibly supported by bacterial contributions (e.g., CO_2 and other metabolites) and involves upregulation of biosynthetic pathways, translation machinery, and genes associated with metabolic exchange. Central to the proposed interaction are changes in amino acid oxidation mediated by L-amino acid oxidase (LAO1), together with SynCom-dependent upregulation of ammonium transport and assimilation, agmatine iminohydrolase-mediated putrescine biosynthesis, and amino acid exchange mediated by a bidirectional amino acid transporter (BAT1). Parts of this response are conserved with evolutionarily distant land plants, suggesting a possible ancestral origin of these responses. The identified candidate genes provide a starting point for functional validation using available *C. reinhardtii* mutants. Elucidating how *C. reinhardtii* interacts with its phycosphere microbiota and identifying mechanisms conserved with land plants will improve our understanding of the evolutionary origins of interactions between photosynthetic organisms and bacteria, and may inform the design of SynComs for biotechnological applications.

3.2 Introduction

Microscopic algae associate and interact with complex bacterial communities similarly to land plants[104,105]. In both aquatic and terrestrial ecosystems, algae assemble distinct phycosphere microbiota, with soil-borne algae recruiting bacteria from soil[104,105]. This ability resembles the establishment of distinct root-associated microbial communities by embryophytes[105], suggesting that parts of the mechanism driving microbiota assembly are shared between green algae and land plants[50]. The rhizosphere and phycosphere, which is the region surrounding an algal cell, are analogous microenvironments, where the exudates change the physicochemical environment close to the root or algal cell, respectively, and can stimulate chemotaxis enabling microbial colonization[77,104]. During plant root colonization, metabolites exuded by plant roots act as signaling molecules mediating the root colonization, and provide a stable environment for the associated plant microbiome[77,99]. The components shaping root community include polyamines, such as putrescine[77,95], isoflavonoids[96], purines[97], coumarins[75], amino acids[95,98,99], organic acids[82], sugars and vitamins[77,99]. Similarly, in aquatic environments, various metabolites, such as organic carbon[109,110], vitamins[106,111], and other micronutrients[112] are exchanged, influencing algal growth and showing similarities to the processes in the plant rhizosphere[104]. Beneficial bacteria can mobilize nutrients[75,76], lessen the effects of abiotic stresses[73,80,81], and protect the plant from pathogens[78,79]. Their growth is supported by the secretion of compounds that act as attractants while at the same time restricting the growth of the non-beneficial bacteria[83,106,111,113,114]. Historically, algal-bacterial research has focused on pairwise interactions between individual bacterial species and algae[111,115–118]. However, recent findings showed that the model alga *C. reinhardtii* can recruit and associate with a community of heterotrophic, soil-borne bacteria in a manner similar to land plants[105], even though it does not possess a complex immune system[46,105], which has been shown before to play a role in the root-associated microbiota assembly in land plants[84]. It is assumed that the main driver of the assembly of soil-derived microbial communities associated with unicellular algae is the secretion of photoassimilates[82,83,119]. Nevertheless, recent findings suggest that the provision of organic carbon compounds cannot be the only cue for the microbiota assembly, and an exchange of metabolites and/or molecular signals is most likely necessary to establish the interaction and support the growth of the phycosphere microbiota[105]. However, the principles of phycosphere microbiota establishment in soil algae have not yet been

elucidated. In this study, we investigate the interaction between *C. reinhardtii* and a soil-derived synthetic community (SynCom) using transcriptomics to identify candidate genes involved in alga-bacteria interactions.

3.3 Results

3.3.1 *C. reinhardtii* assembles a phycosphere microbiota in the photobioreactor

To characterize how the interaction with the bacterial SynCom affects *Chlamydomonas reinhardtii* transcriptional responses, we grew *C. reinhardtii* axenically or in the co-culture with the bacterial SynCom[105] in a gnotobiotic photobioreactor. We collected samples for chlorophyll and cell density measurements, 16S rRNA amplicon sequencing, and RNA sequencing (RNA-Seq) at two different timepoints, in the early exponential growth phase (4 days post-inoculation, dpi) and in the stationary phase (7 dpi) of *C. reinhardtii* growth (Fig. 3.1).

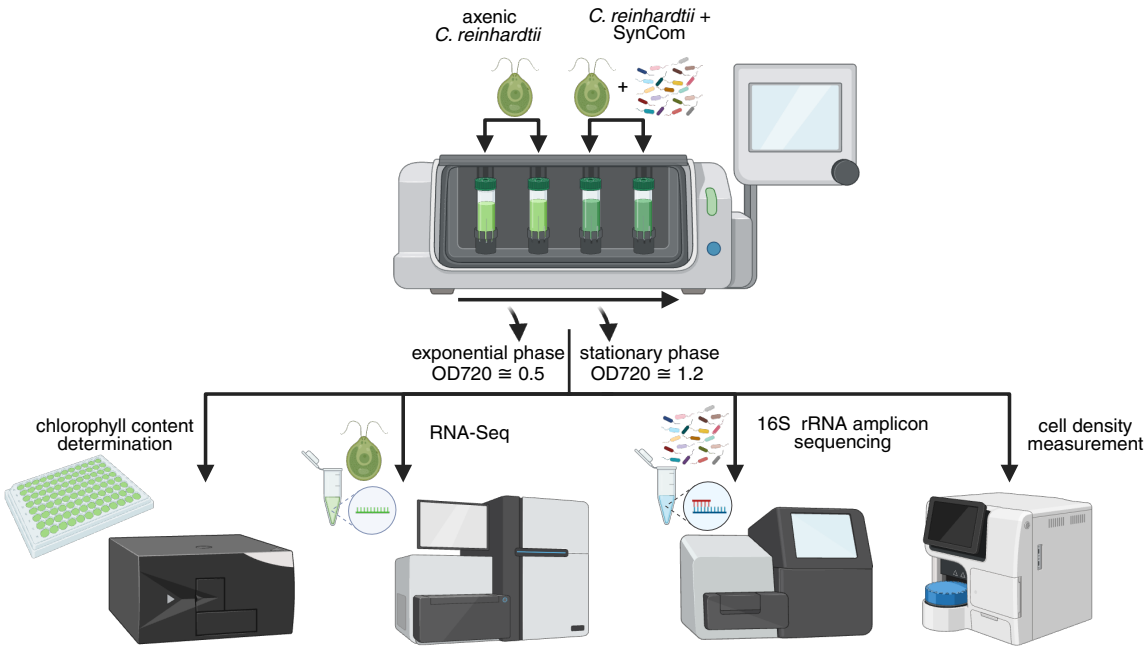


Figure 3.1 Experimental setup scheme. *C. reinhardtii* was cultivated axenically or in a co-culture with the SynCom in a gnotobiotic photobioreactor. Samples for RNA-Seq, 16S rRNA amplicon sequencing, chlorophyll content measurements, and cell density measurements were collected in the exponential and stationary phases of *C. reinhardtii* growth (determined based on the optical densities of the axenic samples).

At both sampling timepoints, we measured total chlorophyll content of the cultures and *C. reinhardtii* and bacterial cell densities (Fig. 3.2). While the total chlorophyll content was not significantly different between *C. reinhardtii* cultured axenically and in the co-culture with the SynCom (Fig. 3.2a), in the stationary phase of growth, we observed significantly higher *C. reinhardtii* cell densities in the co-culture than in the axenic culture (Fig. 3.2b), indicating the beneficial impact of the phycosphere community on *C. reinhardtii* growth. Additionally, in the co-cultures, we observed higher bacterial cell densities in the stationary phase than in the exponential phase, indicating the bacterial populations increased in the course of the experiment (Fig. 3.2c).

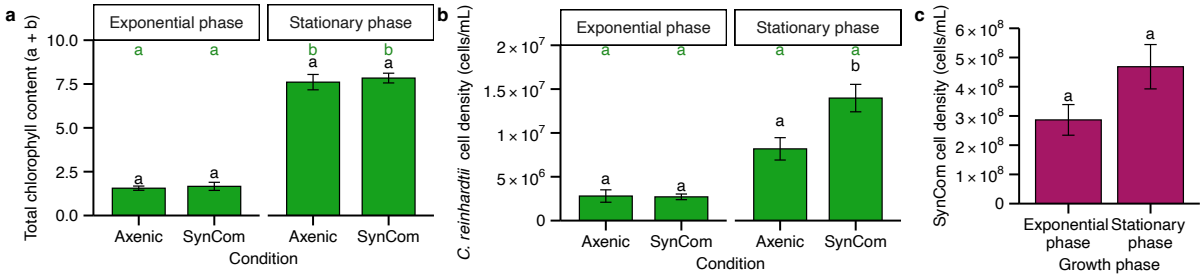


Figure 3.2 Comparison of *C. reinhardtii* growth in axenic culture and in the co-culture with the SynCom. a – Barplot showing the total chlorophyll content of the *C. reinhardti*i cultures. b – Barplot showing the *C. reinhardtii* cell densities. In panels a and b, letters above bars indicate statistically distinct groups ($P \leq 0.05$), with green letters denoting the comparison between Growth phases within the same Treatment (paired Wilcoxon test), and black letters denote the comparison between Treatments within each Growth phase (Mann-Whitney test). c – Bacterial cell densities in the co-cultures of *C. reinhardtii* and SynCom in different phases of *C. reinhardtii* growth. Letters denote the comparison between growth phases (paired Wilcoxon test). $n = 4$ per treatment and growth phase.

Similarly to what was shown before[105], we observed that *C. reinhardtii* assembled a stable community from the initial SynCom, with the community profile distinct from the input SynCom (Fig. 3.3). The samples from the co-cultures separated from the input SynCom along axis 1 and according to growth phase (timepoint) along axis 2, with samples taken in the stationary phase exhibiting higher within-group variance than the samples taken in the exponential phase (Fig. 3.3a). The most abundant SynCom members were Pseudomonas (Pseudomonadaceae), Cupriavidius (Burkholderiaceae) and Chryseobacterium (Flavobacteriaceae) in the exponential phase of growth, with Sphingopyxis (Sphingomonodaceae) increasing in abundance in the stationary phase of growth of *C. reinhardtii* (Fig. 3.3b).

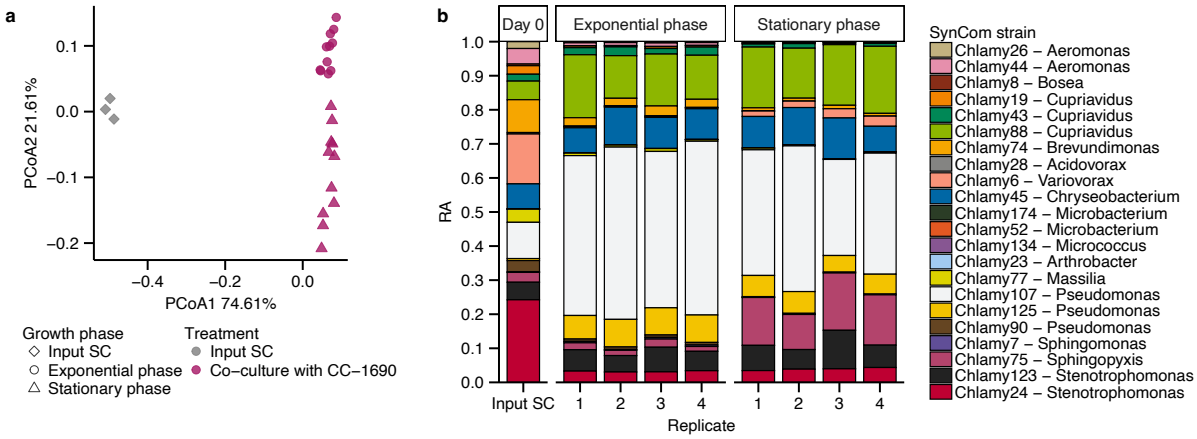


Figure 3.3 Analysis of the SynCom community composition over time. a – Principal Component Analysis of Bray–Curtis dissimilarities (ASV-level). b – Barplot showing the relative abundances of individual SynCom members in the input SynCom and in the exponential and stationary phases of *C. reinhardtii* growth, normalized by the copy number of the 16S rRNA gene of each SynCom member. n = 4 per treatment and growth phase.

3.3.2 *C. reinhardtii* responds transcriptionally to its phycosphere microbiota

To investigate if *C. reinhardtii* responds transcriptionally to its phycosphere microbiota, we compared the transcriptional outputs of *C. reinhardtii* cultured axenically and co-cultured with the SynCom. Since *C. reinhardtii* assembled a stable community from the initial bacterial SynCom, we expected to see separation based on the SynCom presence. Analysis of the RNA-Seq showed that transcriptional outputs separated according to *C. reinhardtii* growth phase along axis 1 and according to SynCom treatment along axis 2, with greater separation between the samples from the axenic and co-cultures with the SynCom in the stationary phase of *C. reinhardtii* growth than in the exponential phase (Fig. 3.4a). We identified 2,621 differentially expressed genes (DEGs) differentially regulated in response to time and to SynCom presence in total. A complete list of DEGs, including gene identifiers, $\log_2$(fold changes) ($\log_2$FC), adjusted *P*-values, and annotations, is provided in Supplementary Table S3.1. Although most SynCom-responsive DEGs were differentially expressed in the stationary phase, some were consistently differentially expressed in both phases of *C. reinhardtii* growth (Fig. 3.4b–c).

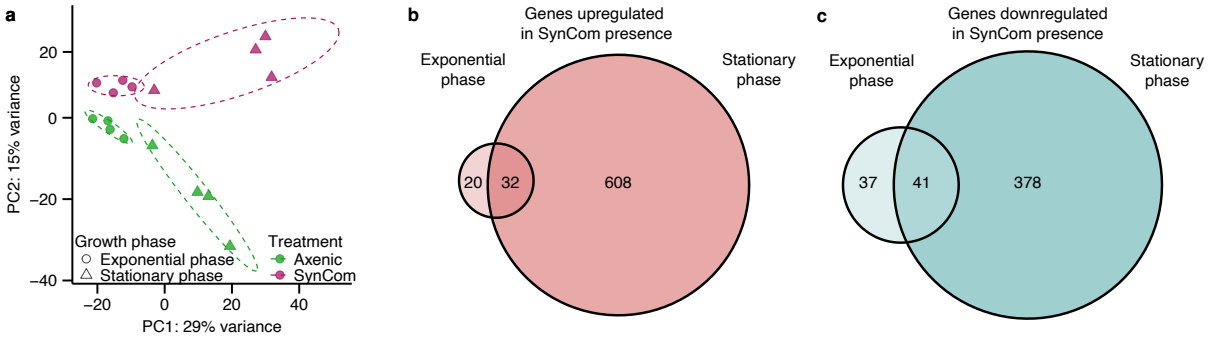


Figure 3.4 SynCom and time-specific transcriptional response of *C. reinhardtii*. a – Whole transcriptome level principal component analysis of the transcriptome (n = 4 per treatment and growth phase), with ellipses representing the 68% confidence interval of a multivariate normal distribution, approximating one standard deviation around the group centroid b, c – Venn diagrams showing the overlap between the DEGs upregulated (b) and downregulated (c) in the exponential and in the stationary phase of *C. reinhardtii* growth. The circle size is proportional to the number of genes in each group.

Further analysis of *k*-means clustering of the DEGs revealed gene clusters associated with general response to the SynCom, independently of the *C. reinhardtii* growth phase, as well as clusters specific both to growth phase and SynCom presence, and clusters specific to the *C. reinhardtii* growth phase only (Fig. 3.5).

3.3.3 Gene Ontology analysis reveals coordinated regulation of metabolism and gene expression

To characterize the functions of the DEGs, we first performed enrichment analysis of Gene Ontology (GO) Biological Process (BP) and Molecular Function (MF) terms. Since the Phytozome annotation provided GO annotations only for 48.5% of the expressed genes, we decided to expand the GO annotation to facilitate the exploratory analysis of the DEGs' functions. We used Gene Ontology Meta Annotator for Plants (GOMAP), a high-throughput, reproducible pipeline that enables genome-scale GO annotation of plant genomes[188], and annotated almost all *C. reinhardtii* expressed genes (Supplementary Figure S3.1). After increasing the coverage of the GO annotation, we performed enrichment analysis of GO BP and MF terms, followed by s semantic similarity-based reduction of GO terms to summarize related processes and improve interpretability. To get an overview of the *C. reinhardtii* biological processes influenced by the SynCom presence, we started with an analysis of the GO BP terms enriched in the *k*-means clusters (Fig. 3.5, Supplementary Figure S3.2,

Supplementary Table S3.2), along with the analysis of the fewer GO MF terms (Supplementary Figure S3.3, Supplementary Table S3.3).

Figure 3.5 Heatmap of genes differentially regulated in response to SynCom presence and over time. Heatmap showing scaled counts of genes arranged according to *k*-means clustering results (only DEGs shown, $|\log_2 FC| > 0$, Benjamini-Hochberg adjusted $P \leq 0.05$), with Gene Ontology Biological Processes enriched in clusters 1 – 4 shown next to the heatmap. Below the heatmap, cell densities of *C. reinhardtii* and SynCom bacteria in each culture at the sampling timepoints are shown. $n = 4$ per treatment and growth phase.

In cluster 1, which includes genes consistently downregulated in the SynCom presence regardless of the growth phase, enriched GO BP terms included terms related to amino acid catabolism, carbohydrate metabolic processes, the glyoxylate cycle and regulation of peroxisome size (Fig. 3.5), whereas among the GO MF terms, we found “hydrolase

activity, hydrolyzing O-glycosyl compounds" and "3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring) activity" (Supplementary Figure S3.3). While the first term is connected to carbohydrate metabolism, the second relates to the branched-chain amino acid catabolic process, as 3-methyl-2-oxobutanoate dehydrogenase converts the valine-derived α-keto acids into acyl-CoA intermediates. Cluster 2, which includes genes upregulated in both phases of *C. reinhardtii* growth, was enriched for GO BP terms related to negative regulation of hydrolase activity and polysaccharide catabolic process. Consistent with the enrichment of the "proteolysis" GO BP term among the DEGs in cluster 1, downregulated in the SynCom presence, in cluster 2, we detected enrichment of the "negative regulation of proteolysis" term (Fig. 3.5), and "endopeptidase regulator activity" among the enriched GO MF terms (Supplementary Figure S3.3). Collectively, these results suggest that *C. reinhardtii* transcriptionally responds to SynCom presence by largely regulating its amino acid, carbohydrate, and protein metabolism. The regulation of amino acid metabolism in the SynCom presence was also evident in the stationary phase, as seen in clusters 3 and 4, both containing genes upregulated in the co-cultures with the SynCom in the stationary phase. In these clusters, we detected enrichment of GO BP terms related to L-amino acid and proteinogenic amino acid biosynthetic and metabolic processes, indicating that apart from downregulating amino acid catabolism, *C. reinhardtii* also upregulates genes annotated to amino acid biosynthesis (Fig. 3.5, Supplementary Figure S3.2). Among the GO MF terms enriched in cluster 4, we observed enrichment of the "valine-tRNA ligase activity" term (Supplementary Figure S3.3), further indicating that, in the presence of the SynCom, amino acids, including branched-chain amino acids (BCAA) such as valine, are channeled into protein synthesis rather than being catabolized for energy.

In addition to the metabolic processes-related terms, supporting the hypothesis of changes in *C. reinhardtii* metabolism in response to the SynCom, we also identified a substantial number of GO BP and MF terms related to gene expression, transcription, and translation (Fig. 3.5, Supplementary Figure S3.2, S3.3). Moreover, we identified several terms related to the response to stimulus to be enriched in clusters of SynCom responsive genes (clusters 1, 3, 4, and 6, which include DEGs responding more strongly to SynCom presence in two out of the four replicates) (Fig. 3.5, Supplementary Figure S3.2). These terms likely reflect environment sensing by *C. reinhardtii*, which

induces transcriptional reprogramming in response to SynCom bacteria. Several GO BP terms enriched in clusters 3 and 4, such as ribonucleotide metabolic process and ribonucleoprotein complex biogenesis, are related to ribosome biogenesis, which is an extremely energy-costly process for the cells, and thereby needs to be precisely controlled[189,190]. Enrichment of these terms indicates that *C. reinhardtii* may be investing energy into growth and cell division, which aligns with the observation of higher *C. reinhardtii* cell densities in the stationary phase in the co-cultures than in the axenic cultures (Fig. 3.2b). Combined with the downregulation of amino acid catabolic processes in both phases of *C. reinhardtii* growth in cluster 1, the upregulation of amino acid synthesis in the stationary phase of *C. reinhardtii* growth suggests resource availability in the co-cultures, allowing the use of nitrogen for cell growth, as cell growth requires a constant supply of *de novo* biosynthesized amino acids[191].

Apart from the analysis of GO terms enrichment in the *k*-means clustered DEGs, we also specifically compared GO terms enrichment among DEGs either up- or downregulated between the two timepoints (stationary vs exponential growth phase) and the treatments (co-culture with the SynCom vs axenic cultures in both growth phases of *C. reinhardtii*). Although we performed the functional enrichment also on the DEGs exhibiting differential timepoint response while comparing the timepoint response in the SynCom cultures to the timepoint response in the axenic cultures, for this subset of DEGs, we did not identify any specifically enriched terms, which might be due to the low number of DEGs in the interaction contrast results (only 37 DEGs).

Consistent with the higher number of SynCom-responsive DEGs in the stationary phase (Fig. 3.4b–c), we also observed a greater number of enriched GO BP terms in the stationary phase (Fig. 3.6, Supplementary Table S3.4). Similarly to what we detected in the analysis of the GO terms enrichment in the *k*-means clustered DEGs, the GO BP terms enriched among the upregulated genes in the stationary phase included mainly those related to gene expression and biosynthetic processes, mainly those also related to gene expression, such as “nucleic acid biosynthetic process” and “organophosphate biosynthetic process” (Fig. 3.6).

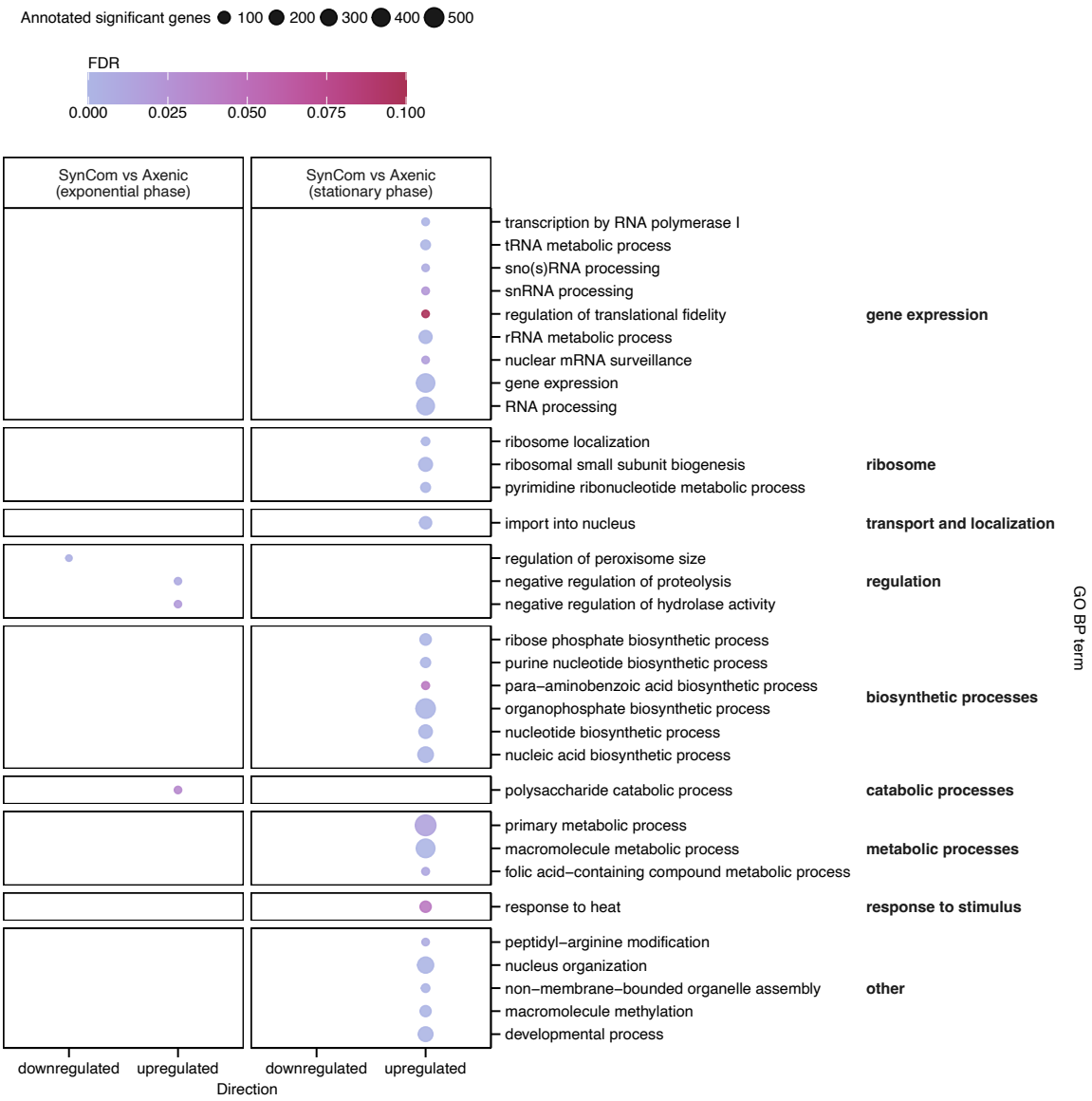


Figure 3.6 Gene Ontology Biological Process enrichment for SynCom vs axenic comparisons. Bubble plot showing GO BP terms enriched among significantly upregulated and downregulated genes in *C. reinhardtii* grown in the co-culture with the SynCom relative to axenic controls in exponential and stationary growth phases. Terms were semantically clustered and reduced, and representative parent terms are displayed. Bubble size denotes the number of significant genes annotated to each term, and color indicates adjusted *P*-values (FDR). Facets separate the SynCom-related comparisons: SynCom vs axenic (exponential phase), SynCom vs axenic (stationary phase), and the SynCom timepoint effect (stationary vs exponential).

Analogous pattern was also evident while analyzing GO BP and MF terms enriched among DEGs differentially regulated between the stationary and exponential phase separately in the axenic and the co-cultures (Supplementary Figure S3.4, S3.5, Supplementary Table S3.4, S3.5). While the terms related to gene expression and amino acid-related processes were significantly enriched among the upregulated DEGs in the co-cultures with the SynCom, they were not enriched in the axenic cultures,

further indicating that SynCom presence causes major transcriptional changes in *C. reinhardtii,* especially in the stationary phase of growth. Consistent with the GO BP terms enriched in SynCom-responsive clusters 1 and 2 (Fig. 3.5), the GO BP term "regulation of peroxisome size" was significantly enriched among the DEGs downregulated in the exponential phase in response to the SynCom, whereas terms "negative regulation of proteolysis", "negative regulation of hydrolase activity", and "polysaccharide catabolic process" were significantly enriched among the DEGs upregulated in the exponential phase in response to the SynCom (Fig. 3.6). A similar pattern emerged in the analysis of GO MF terms, with several gene-expression-related GO MF terms upregulated in the stationary phase in SynCom presence, and enrichment of "endopeptidase regulator activity" in the exponential phase (Fig. 3.7).

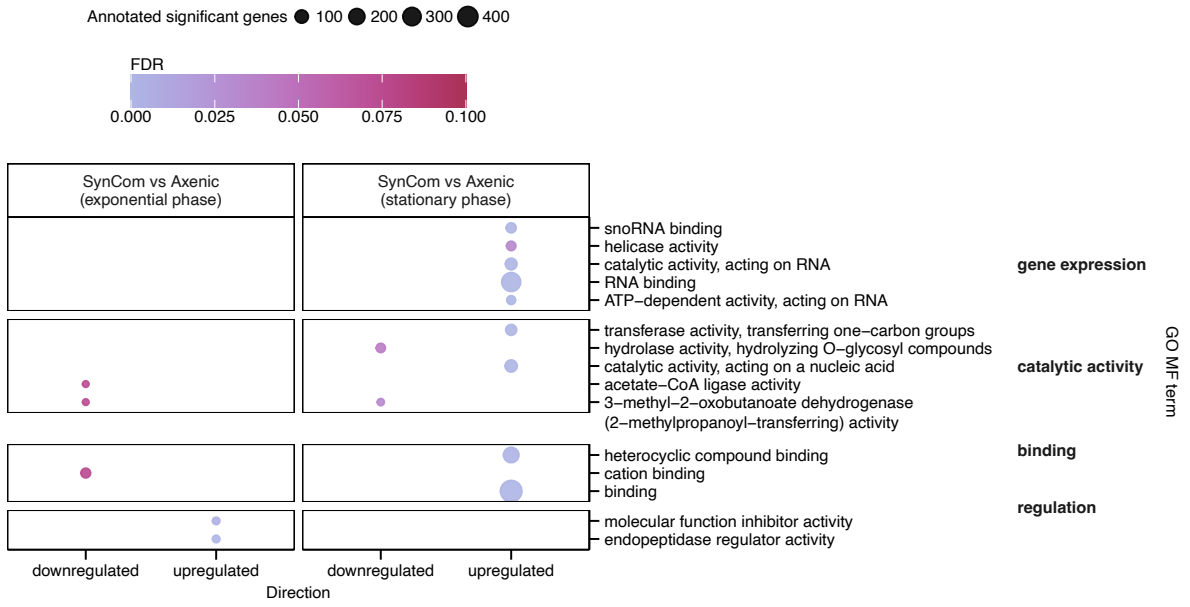


Figure 3.7 Gene Ontology Molecular Function enrichment for *C. reinhardtii* grown in the co-culture with the SynCom relative to axenic controls. Bubble plot showing GO MF terms enriched among significantly upregulated and downregulated genes in *C. reinhardtii* grown in the co-culture with SynCom relative to axenic controls in exponential and stationary growth phases. Terms were semantically clustered and reduced, and representative parent terms are displayed. Bubble size denotes the number of significant genes annotated to each term, and color indicates adjusted *P*-values (FDR). Facets separate the SynCom-related comparisons: SynCom vs axenic (exponential phase), SynCom vs axenic (stationary phase), and the SynCom timepoint effect (stationary vs exponential).

The term "3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring) activity", enriched in the cluster of genes downregulated in response to the SynCom in both phases of growth, was also enriched among the DEGs downregulated in the specific comparison of the *C. reinhardtii* co-culture with the

SynCom vs axenic cultures in both phases of *C. reinhardtii* growth (Fig. 3.5, 3.7, Supplementary Figure S3.3). Interestingly, the terms enriched in the DEGs downregulated in the exponential phase in the SynCom presence also included "acetate-CoA ligase activity" (Fig. 3.7). Acetate-CoA ligase activity is a key biochemical process, in which the acetate, Coenzyme A, and ATP are converted into acetyl-CoA, which is essential both for energy production in the tricarboxylic acid (TCA) cycle, as well as for biosynthesis pathways. The downregulation of these two enzymes further suggests that in the SynCom presence, *C. reinhardtii* downregulates alternative to photosynthesis pathways for energy production.

3.3.4 KEGG pathway analysis supports a shift from catabolism to photosynthesis-based metabolism

We further characterized *C. reinhardtii* transcriptional response to SynCom presence with the Kyoto Encyclopedia of Genes and Genomes (KEGG)[192]. First, because only 13% of the expressed genes were annotated with KEGG terms in the Phytozome annotation[193], we extended the Phytozome KEGG annotation using various KEGG annotation tools[194–196], which ultimately resulted in nearly 40% of the genes being annotated (Supplementary Figure S3.6). After expanding the annotation, we conducted a Gene Set Enrichment Analysis (GSEA) to identify functional modules and pathways enriched among the DEGs in each comparison (Fig. 3.8, Supplementary Table S3.7, S3.8).

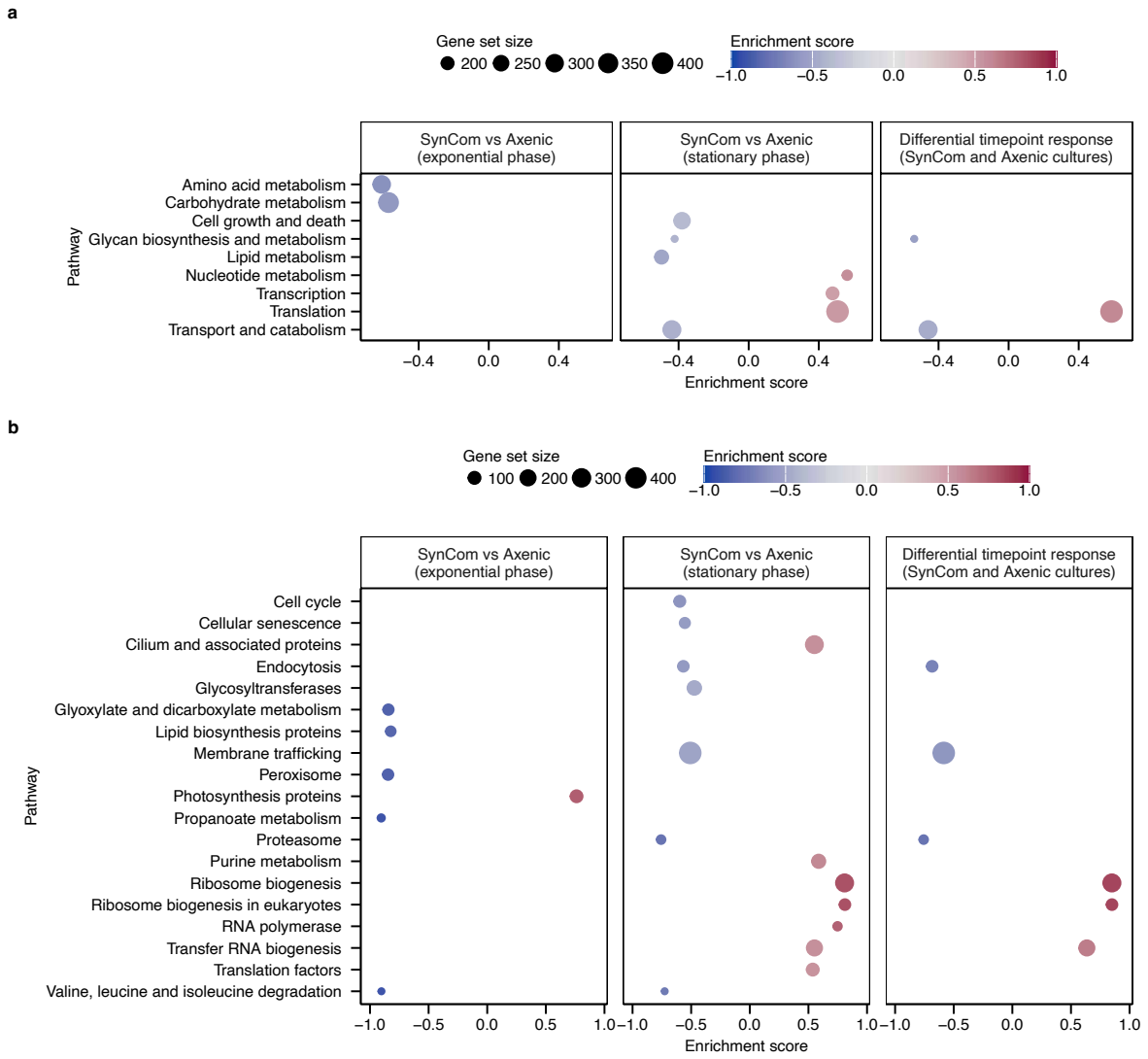


Figure 3.8 Curated KEGG functional category and pathway enrichment for *C. reinhardtii* grown in the co-culture with the SynCom relative to axenic controls. a – GSEA at the KEGG functional category (level B) for *C. reinhardtii* grown in the co-culture with SynCom relative to axenic controls in exponential and stationary phases. b – GSEA at the KEGG pathway level for the same comparisons. Both panels are faceted by comparisons: SynCom vs axenic (exponential phase), SynCom vs axenic (stationary phase), and the SynCom timepoint effect (stationary vs exponential). KEGG annotations were curated to exclude pathways not relevant to *C. reinhardtii* biology, i.e., animal-specific processes. Points are sized according to gene set sizes and colored by enrichment scores, with shades of blue indicating downregulation of a process or pathway, and red – upregulation.

GSEA revealed distinct enrichment patterns depending on the SynCom presence and *C. reinhardtii* growth phase. Among the processes with negative enrichment scores in the co-cultures with the SynCom in the exponential phase were amino acid and carbohydrate metabolism, and in the stationary phase, lipid metabolism and glycan biosynthesis and metabolism (Fig. 3.8a). The latter is connected to the GO MF term "hydrolase activity, hydrolyzing O-glycosyl compounds" (Fig. 3.7) and the KEGG

"glycotransferases" term negatively enriched in the stationary phase in response to the SynCom (Fig. 3.8b). Together, they indicate that *C. reinhardtii* reduced glycan turnover in the SynCom presence. As the *C. reinhardtii* cell wall is rich in glycoproteins[197,198], this pattern suggests that in the SynCom presence, the cell wall structure rich in hydroxyproline O-glycans may be stabilized, instead of being actively remodeled. In previous research on *Arabidopsis thaliana* roots, SynCom response included upregulation of genes involved in cell wall remodeling, including cell wall polysaccharide biosynthetic and metabolic processes, and biosynthesis of cell wall components[84]. In *C. reinhardtii*, although SynCom presence seems to induce transcriptional regulation of genes encoding proteins important for the cell wall structure, glycan turnover is decreased, and cell walls are rather stabilized than newly synthesized or remodeled.

While no functional modules showed a positive enrichment score in the co-cultures with the SynCom in the exponential phase, in the stationary phase, we detected positive enrichment of nucleotide metabolism, transcription, and translation. Although translation was positively enriched in the timepoint effect comparisons in both Axenic and SynCom treatments (Supplementary Figure S3.7), it showed positive enrichment in the differential timepoint response between SynCom and Axenic cultures comparison, indicating that translation dynamics differed between *C. reinhardtii* cultivated axenically and with the SynCom, and the magnitude of increase was higher in the SynCom presence (Fig. 3.8a).

To get a better understanding of the processes enriched among the *C. reinhardtii* DEGs, we analyzed the KEGG pathway enrichment. The gene set "photosynthesis proteins" was positively enriched in the co-cultures with the SynCom in the exponential phase, suggesting that *C. reinhardtii* photosynthetic capacity may be increased in the SynCom presence, possibly due to CO_2 provision by the bacteria (Fig. 3.8b, Supplementary Figure 3.8). This observation is consistent with the hypothesis that the secreted organic carbon drives the assembly of microbial communities associated with the alga[105]. We also detected negative enrichment of "propanoate metabolism", "glyoxylate and dicarboxylate metabolism" and "valine, leucine and isoleucine degradation" pathways (Supplementary Figure S3.9–S3.12), corresponding to the patterns observed in the analysis of GO BP terms enrichment in

the clusters of SynCom-responsive DEGs (Fig. 3.5), in which we detected both downregulation of glyoxylate cycle and BCAA catabolism. Together, the results of GO BP term and KEGG enrichment indicate that in the SynCom presence, *C. reinhardtii* metabolism shifts from catabolism-based to photosynthesis-based.

Additionally, consistent with the enrichment of GO BP term "regulation of peroxisome size" among the downregulated genes in the SynCom vs axenic comparison in the exponential phase (Fig. 3.6), we also detected negative enrichment of the "peroxisome" KEGG pathway in the comparison of SynCom vs axenic cultures in the exponential phase (Fig. 3.8b, Supplementary Figure S3.13), suggesting that in the co-cultures with the SynCom, expression of genes associated with peroxisomal metabolism and controlling peroxisome volume is reduced. Peroxisomes in *C. reinhardtii* are involved in β-oxidation of fatty acids[199,200], which is in line with the negative enrichment of "Lipid metabolism" module and "Lipid biosynthesis proteins" in the SynCom presence (Fig. 3.8).

3.3.5 SynCom induces metabolic changes in *C. reinhardtii*

Integrating GO and KEGG enrichment results reveals that in response to SynCom presence, *C. reinhardtii* transcriptional responses result in major changes in metabolism. In the SynCom presence, *C. reinhardtii* upregulated photosynthesis (Fig. 3.8b) and simultaneously downregulated pathways such as the glyoxylate cycle and BCAA catabolism, used while sourcing energy from catabolic reactions (Fig. 3.5–3.8), suggesting SynCom might contribute to CO_2 and nutrient availability, allowing *C. reinhardtii* to invest energy in biosynthesis and growth (Fig. 3.2, 3.5–3.8). Several genes encoding enzymes playing important roles in alternative to photosynthesis pathways leading to energy generation were downregulated in *C. reinhardtii* in the SynCom presence, including isoamylase, 3-methyl-2-oxobutanoate dehydrogenase, acetate-CoA ligase, and isocitrate lyase. Isoamylase, encoded by the *C. reinhardtii* gene Cre03.g155001, was significantly downregulated in the stationary phase in response to the SynCom (Supplementary Table S3.1). Isoamylase catalyzes the reaction of starch degradation to maltodextrin. In *C. reinhardtii*, starch is produced through photosynthetic carbon fixation during photoautotrophic growth and serves as a long-term storage carbohydrate that can be utilized as an energy source under stress conditions[201,202]. Apart from starch, amino acids can also serve as an energy source in

carbon-limiting or stress conditions[203,204]. In the SynCom presence, genes involved in amino acid biosynthesis were upregulated, whereas genes involved in amino acid catabolism were downregulated, including those playing a role in the BCAA degradation (Fig. 3.5, 3.8, Supplementary Figure S3.11, S3.12). Moreover, the genes encoding the key enzyme involved in BCAA degradation, 3-methyl-2-oxobutanoate dehydrogenase (Cre06.g311050, Cre12.g539900), were significantly downregulated in the SynCom presence (Fig. 3.5, 3.7, Supplementary Table S3.1). BCAA can be used as respiratory substrates[203], and BCAA catabolism was shown to contribute to the biosynthesis and metabolism of triacylglycerols in *C. reinhardtii*[204]. In our experiment in a nitrogen-rich medium[205], we also identified negative enrichment of "lipid metabolism" and "lipid biosynthesis proteins", suggesting that in the SynCom presence, *C. reinhardtii* does not activate lipid biosynthesis to accumulate carbon in the form of lipids.

Another important change in *C. reinhardtii* metabolism in the SynCom presence is the downregulation of the glyoxylate cycle genes (Fig. 3.5, 3.8). Since in the gnotobiotic photobioreactor, *C. reinhardtii* grows photoautotrophically on an acetate-free medium, downregulation of glyoxylate cycle genes does not indicate changes in metabolism of supplied acetate, but rather changes in the endogenous acetate metabolism. Acetate is a major fermentation product in *C. reinhardtii* and can be used for the biosynthesis of other metabolites[206]. The key enzyme of the glyoxylate cycle, isocitrate lyase (Cre06.g282800, Supplementary Table S3.1), which catalyzes the conversion of isocitrate to glyoxylate, was linked to changes in response to oxidative stress under mixotrophic growth[207]. The inactivation of isocitrate lyase in *C. reinhardtii* led to a decrease in the amounts of glyoxylate cycle and gluconeogenesis enzymes, and, consequently, to a stimulation of amino acid synthesis and to an increase in fatty acid content, thereby enabling adaptation to oxidative stress[207]. Moreover, under nitrogen (N) deprivation, acetate was shown to be directed into fatty acid biosynthesis and triacylglycerol (TAG) accumulation, allowing cells to store carbon during nitrogen-limiting conditions, when growth is limited[208]. In our experiment, there was no nitrogen limitation[205,208], and consequently, we did not observe induction of TAG biosynthesis. The significant downregulation of the gene encoding the isocitrate lyase in the SynCom presence in the stationary phase, and at the same time, the negative enrichment of the glyoxylate metabolism (Fig. 3.5, 3.8),

accompanied by the downregulation of BCAA catabolism (Fig. 3.5, 3.7, 3.8, Supplementary Figure S3.11, S3.12) and lipid biosynthesis proteins (Fig. 3.8), suggest that in the co-cultures with the SynCom, SynCom bacteria might consume acetate produced by *C. reinhardtii*, minimizing the need for acetate recycling in the glyoxylate cycle. In the co-culture with the SynCom, carbon is used for biosynthesis (Fig. 3.5, 3.6, 3.8), enabling *C. reinhardtii* cells to grow more than in axenic cultures (Fig. 3.2). This further suggests that in the co-culture with the SynCom, nutrients are replete and photosynthesis can be upregulated (Fig. 3.8), possibly due to CO_2 provided by the bacteria, which together suggest a beneficial impact of the SynCom on *C. reinhardtii* growth.

3.3.6 *C. reinhardtii* transcriptional responses to SynCom show evolutionary conservation with *A. thaliana* and *L. japonicus*

After functional analysis of DEGs, we examined the conservation of the *C. reinhardtii* transcriptional response to SynCom with land plants to determine the extent to which transcriptional responses to bacterial presence are conserved among evolutionarily distant photosynthetic organisms. We decided to compare *C. reinhardtii* data with data collected previously in experiments with *A. thaliana* and *Lotus japonicus*, in which each plant was grown either axenically or with its native (*A. thaliana*-derived for *A. thaliana*, *L. japonicus*-derived for *L. japonicus*) or non-native SynCom (derived from the other plant) in a soil-based gnotobiotic system, and root transcriptomes were analyzed[84]. To compare the transcriptional responses in the three host organisms, we first performed a BLASTP[169,170] search for Reciprocal Best Hits (RBH) for each pair of organisms, *C. reinhardtii – A. thaliana, C. reinhardtii – L. japonicus,* and *A. thaliana – L. japonicus*, and identified genes conserved within each pair and within all three organisms. We expected more genes to be conserved between the two more closely related land plants, *A. thaliana* and *L. japonicus*, than between *C. reinhardtii* and one of the land plants or among all three organisms. Consistent with our expectations, 84.1% of *C. reinhardtii* genes were not conserved, but only 57.6% and 61.2% of *A. thaliana* and *L. japonicus* genes were not conserved, respectively (Fig. 3.9a, Supplementary Table S3.8). Among the conserved genes in *A. thaliana* and *L. japonicus,* above 30% were conserved between the two, more than 6% were conserved between all three organisms, and only a small proportion was conserved

between each plant and *C. reinhardtii*. Among the *C. reinhardtii* genes, 10.6% were conserved between all three organisms (Fig. 3.9a, Supplementary Table S3.8).

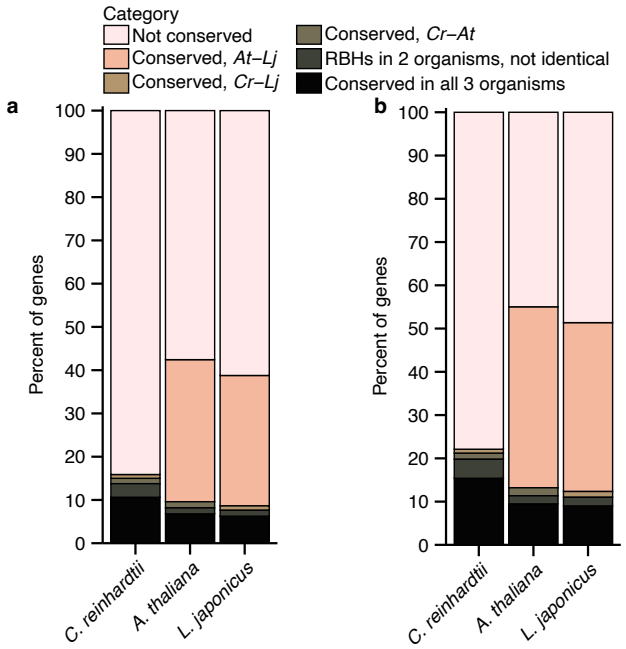


Figure 3.9 Conservation of *C. reinhardtii* genes with *A. thaliana* and *L. japonicus*. a – Percentages of *C. reinhardtii*, *A. thaliana*, and *L. japonicus* genes conserved between the organisms and the ones unique to each organism. b – Conservation of genes differentially expressed in response to SynCom presence in *C. reinhardtii*, *A. thaliana*, and *L. japonicus*.

After analyzing the conservation of all genes among the three organisms, we specifically focused on the SynCom-responsive DEGs. The proportions of conserved genes were higher for the DEGs than for all genes, with 15.4% *C. reinhardtii* and 9.5% and 9.0% *A. thaliana* and *L. japonicus* SynCom-responsive DEGs being genes conserved among all three organisms, respectively (Fig. 3.9b, Supplementary Table S11). In *A. thaliana* and *L. japonicus*, 41.8% of *A. thaliana* DEGs and 39.0% of *L. japonicus* DEGs belonged to genes conserved between the two plants (Fig. 3.9b, Supplementary Table S3.9). These observations suggest that some parts of the transcriptional response to SynCom involve genes conserved among *C. reinhardtii* and distantly related land plants.

In both land plants, the native SynComs induced a specific transcriptional response involving components of the host immune response[84]. Since the *C. reinhardtii* immune system is much less developed than that of land plants[46], we did not expect a large overlap among the genes responding transcriptionally to the SynComs across all three

analyzed organisms. To investigate the degree of conservation of the *C. reinhardtii* transcriptional response to the SynCom, we analyzed down- and upregulated DEGs separately and compared their conservation and similarity in regulation in response to the SynCom treatment. Among the downregulated *C. reinhardtii* DEGs, slightly above 20% belonged to genes with some degree of conservation (not unique to *C. reinhardtii*), and within those, more than half were conserved at least in two organisms (*C. reinhardtii – A. thaliana, C. reinhardtii – L. japonicus*, or conservation among all three organisms) and differentially expressed in response to the SynCom treatment (Fig. 3.10a-b).

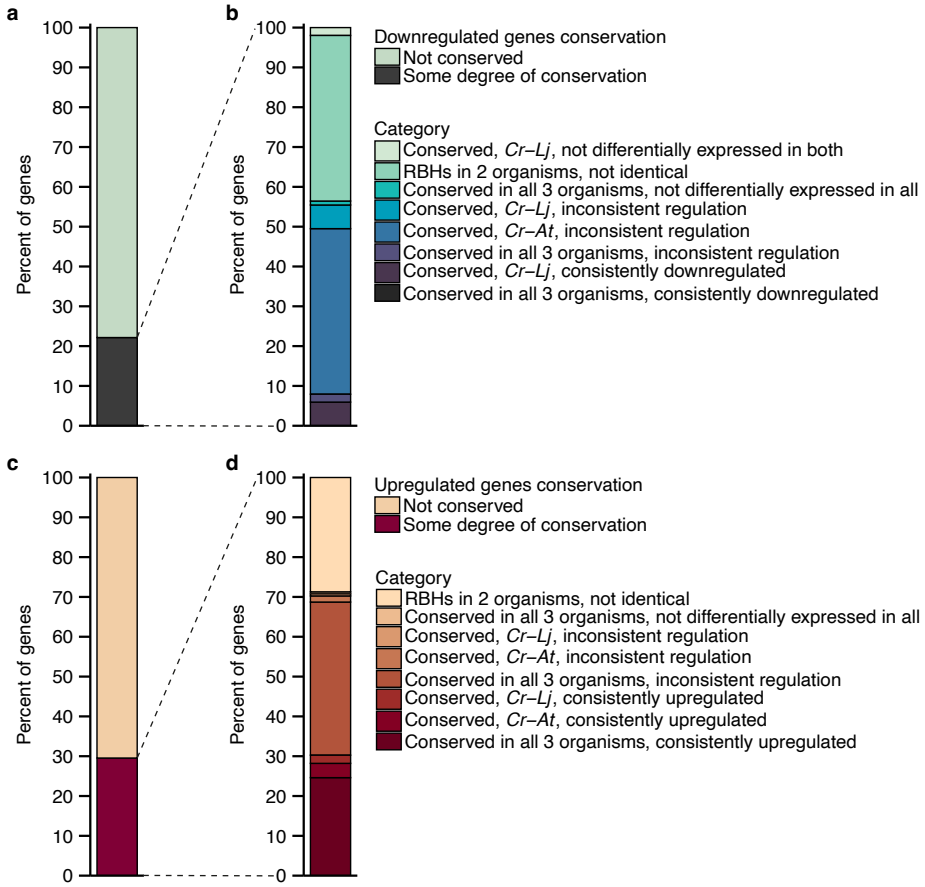


Figure 3.10 Conservation of *C. reinhardtii* genes down- and upregulated in response to SynCom presence. a – Percentage of *C. reinhardtii* downregulated DEGs with an RBH in *A. thaliana*, *L. japonicus*, or both organisms. b – Detailed description of the conservation level of the conserved downregulated genes. c – Percentage of *C. reinhardtii* upregulated DEGs with an RBH in *A. thaliana*, *L. japonicus*, or both organisms. d– Detailed description of the conservation level of the conserved upregulated genes.

Interestingly, 5.9% *of C. reinhardtii* DEGs belonged to genes conserved among all three organisms and were consistently downregulated in the presence of SynCom

(Supplementary Table S3.10), suggesting that some parts of the transcriptional response are deeply conserved. While slightly more than 20% of downregulated *C. reinhardtii* DEGs belonged to genes with some degree of conservation, for the upregulated *C. reinhardtii* DEGs, this proportion was higher — almost 30% (Fig. 3.10c). Among these, slightly above 70% were differentially expressed in response to the SynCom, and 30.2% were conserved and upregulated in response to the SynCom in at least two organisms (Fig. 3.10d, Supplementary Table S3.11). Surprisingly, 24.6% of *C. reinhardtii* genes upregulated in response to the SynCom were also upregulated in *A. thaliana* and *L. japonicus*[84]. These observations indicate that parts of the response to commensal bacteria are conserved between *C. reinhardtii* and evolutionarily distant land plants *A. thaliana* and *L. japonicus*, suggesting these responses might represent ancestral responses to commensal bacteria.

3.3.7 Candidate genes reveal conserved programs for growth, translation, and metabolic exchange

After analyzing the conservation of DEGs among *C. reinhardtii*, *A. thaliana,* and *L. japonicus*, we decided to further narrow down the list of candidate genes. To prioritize genes, each DEG was assigned a score (0–9) based on: conservation (RBH in *A. thaliana* and *L. japonicus*), consistent upregulation in response to SynCom across species, and membership in enriched functional categories (KEGG pathways, GO BP and MF terms) in *C. reinhardtii*. We focused on genes that were conserved between *C. reinhardtii*, *A. thaliana,* and *L. japonicus* and (1) upregulated in response to bacteria in both *A. thaliana* and *L. japonicus*, with a $\log_2$FC of at least 1.25 in the SynCom vs axenic comparison in either exponential or stationary growth, or (2) upregulated in response to the native SynCom treatment in both *A. thaliana* and *L. japonicus*, with a $\log_2$FC of at least 0.5 in the SynCom vs axenic comparison in either exponential or stationary. We added one non-conserved *C. reinhardtii* gene (Cre12.g551352) to the candidate gene list based on its exceptionally strong upregulation in the stationary phase of *C. reinhardtii* growth ($\log_2$FC = 10.2) and previous reports on its importance in *C. reinhardtii*–bacteria interactions[116]. This approach allowed us to narrow down the candidate gene list to 21 genes, which can be grouped into four coordinated programs, manually curated based on the similarity of

the gene functions: growth preparation, translation machinery, protein homeostasis, and metabolic exchange (Fig. 3.11, Table 3.1, Supplementary Table S3.12).

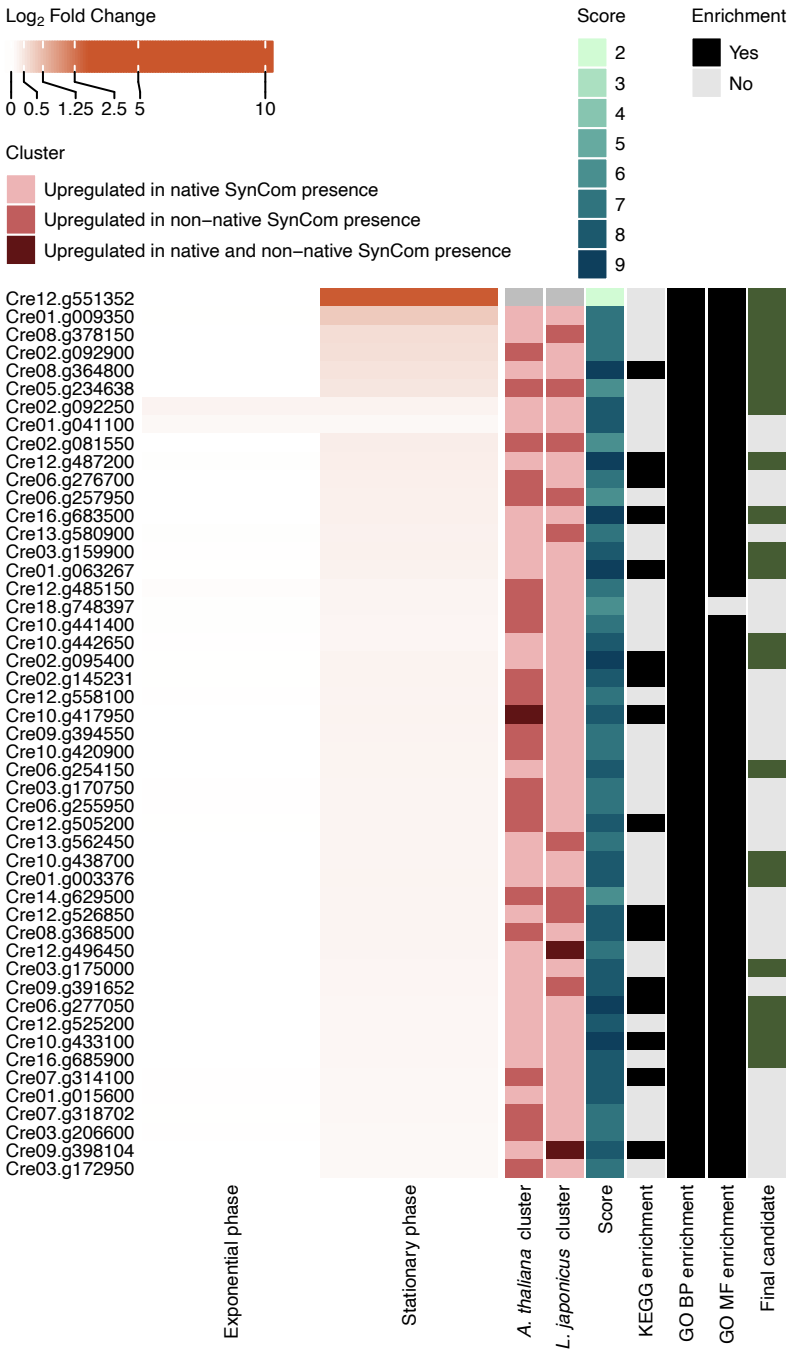


Figure 3.11 Heatmap of genes differentially regulated in response to SynCom presence in *C. reinhardtii*, conserved between *C. reinhardtii*, *A. thaliana*, *L. japonicus,* and upregulated in *A. thaliana*, *L. japonicus* in response to bacteria, and a candidate gene added based on its strong upregulation in SynCom presence in the stationary phase of *C. reinhardtii* growth. DEGs were scored (0–9 points) based on conservation with *A. thaliana* and *L. japonicus,* cross-species SynCom-responsive upregulation, and membership in enriched KEGG/GO terms. See Materials and methods for details. Heatmap shows log_2FC in the exponential and stationary phases of growth, expression clusters of the RBHs in

A. thaliana and *L. japonicus*, gene scores, membership in the enriched KEGG/GO terms, and presence in the final candidate list.

Table 3.1 Candidate genes differentially expressed in response to SynCom presence, with functional annotation, expression changes across growth phases, KEGG pathway assignment, and conservation in *A. thaliana* and *L. japonicus*.

Gene	**Description**	**Program**	**Log_2FC, SynCom vs axenic, exponential phase**	**Log_2FC, SynCom vs axenic, stationary phase**	**KEGG pathway**	***A. thaliana* RBH**	***L. japonicus* RBH**	**Conservation**
Cre02.g092250	FKBP42, Peptidylprolyl isomerase	protein homeostasis	0.71	0.72	Chaperones and folding catalysts	AT3G21640	LotjaGi3g1v0550000	Conserved in all 3 organisms
Cre01.g003376	DNAJC2; DnaJ homolog subfamily C member 2	protein homeostasis	0.00	0.66	Chaperones and folding catalysts	AT3G11450	LotjaGi6g1v0362500	Conserved in all 3 organisms
Cre16.g683500	DEAD/DEAH box helicase, ATP-dependent RNA helicase DDX49/DBP8 (*A. thaliana*: RH36)	translation machinery	-0.03	0.88	Ribosome biogenesis	AT1G16280	LotjaGi3g1v0134400	Conserved in all 3 organisms
Cre01.g009350	agmatine iminohydrolase (aguA)	metabolic exchange	0.02	3.24	Arginine and proline metabolism	AT5G08170	LotjaGi4g1v0258800	Conserved in all 3 organisms
Cre01.g063267	bystin-like protein	translation machinery	0.02	0.80	Ribosome biogenesis	AT1G31660	LotjaGi2g1v0100300	Conserved in all 3 organisms
Cre02.g092900	GMP synthase (glutamine-hydrolyzing)	growth preparation	0.00	1.98	Purine metabolism, Peptidases and inhibitors	AT1G63660	LotjaGi4g1v0426900	Conserved in all 3 organisms
Cre02.g095400	ribosome biogenesis protein (BRX1, BRIX1)	translation machinery	0.02	0.75	Ribosome biogenesis	AT3G15460	LotjaGi1g1v0523100	Conserved in all 3 organisms
Cre03.g159900	RIO kinase 2 (RIOK2)	translation machinery	0.03	0.83	Ribosome biogenesis in eukaryotes, Protein kinases, Ribosome biogenesis	AT3G51270	LotjaGi1g1v0187700	Conserved in all 3 organisms
Cre03.g175000	3'-5' exoribonuclease PH component of the exosome (RRP42, EXOSC7)	translation machinery	0.00	0.57	RNA degradation, Messenger RNA biogenesis	AT3G60500	LotjaGi1g1v0741200	Conserved in all 3 organisms
Cre05.g234638	amidophosphoribosyltransferase (purF, PPAT)	growth preparation	0.00	1.56	Purine metabolism, Alanine, aspartate and glutamate metabolism, Peptidases and inhibitors	AT2G16570	LotjaGi3g1v0515700	Conserved in all 3 organisms

Cre06.g254150	Nucleolar protein, component of the U3 processome; periodic tryptophan protein 2 (PWP2, UTP1)	translation machinery	0.00	0.69	Ribosome biogenesis in eukaryotes, Ribosome biogenesis	AT1G15440	LotjaGi1g1v0124400	Conserved in all 3 organisms
Cre06.g277050	protein SDA1 (SDA1, SDAD1); ARM repeat superfamily protein	translation machinery	0.01	0.54	Ribosome biogenesis	AT1G13160	LotjaGi1g1v0302300	Conserved in all 3 organisms
Cre08.g364800	phosphoribosylformylglycinamidine synthase (A. thaliana PUR4)	growth preparation	0.00	1.68	Purine metabolism	AT1G74260	LotjaGi6g1v0063600	Conserved in all 3 organisms
Cre08.g378150	glucose-6-phosphate 1-dehydrogenase (G6PD)	growth preparation	0.00	2.09	Pentose phosphate pathway, Glutathione metabolism, Exosome	AT5G13110	LotjaGi1g1v0531700	Conserved in all 3 organisms
Cre10.g433100	GTP-binding protein (obgE)	translation machinery	0.00	0.53	Ribosome biogenesis	AT5G18570	LotjaGi6g1v0307800	Conserved in all 3 organisms
Cre10.g438700	U3 small nucleolar RNA-associated protein 12 (DIP2, UTP12, WDR3)	translation machinery	0.01	0.64	Ribosome biogenesis in eukaryotes, Ribosome biogenesis	AT3G21540	LotjaGi1g1v0685500	Conserved in all 3 organisms
Cre10.g442650	ATP-dependent RNA/DNA helicase	translation machinery	0.05	0.58	Ribosome biogenesis, Translation factors	AT2G03270	LotjaGi1g1v0721200	Conserved in all 3 organisms
Cre12.g487200	ribosome assembly protein (SQT1)	translation machinery	0.06	0.99	Ribosome biogenesis	AT1G71840	LotjaGi5g1v0344600	Conserved in all 3 organisms
Cre12.g525200	nucleolar protein 56 (NOP56)	translation machinery	0.02	0.53	Ribosome biogenesis in eukaryotes, Ribosome biogenesis	AT1G56110	LotjaGi4g1v0223200	Conserved in all 3 organisms
Cre12.g551352	periplasmic L-amino acid oxidase 1 (LAO1)	metabolic exchange	0.00	10.20	NA	NA	NA	Chlamydomonas only
Cre16.g685900	protein arginine methyltransferase (PRMT3)	translation machinery	-0.05	0.51	Chromosome and associated proteins	AT3G12270	LotjaGi3g1v0107900	Conserved in all 3 organisms

The first program, growth preparation, includes genes important for the purine biosynthesis pathway: amidophosphoribosyltransferase (Cre05.g234638), guanosine monophosphate synthase (GMP-synthase, Cre02.g092900), phosphoribosylformylglycinamide synthase (Cre08.g364800), and glucose-6-phosphate 1-dehydrogenase (G6PD, Cre08.g378150), (Table 3.1, Supplementary Table S3.12). *De novo* purine biosynthesis is a metabolically costly process that needs to be under tight control, and the biosynthesized purines play an important role in cell growth, proliferation, and energy homeostasis[209–211] . The biosynthesis of adenosine monophosphate (AMP) and guanosine monophosphate (GMP) requires inosine 5'-monophosphate (IMP), and amidophosphoribosyltransferase (EC:2.4.2.14), catalyzing one of the first reactions leading to IMP production, was significantly upregulated in the stationary phase of *C. reinhardtii* growth in the SynCom presence. Amidophosphoribosyltransferase catalyzes the conversion of 5-phosphoribosyl-1-pyrophosphate (PRPP) into phosphoribosylamine, using the amine group from a glutamine side-chain[212] (Supplementary Figure S3.14). PRPP is supplied from the pentose phosphate pathway, where another one of the candidate genes, G6PD (EC:1.1.1.49 1.1.1.363), catalyzes the conversion of β-D-glucose 6-phosphate to D-glucono-1,5-lactone 6-phosphate. Moreover, the genes encoding enzymes catalyzing the later steps of PRPP synthesis, 6-phosphogluconate dehydrogenase (EC:1.1.1.44 1.1.1.343, Cre12.g526800) and ribose-phosphate pyrophosphokinase (EC:2.7.6.1, Cre08.g37880) were significantly upregulated in the SynCom presence in the stationary phase. In the later steps of GMP synthesis, phosphoribosylformylglycinamide synthase (EC:6.3.5.3) catalyzes the conversion of 5'-phosphoribosyl-N-formylglycinamide to 2-(formamido)-N1-(5'–phosphoribosyl)acetamidine. The rate-limiting step of GMP synthesis is the oxidative reaction in which IMP is transformed into xanthosine 5'-monophosphate (XMP) catalyzed by IMP dehydrogenase[213] (IMPDH, EC:1.1.1.205, Cre14.g614300), which was also upregulated in the stationary phase in the SynCom presence. Finally, GMP-synthase (EC:6.3.5.2) aminates XMP to guanosine monophosphate (GMP)[214]. Additionally, the adenylate kinase (EC:2.7.4.3, Cre02.g081550), which catalyzes the reversible conversion of two ADP molecules into one ATP and one AMP, was significantly upregulated in the stationary phase in the SynCom presence. A constant supply of purines is crucial for the maintenance of growth and proliferation, as well as for energy homeostasis in cells, and *de novo* purine biosynthesis is a metabolically costly process that is tightly controlled[209–211] . Upregulation of multiple genes

important for the *de novo* purine biosynthesis pathway, supplying nucleotides — the building blocks of DNA, RNA, and metabolites central for metabolism, such as ATP or NADH[214] – suggests that in the SynCom presence, *C. reinhardtii* can afford investing into an energetically expensive pathway important for growth. Interestingly, although all these genes were upregulated in response to the native SynCom in *C. reinhardtii*, their expression was not consistent in *A. thaliana* and *L. japonicus*. In *C. reinhardtii* and *L. japonicus*, genes encoding GMP-synthase were upregulated in the native SynCom presence; in *A. thaliana*, the gene was upregulated in the non-native SynCom presence. The opposite pattern was true for the genes encoding G6PD, and the genes encoding amidophosphoribosyltransferase were upregulated in response to the non-native SynCom in *A. thaliana* and *L. japonicus*, while *C. reinhardtii* gene was upregulated in response to the native SynCom[84]. The observed patterns, slightly varying across the three photosynthetic organisms, suggest that increased expression of purine biosynthesis genes may be a conserved response to the presence of possibly commensal bacteria, rather than being limited to the native bacterial communities.

One process closely linked to *de novo* purine biosynthesis is ribosome biogenesis[215]. Among the 21 candidate genes, 12 can be assigned to the "translation machinery" program, including predominantly genes involved in ribosome biogenesis (Table 3.1, Supplementary Table S3.12, Supplementary Figure S3.15). Genes important for the 40S (small subunit) ribosome biogenesis include DEAD/DEAH box helicase (Cre16.g683500), bystin-like protein (Cre01.g063267), nucleolar protein (Cre06.g254150) and RIO kinase 2 (RIO2, Cre03.g159900). In *A. thaliana*, bystin-like protein (AT1G31660, RBH to Cre01.g063267) was found to be a component of the pre-40S ribosome, essential for proper female and male gametophyte development[216]. Two other proteins, RH36 DEAD/DEAH box helicase (AT1G16280, RBH to Cre16.g683500) and Pwp2 (AT1G15440, RBH to Cre06.g254150), were found to be essential for female gametophyte development[216,217]. RH36 is homologous to yeast Dbp8p helicase, which is essential for viability and proper rRNA processing, and its depletion causes a decrease in the levels of the small ribosomal subunit[218]. RIO2 was shown to be essential for 40S ribosomal subunit maturation in both yeast and human cells[219–221], and to be important for the prevention of premature translation initiation[222] and cell cycle progression in cancer cells[223,224]. The upregulation of the expression of the genes important for the 40S ribosome subunit generation in the

SynCom presence suggests that *C. reinhardtii* invests in translation machinery instead of transitioning into a quiescent state. The increased expression of the 40S subunit ribosome biosynthesis genes is accompanied by upregulation of genes involved in 60S large ribosome subunit generation: ribosome biogenesis protein (Cre02.g095400, BRX1/BRIX1), 3'-5' exoribonuclease PH component of the exosome (Cre03.g175000), SDA1 ARM repeat superfamily protein (Cre06.g277050, SDA1)[225], nucleolar protein 56 (Cre12.g525200, NOP56), playing a role in directing 2′-O-methylation of rRNA during ribosome biogenesis[226,227], and ribosome assembly protein (Cre12.g487200, SQT1) participating in a late step of 60S subunit assembly or modification in *S. cerevisiae*[228]. Associated with the pre-60S ribosomal subunit BRX1 is involved in pre-rRNA processing and is important for early plant development[229]. AT3G60500, *A. thaliana* RBH of Cre03.g175000, the 3'-5' exoribonuclease PH component of the exosome, is a component of the conserved core complex of the RNA exosome, Exo9, which in *A. thaliana*, associates with ribonucleases and other cofactors, and contributes to rRNA maturation processes[230,231]. Another gene important for the pre-rRNA processing is a U3 small nucleolar RNA-associated protein 12 (Cre10.g438700)[232]. Cre10.g433100 encodes a GTP-binding protein homologous to AT5G18570, which encodes a chloroplast-targeting GTPase participating in chloroplast ribosome biogenesis that is essential for early embryogenesis[233]. Cre10.g442650 encodes a putative helicase, whose exact function remains to be characterized[234,235]. Another candidate gene important for translation is Cre16.g685900, encoding protein arginine methyltransferase (PRMT3), which in *A. thaliana* was shown to regulate pre-rRNA processing and to promote the translation of housekeeping mRNAs while simultaneously repressing stress-related mRNAs related[236,237].

The coordinated upregulation of both 40S and 60S ribosome biogenesis genes, as well as genes encoding proteins important for ribosome quality control, such as helicases, indicates that in the SynCom presence, *C. reinhardtii* translational capacity is maintained.

Another group of genes upregulated in the SynCom presence are genes related to protein homeostasis: peptidylprolyl isomerase Cre02.g092250, important for proper protein folding[238] and Cre01.g003376 encoding a DnaJ homolog subfamily C member 2 (DNAJC2) protein (Table 3.1, Supplementary Table S3.12). *A. thaliana*

Cre01.g003376 RBH is a widely expressed chromatin regulator important for multiple processes of plant growth and development, and acts as a co-chaperone[239–241].

Apart from the groups of genes related to growth, ribosome biogenesis, translation, and protein homeostasis, two other candidate genes, Cre12.g551352 and Cre01.g009350 encode proteins involved in metabolic exchange (Table 3.1, Supplementary Table S3.12). Cre12.g551352 encodes a periplasmic L-amino acid oxidase previously shown to be involved in indole-3-acetic acid (IAA) production and to support mutualistic interaction between *C. reinhardtii* and a plant growth-promoting bacterium, *Methylobacterium aquaticum*, under N-limiting conditions[116]. LAO1 protein deaminates a broad range of amino acids, producing ammonium, which is imported into the *C. reinhardtii* cells and α-keto acids, which remain extracellular[116,124]. These extracellular α-keto acids may serve multiple ecological functions in the phycosphere. Depending on the amino acid substrate, LAO1-derived α-keto acids can act as carbon sources for nearby bacteria, be precursors for hormone biosynthesis, for instance, IPyA from L-tryptophan is a precursor for IAA biogenesis[116], scavenge hydrogen peroxide, like pyruvate from L-alanine or oxaloacetate from L-aspartate[242], or act as siderophores improving iron nutrition, like two α-keto acids from BCAA: α-ketoisocaproic from L-leucine and α-ketoisovaleric from L-valine[243]. The potential production of BCAA-derived α-keto acids is consistent with the downregulation of BCAA catabolism in the co-cultures with the SynCom (Fig. 3.8), suggesting that BCAA may be preferentially channeled toward LAO1-mediated oxidation rather than intracellular degradation.

Although the LAO1 gene is not conserved between *C. reinhardtii*, *A. thaliana*, and *L. japonicus*, it was included based on its exceptionally strong upregulation in the stationary phase of *C. reinhardtii* growth ($log_2FC = 10.2$) and previous reports on its importance for *C. reinhardtii*–bacteria interactions. Since LAO1 has been shown to be important for the *C. reinhardtii – M. aquaticum* interaction under N-limiting conditions[116], we analyzed the DEGs involved in nitrogen uptake and assimilation. The medium used in our experiments, TP10, contains 6.75 mM KNO_3 and 0.75 mM NH_4Cl, supplying both ammonium and nitrate (Supplementary Table S3.13). However, ammonium is preferred over nitrate as an N source, and negatively regulates the expression of nitrate assimilation genes[244,245]. We observed upregulation of

ammonium transporters, nitrate/nitrite transporters, nitrate reductases, and nitrite reductases over time, with some upregulated only in the SynCom presence (Supplementary Table S3.1, S3.14). Additionally, we detected enrichment of the GO BP term "response to nitrate" in the clusters of genes upregulated in the SynCom presence (Fig. S3.2), further supporting the observed SynCom-dependent changes in N-metabolism and uptake. Interestingly, although we detected some upregulation of two genes encoding ammonium transporters over time, independent of bacteria presence, five others were induced only in the co-cultures (Supplementary Table S3.14). As *C. reinhardtii* preferentially uses ammonium over nitrate, and usually, ammonium presence represses the nitrate assimilation pathway[245], we can hypothesize that over time, ammonium in the medium is depleted, and the *C. reinhardtii* cells upregulate ammonium transporters to scavenge the remaining ammonium, simultaneously activating the expression of genes involved in the assimilation pathway of the less preferred N source – nitrate. Additionally, since in the SynCom presence, *C. reinhardtii* strongly upregulates the LAO1 gene (Table 3.1, Supplementary Table S3.12), and the LAO1 protein produces ammonium, increasing the pool of ammonium available to the alga, ammonium can be imported into the cell, which activates the expression of ammonium transporters. While the differences in the expression of nitrite and nitrate reductases between C. *reinhardtii* cultivated axenically or in the co-culture are rather minor, the GS/GOGAT genes important for the final steps of N assimilation — incorporation of ammonium into carbon skeletons[245,246] — were upregulated in a SynCom-dependent manner over time (Supplementary Table S3.14), which might be connected to the higher needs for N in the presence of the SynCom, caused by an upregulation of processes requiring N, such as amino acid biosynthesis and ribosome biogenesis. In the co-cultures, we also observed downregulation of genes involved in amino acid catabolism and increased expression of amino acid biosynthesis genes (Fig. 3.5). Since amino acids are recycled and used for protein synthesis under nutrient limiting conditions[203], and we observed both downregulation of amino acid catabolism genes in SynCom presence and upregulation of genes important for growth, LAO1 upregulation cannot be explained solely by ammonium depletion. Strong, SynCom-dependent LAO1 induction suggests that the SynCom presence alters *C. reinhardtii* response to nitrogen availability.

The other metabolic exchange candidate gene, Cre01.g009350, was highly upregulated in the SynCom presence in the stationary phase (log2FC = 3.24) and conserved between *C. reinhardtii*, *A. thaliana*, and *L. japonicus* (Table 3.1, Supplementary Table S3.12). It encodes a putative agmatine iminohydrolase. Agmatine iminohydrolase in plants catalyzes the conversion of agmatine to N-carbamoylputrescine, which is a step leading to the synthesis of putrescine, the smallest polyamine[247] (Supplementary Figure S3.16). In plants, polyamines participate in growth, development, and stress responses, including responses to pathogens[248,249]. Putrescine was shown to be important for the interaction of the endophytic root fungus *Piriformospora indica* with tomato and *A. thaliana*[95]. In both hosts, *P. indica* colonization promoted host growth in a putrescine-dependent manner, and arginine decarboxylase (ADC), an enzyme involved in the putrescine biosynthesis from arginine[95], was shown to be important for the interaction[95]. *P. indica* increases host ADC expression, which leads to a higher putrescine biosynthesis, which in turn stimulates *P. indica* growth and increases the abundance of IAA and gibberellic acid (GA), which lead to plant host growth promotion[95,250]. While this interaction involves a fungal partner, putrescine is also important for plant-associated bacteria[251]. Putrescine was suggested to serve as a signaling molecule in the rhizosphere, as a *Pseudomonas fluorescens* mutant unable to metabolize putrescine was impaired in its ability to colonize *A. thaliana* roots[103]. Additionally, ABC-type transporters of putrescine were enriched in the genomes of robust plant colonizer[251].

Interestingly, RBHs of agmatine iminohydrolase, significantly upregulated in the SynCom presence in *C. reinhardtii*, were also upregulated in native SynCom presence in *A. thaliana* and *L. japonicus*[84] (Supplementary Table S3.12). In addition to the upregulation of genes involved in putrescine biosynthesis, we identified the induction of genes encoding proteins involved in the arginine biosynthesis pathway: Cre06.g257950, Cre09.g400550 and Cre09.g416050 (Supplementary Table S3.1, Supplementary Figure S3.17). Cre06.g257950, upregulated in the SynCom presence, encodes an aspartate aminotransferase (EC:2.6.1.1), which catalyzes the conversion of 2-oxoglutarate to glutamate in the arginine biosynthesis pathway (Supplementary Figure S3.17). We identified Cre06.g257950 RBHs in both *A. thaliana* and *L. japonicus*, and in both land plants, these genes were upregulated in response to non-native SynComs. Two other genes active in the later steps of arginine biosynthesis,

Cre09.g400550 (EC:1.14.13.39) and Cre09.g416050, encoding argininosuccinate synthase, were upregulated in the stationary phase in the SynCom presence. Cre09.g416050 RBH in *A. thaliana* (AT4G24830) was also upregulated in the presence of the native SynCom[84]. The significant upregulation of several genes involved in the arginine biosynthesis pathway in *C. reinhardtii*, with some of them also upregulated in the presence of bacteria in land plants, suggests that upregulation of arginine biosynthesis in bacterial presence might be a conserved transcriptional response. In previous research, arginine was shown to be a chemoattractant for the bacterium *Azorhizobium caulinodans* colonizing legume roots[98]; however, in our experiment, the increased expression of these genes was coupled with the upregulation of agmatine iminohydrolase, which is involved in putrescine biosynthesis, indicating that the higher levels of arginine might be directly used for further putrescine production. The conserved upregulation of agmatine iminohydrolase in response to the native SynCom, along with the described putrescine–mediated growth promotion induced by an endophytic fungus colonization[95], indicates that agmatine iminohydrolase might be commonly important for plant-microbe interactions, and that it was present in the common ancestor of *C. reinhardtii* and land plants.

While only two of the candidate genes are involved in metabolic exchange (Table 3.1, Supplementary Table S3.12), another gene, Cre02.g108550, was highly upregulated in the SynCom presence in the stationary phase ($log_2FC = 2.84$) (Supplementary Table S3.1). While it is also conserved between *C. reinhardtii*, *A. thaliana* and *L. japonicus*, it was not consistently upregulated in the native SynCom presence. Cre02.g108550 encodes a bidirectional amino acid transporter 1 (BAT1), which in *A. thaliana* is active in the uptake of alanine and arginine, and export of glutamate and lysine[252]. In *C. reinhardtii,* arginine is the only amino acid known to be efficiently imported[253]. Additionally, Cre02.g108550 was recently shown to be upregulated in arginine-fed compared with ammonium-fed *C. reinhardtii*[191]. However, the uptake of L-arginine by *C. reinhardtii* is induced by the absence of nitrogen and inhibited by the presence of ammonium and nitrate[124,253,254]. Therefore, the upregulation of BAT1 in our system may either be connected to amino acid export or SynCom-dependent arginine uptake induced in ammonium and nitrate presence, contrary to what was described before in axenic cultures[124]. In plant-microbe interactions, amino acids can act as chemoattractants during the root colonization process[94,98,255]. Considering the strong

upregulation of the agmatine iminohydrolase gene, which is involved in the synthesis of putrescine from arginine, and BAT1 upregulation, it can be hypothesized that, besides the induction of arginine biosynthesis, *C. reinhardtii* imports arginine produced by the SynCom. However, based on the gene expression analysis only, it cannot be determined which amino acids are imported and exported. Both glutamate and lysine may be exported to facilitate the interactions between *C. reinhardtii* and SynCom, especially since both lysine and glutamic acid were shown to be chemoattractants during tomato root colonization[255], and amino acids released by the algae are known to be involved in both algae-bacteria and plant-bacteria interactions[82,256,257].

3.4 Discussion

Our results show that *C. reinhardtii* responds transcriptionally to its associated phycosphere community, extending our previous work demonstrating that *C. reinhardtii* can associate with soil-borne bacteria[105] and providing the first transcriptomic characterization of *C. reinhardtii* response to its phycosphere community. In the bacterial SynCom presence, *C. reinhardtii* growth is increased, and gene expression shifts from genes suggesting catabolism–based, stress-associated metabolism[201–204], towards growth– and biosynthesis–oriented metabolism, with coordinated upregulation of photosynthesis, ribosome biogenesis, translation machinery, protein homeostasis, and amino acid biosynthesis genes, and downregulation of genes involved in catabolism. Notably, parts of this response are conserved between *C. reinhardtii* and evolutionarily distant land plants, suggesting a possible ancestral origin. Below, we discuss two interconnected patterns emerging from our data: (1) the growth program shift and its conservation, and (2) a coordinated nitrogen and amino acid metabolism system involving LAO1, agmatine iminohydrolase, and BAT1.

In the SynCom presence, *C. reinhardtii* upregulated expression of genes involved in ribosome biogenesis, translation, and protein homeostasis in a coordinated manner (Fig. 3.5–3.8, Table 3.1). These included 40S and 60S ribosomal subunit biogenesis genes, rRNA maturation factors, and protein folding chaperones. In addition, photosynthesis and biosynthesis genes (Fig. 3.5–3.8, Table 3.1) were upregulated in the presence of SynCom. Concurrently, *C. reinhardtii* downregulated expression of

genes important for catabolism of storage compounds (starch, lipids, amino acids), suggesting a shift from nutrient recycling to active biosynthesis. This pattern indicates higher nutrient availability in the co-cultures with the SynCom than in the axenic *C. reinhardtii* cultures, possibly due to bacterial provision of CO_2 or other metabolites.

Notably, the native SynCom-dependent upregulation of *de novo* purine biosynthesis, ribosome biogenesis, and protein homeostasis genes was conserved across *C.* reinhardtii, *A. thaliana,* and *L. japonicus*[84]. While purines can act as bacterial chemoattractants[97], the coordinated upregulation of *de novo* purine biosynthesis genes with translation machinery suggests that the increased purine biosynthesis primarily supports *C. reinhardtii* growth rather than signaling. Nevertheless, whether *C. reinhardtii* also secretes some biosynthesized purines to attract bacteria remains to be determined. Importantly, conservation of the response at the transcriptional level does not necessarily imply identical downstream effects in *C. reinhardtii*, *A. thaliana*, and *L. japonicus*. While the increased translational activity in response to the native SynComs is conserved between the three organisms, the specific proteins translated later might differ.

The metabolic candidate gene with the most striking response to the bacterial SynCom was LAO1 (stationary phase $\log_2$FC = 10.2). LAO1 encodes an L-amino acid oxidase, highly expressed in nitrogen-depleted conditions, in which the LAO1 protein deaminates a broad range of amino acids, producing ammonium reimported by *C. reinhardtii*, and α-keto acids, which remain extracellular[124,258]. While LAO1 is typically upregulated under N-limiting conditions, several observations suggest it might have another function in our N-replete system.

Firstly, upregulation of LAO1 expression is accompanied by downregulation of amino acid catabolism genes and upregulation of amino acid biosynthesis genes (Fig. 3.5, 3.8), contradicting a starvation response. Secondly, some genes encoding ammonium transporters and GS/GOGAT pathway enzymes important for efficient ammonium assimilation[245,246] were upregulated in a SynCom-dependent manner over time (Supplementary Table S3.14), suggesting active assimilation of ammonium produced by LAO1 protein. Lastly, growth-related genes were upregulated, indicating metabolic

investment by *C. reinhardtii* in growth rather than in survival, stress–induced response.

Recent research showed that LAO1 mediates a mutualistic interaction between *C. reinhardtii* and a plant growth-promoting bacterium *M. aquaticum* through LAO1-mediated deamination of L-tryptophan into IPyA which is the first step of IAA biosynthesis[116]. While the accumulation of high IAA levels inhibits *C. reinhardtii* growth, IAA degradation by *M. aquaticum* can facilitate *C. reinhardtii* growth[116]. A similar mechanism might mediate the *C. reinhardtii*–SynCom interaction in our system; however, based on the conducted experiment, we cannot determine which specific amino acids are oxidized by LAO1 in our system, as LAO1 has a broad substrate range[258]. We propose that in our system, LAO1 functions as a metabolic exchange mediator rather than solely a nitrogen scavenger. In SynCom presence, *C. reinhardtii* produces α-keto acids that might act as chemoattractants for bacteria or be further metabolized by bacteria, and ammonium, which is reimported and assimilated by *C. reinhardtii*.

The LAO1 gene was not the only metabolic exchange candidate gene we identified. The other was Cre01.g009350, which encodes an agmatine iminohydrolase catalyzing the conversion of agmatine to N-carbamoylputrescine during the synthesis of putrescine[247]. The agmatine iminohydrolase gene was highly upregulated in the stationary phase in the SynCom presence, and its response to the SynCom was conserved between *C. reinhardtii*, *A. thaliana* and *L. japonicus* (Table 3.1, Supplementary Table S3.12)[84]. Putrescine can function as a bacterial chemoattractant in the rhizosphere[103], and was shown to be important for plant growth promotion by the endophytic root fungus *P. indica*[95]. In addition to the upregulation of agmatine iminohydrolase, we detected an increase in expression of genes involved in arginine biosynthesis in *C. reinhardtii*, suggesting an increased metabolite flux towards putrescine production.

The amino acid exchange between *C. reinhardtii* and its associated phycosphere bacteria may be facilitated by the bidirectional amino acid transporter BAT1, upregulated in the SynCom presence in *C. reinhardtii* (Supplementary Table S3.1). However, this gene was not included in our candidate gene list because, although it

was conserved among *C. reinhardtii*, *A. thaliana,* and *L. japonicus*, it was not consistently upregulated in all the organisms in response to the SynComs. In our system, BAT1 is most likely active in the uptake of arginine and may be active in the export of glutamate and lysine[191,252,253]. In plant-microbe interactions, amino acids can act as chemoattractants during root colonization process[94,98,255]. Combined with LAO1-mediated extracellular amino acid processing, this suggests a coordinated system for amino acid-based communication with bacteria, in which BAT1 might export amino acids that act as bacterial chemoattractants or serve as substrates for LAO1, especially since LAO1 was shown before to oxidize L-arginine with activity similar to that of L-tryptophan oxidation[259]. The α-keto acids produced by LAO1 may serve diverse functions, including acting as chemoattractants, carbon source for bacteria, intermediate for IAA biosynthesis[116] or siderophores improving nutrition[243], further shaping the phycosphere community. However, based on transcriptomics alone, we cannot determine which specific amino acids are transported by BAT1 and later metabolized in a LAO1-dependent or LAO1-independent manner. In addition, we cannot determine whether the amino acids oxidized by LAO1 are produced by *C. reinhardtii* or SynCom members.

Based on our findings, we propose a working model for the *C. reinhardtii* – SynCom interaction. In the presence of phycosphere bacteria, *C. reinhardtii* responds transcriptionally and shifts from expression of genes related to a catabolic, stress-associated metabolism toward an anabolic, growth-oriented state. This shift is supported by bacterial contributions (possibly CO_2 and metabolite exchange) and involves coordinated upregulation of biosynthetic pathways, translation machinery, and metabolic exchange genes. Central to the proposed interaction are LAO1-mediated amino acid oxidation along SynCom-dependent upregulation of ammonium transport and assimilation, agmatine iminohydrolase-mediated putrescine biosynthesis, and BAT1-mediated amino acid exchange (Fig. 3.12).

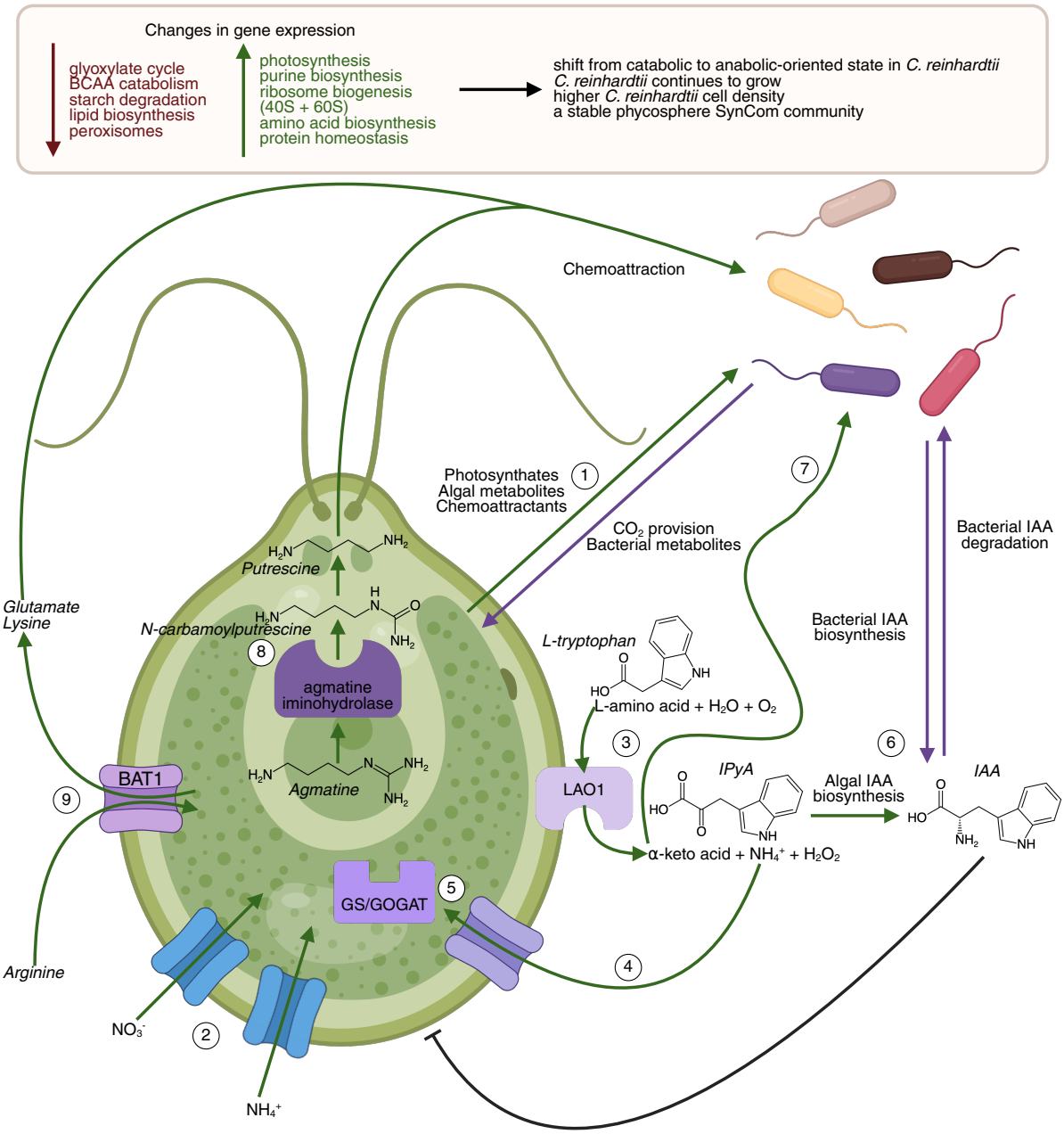


Figure 3.12 Working model of the *C. reinhardtii* – SynCom interaction. The upper rectangular panel shows a summary of the transcriptional changes in the SynCom presence with the downregulated processes (red), upregulated processes (green), and the observed outcomes on the right. Below, the proposed model of possible metabolic exchange between *C. reinhardtii* and phycosphere bacteria is shown. This model is based on transcriptomics data; metabolite fluxes and enzymatic activities were not directly measured. Processes differentially regulated in the SynCom presence are shown in purple, SynCom-independent, time-dependent changes in blue. Green arrows indicate *C. reinhardtii*-mediated processes, purple arrows indicate SynCom-mediated processes. Specific metabolite names are shown in italics. Numbers on the scheme refer to: (1) – proposed mutualistic exchange of CO_2, photosynthates, metabolites and chemoattractants; (2) – time-dependent, SynCom-independent upregulation of nitrate and ammonium transporters; (3) – SynCom-specific LAO1 induction, with LAO1 producing ammonium and α-keto acids; (4) – SynCom-specific ammonium transporter induction, possibly importing ammonium produced in the LAO1-catalyzed reaction; (5) ammonium assimilation via the GS/GOGAT pathway; (6) LAO1-mediated oxidation of L-tryptophan to IPyA, which can be converted to IAA; IAA

may be degraded by SynCom bacteria as described for the *C. reinhardtii – M. aquaticum* interaction[116] (this pathway is shown as a specific example; LAO1 can oxidize other amino acids to produce different α-keto acids); (7) – possible exchange of α-keto acids produced by LAO1, which may serve diverse ecological functions, including as bacterial carbon sources, siderophores, or signaling molecules; (8) – agmatine iminohydrolase upregulation and agmatine iminohydrolase-dependent production of putrescine, which can act as a chemoattractant[251]; (9) – SynCom-dependent BAT1 upregulation and proposed amino acid exchange.

Although we observed interesting and significant changes in *C. reinhardtii* gene expression in response to the bacterial SynCom, several limitations of this study should be noted. Our analysis is based solely on transcriptomics, which reflects mRNA abundance, but not protein levels, metabolic fluxes, or enzyme activities. Thus, we cannot determine which metabolites are exchanged between *C. reinhardtii* and the SynCom. We also cannot disentangle whether the growth program shifts are a direct response to bacterial signals or an indirect consequence of altered nutrient availability in the presence of bacteria. Another limitation is that the analysis is restricted to the algal transcriptome only, and the bacterial transcriptional response remains uncharacterized. Moreover, while we observed correlations between changes in the expression of specific genes in SynCom presence, causality cannot be established without genetic perturbation.

Our list of candidate genes, focusing on *C. reinhardtii* DEGs upregulated in the SynCom presence and conserved across *C. reinhardtii*, *A. thaliana*, and *L. japonicus*, provides specific targets for functional validation, most of which can be validated using available *C. reinhardtii* mutants[32–34] (Supplementary Table S3.15; Appendix, Section 3.8). For genes in the growth preparation and translation machinery programs (Table 3.1), mutant phenotypes may be pleiotropic, thereby complicating interpretation. Therefore, priority candidates include LAO1 and agmatine iminohydrolase mutants. Comparing the growth of wild-type *C. reinhardtii*, and the bacterial communities assembled by the wild-type *C. reinhardtii* and the mutants will help us determine whether LAO1 and/or agmatine iminohydrolase are necessary for SynCom-enhanced growth or bacterial attraction. Additionally, metabolomic profiles of the wild-type *C. reinhardtii* and mutants in the LAO1 gene might be compared to elucidate which α-keto acids mediate the *C. reinhardtii*–SynCom interaction. However, metabolic flux analysis presents technical challenges, as in the gnotobiotic photobioreactor, *C. reinhardtii* and SynCom share a single compartment where metabolites are continuously produced and consumed. Systems that enable spatial

separation may help address it in the future. One of such systems is the MetaFlowTrain, a fluidic system in which microorganisms can be separated in gnotobiotic microchambers, allowing easy flushing with medium and convenient sampling[260]. However, previous research has shown that the assembly and growth of *C. reinhardtii* phycosphere microbiota beneficial to its host may require both physical proximity and bidirectional exchange of metabolites or signals, making future metabolic profiling challenging. Another interesting line of research would be the analysis of the bacterial side of the interaction, which may be initiated by analyzing genomes of the SynCom bacteria to find genes that could be important for the interaction with *C. reinhardtii*.

In summary, our study reveals that the phycosphere community triggers a coordinated shift in *C. reinhardtii* gene expression, suggesting a shift from catabolism, showing parallels to stress-induced metabolism[201–204], toward growth and biosynthesis-oriented metabolism. Parts of this response are conserved with evolutionarily distant land plants, suggesting their possible ancestral origin. The identified candidate genes provide a starting point for future functional validation using available *C. reinhardtii* mutants[32–34]. Understanding how *C. reinhardtii* interacts with its phycosphere microbiota and identifying which of these mechanisms are conserved with land plants might shed light on the evolutionary origins of interactions between photosynthetic organisms and bacteria, and inform the design of SynComs for biotechnological applications.

3.5 Materials and methods

3.5.1 *C. reinhardtii* culture conditions

C. reinhardtii CC-1690 was cultured in TP10 medium (Supplementary Table S3.13), pH 7.0, at 25 °C, and continuous illumination at 125 µmol $m^{-2}s^{-1}$ with shaking on a rotary shaker at 70 RPM[105]. Cell growth was determined either by measuring samples in a Multisizer 4e Coulter counter (Beckman Coulter Inc., California, USA) particle counter with the Beckman Coulter Multisizer software (v4.03) or using an Infinite M2000Pro (TECAN Austria GmbH, Grödig, Austria) plate reader with the TECAN i-control software (v2.0.10.0), to determine either absorbance at 750 nm or chlorophyll fluorescence used as a proxy for growth (excitation 440/9 nm, emission 680/20 nm)[105].

3.5.2 SynCom inocula preparation

SynCom strains were prepared as described before[105]. Bacterial cultures from 50% TSA agar plates incubated at 25 °C for 5 days were used to inoculate 3 sterile cultivation tubes containing 5 mL 50% TSB medium per bacterial strain. The tubes were incubated at 25 °C and 230 rpm. After 6 days, the cultures were washed three times by centrifugation at 3,000 × g for 5 min at room temperature (RT), the supernatant discarded, and the pellet resuspended in 1 mL TP10 media. The washed cultures were further incubated with shaking at 25 °C for an additional day. Bacterial concentration in the washed cultures was determined by measuring absorbance at 720 nm using the plate reader and then bacteria were pooled in equal ratios based on the absorbance. Subsequently, the cell density of the pooled SynCom was measured using the Multisizer 4e Coulter particle counter and the SynCom was used to inoculate the photobioreactor cultures. Samples for DNA extraction were taken in triplicates from the individual SynCom strain cultures and the pooled SynCom.

3.5.3 Photobioreactor experiment

C. reinhardtii cells were grown photoautotrophically in TP10 medium at 25 °C and the illumination of 80 µmol $m^{-2}s^{-1}$ under continuous light conditions axenically or in a co-culture with the SynCom[105] (*C. reinhardtii* cells:bacteria cells inoculum ratio 1:100, culture volume 85 mL) in a Multi Cultivator MC1000-OD photobioreactor

(PSI (Photon Systems Instruments) spol. s r.o.), with an aeration air speed set to 50% (Multi Cultivator MC1000-OD photobioreactor manufacturer setting). Cell growth was determined by measuring samples in a Multisizer 4e Coulter particle counter. Sampling timepoints were chosen based on optical density (OD 680 nm, OD 720 nm) measurements taken in the photobioreactor (OD 720 nm exponential phase threshold 0.5, OD 720 nm stationary phase threshold 1.2, determined empirically in the past). For each sampling timepoint, samples for 16S rRNA profiling, RNA-Seq, chlorophyll fluorescence determination, and culture absorbance and fluorescence measurements were taken.

3.5.4 Chlorophyll extraction and fluorescence determination

From the photobioreactor samples, chlorophyll was extracted as follows: 100 µL of the culture was added to 400 µL 100% ethanol, incubated at RT for 2 min, and centrifuged at 17,000 × g for 1 min (RT). 150 µL of the supernatant was recovered, and absorbance at 665 nm and 649 nm was measured[261]. Chlorophyll a + Chlorophyll b concentration in µg mL^{-1} was then calculated from $Chla+b = 6.10A_{665} + 20.04A_{649}$[262].

3.5.5 DNA extraction

Bacterial DNA was extracted using alkaline lysis[105,147]. Briefly, 12 µL of the sample (either pure bacterial culture, SynCom mix or co-culture of *C. reinhardtii* with the SynCom) were diluted in 20 µL of Buffer I (NaOH 25 mM, EDTA(Na) 0.2 mM, pH 12), mixed by pipetting, and incubated at 94 °C for 30 min. Next, 20 µL of Buffer II (Tris-HCl 40 mM, pH 7.46) was added to the mixture and stored at −20 °C.

3.5.6 16S rRNA gene amplicon sequencing

16S rRNA profiling data was analyzed as described previously[105]. Briefly, the V5–V7 fragments of the 16S rRNA gene from the extracted DNA were amplified in a two-step PCR amplification using primers 799F (AACMGGATTAGATACCCKG) and 1192R (ACGTCATCCCCACCTTCC)[105,172,173], followed by indexing of the PCR products using Illumina-barcoded primers. In the first step, V5–V7 region of the bacterial 16S rRNA was amplified in a 25 µL reaction volume containing 2 U Taq DNA Polymerase, 1× incomplete buffer, 2 mM $MgCl_2$ (Bioron GmbH, Ludwigshafen, Germany), 0.3% BSA, 0.2 mM dNTPs (ThermoFisher Scientific) and 0.3 µM primers 799F[172] and

1192R[173]. PCR was performed (initial denaturation at 94 °C/2 min, followed by 25 cycles of 94 °C/30 s, 55 °C/30 s, 72 °C/60 s, and final elongation at 72 °C/5 min). 3 µL of the supernatant was used as the DNA template for a second PCR, prepared in the same way as described but with primers including barcodes and Illumina adapters, and cycles reduced to 10. 5 µL of each PCR product were loaded on a 1% agarose gel to assess the band intensities and the PCR products were pooled in similar ratios based on the visual assessment of the bands intensities. The pooled sample was loaded on 1.5% agarose gel and run for 40 min at 90 V (small gel), the bands were excised and DNA was purified using the Gel and PCR-Clean-up kit (QIAquick PCR & Gel Cleanup Kit, QIAGEN, Hilden, Germany). DNA concentration was determined from fluorescence measurements with the QuantiFluor® dsDNA System (Promega GmbH) according to the manufacturer's protocol. Paired-end Illumina sequencing was performed in-house using the MiSeq sequencer and custom sequencing primers[105].

3.5.7 16S rRNA amplicon profiling analysis

Amplicon sequencing data were processed using the QIIME2 2021.11 pipeline[174]. Briefly, data were imported into QIIME2, demultiplexed according to their barcode sequences and trimmed. Paired-end sequencing reads were first merged using FLASH2 (v2.2.00)[263], combining forward and reverse reads for each sample with a maximum overlap length of 250 and mismatch ratio of 0.25. Merging was performed on demultiplexed FASTQ files, and output files containing successfully merged reads were retained for downstream analyses. Intermediate histogram files and reads that could not be merged were discarded. Merged reads were subjected to quality filtering to remove sequences containing ambiguous nucleotides. Specifically, reads with one or more N bases were excluded using USEARCH (v10.0.240_i86linux32)[264] with a maximum allowed number of ambiguous bases set to zero. Only high-quality merged reads passing this filter were retained for further processing. Quality-filtered merged paired-end reads were then aligned to a reference set of sequences extracted from the whole-genome assemblies of every strain included in the SynCom[105], using Rbec (v1.0.0)[265]. In the end, we generated a count table that was used for downstream analyses in R (v4.4.0)[179] with the package *vegan*[183].

3.5.8 RNA extraction

Samples for RNA-sequencing (RNA-Seq) taken from the photobioreactor cultures (1×10^7 cells, based on the measurements with the Multisizer 4e Coulter particle counter) were spun down at 3,202 × g (8 min, RT) and resuspended in 1 mL DNA/RNA Shield (Quick-DNA/RNA Miniprep Plus Kit, D7003, Zymo Research, USA). The resuspended samples were transferred to Lysing Matrix E tubes (MP Biomedicals), snap frozen in liquid nitrogen, and stored at −80 °C until the RNA was isolated using the Quick-DNA/RNA Miniprep Plus Kit (D7003, Zymo Research, USA) according to the manufacturer's instructions, including DNase and proteinase K treatments. The assessment of the isolated RNA quality on the Bioanalyzer 2100 system (Agilent Technologies, CA, USA), further library preparation, and transcriptome sequencing (paired-end, 150 bp) were performed by Novogene (Cambridge, UK), yielding 51.7–116.9 million raw reads per sample with >97% Q20 and >92% Q30 quality scores.

3.5.9 Transcriptomics data analysis

Transcript expressions were quantified using Salmon (v.1.9.0)[266] with default parameters, using an index built on a reference *C. reinhardtii* transcriptome from the Chlamydomonas Genome project version 6[29]. Count data were imported into R (v4.4.0)[179] using the *tximport*[267] package, and differential gene expression analysis was performed with *DESeq2*[267,268]. Data were pre-filtered to keep only transcripts with at least 10 total reads. A generalized linear model was fitted using the following design: Condition + Timepoint + Condition:Timepoint, where Condition was the presence or absence of the SynCom (co-cultures with the SynCom or axenic *C. reinhardtii* cultures), and Timepoint referred to the sampling timepoint – *C. reinhardtii* growth phase (exponential or stationary). Exponential phase of growth and axenic *C. reinhardtii* cultures were used as reference factor levels. Wald tests were used to assess significance for all contrasts, and shrinkage of effect size was performed using the *apeglm* method to reduce bias for lowly expressed genes[269]. Differentially expressed genes (DEGs) were defined as genes with $\log_2FC \neq 0$ and Benjamini-Hochberg adjusted *P*-value < 0.05. From the fitted model, the following biologically relevant contrasts were extracted: co-culture with SynCom vs axenic culture in exponential phase (CondSCvsA_exp) – effect of the SynCom presence during

exponential growth, referred to as SynCom vs axenic (exponential phase);; co-culture with SynCom vs axenic culture in stationary phase (CondSCvsA_stat) – effect of the SynCom presence during stationary growth, referred to as SynCom vs axenic (stationary phase); stationary vs exponential phase in the axenic *C. reinhardtii* cultures (timeAxenicSvsE) – timepoint-dependent expression changes in the axenic *C. reinhardtii* cultures; stationary vs exponential phase in the *C. reinhardtii* co-cultures with the SynCom (timeSCSvsE) – timepoint-dependent expression changes in the *C. reinhardtii* co-cultures with the SynCom; Condition × Timepoint interaction (timepointEffect_SCvsA) – genes for which the effect of timepoint differs between the co-cultures with the SynCom and axenic cultures. The distribution of unadjusted and adjusted *P*–values was examined to assess model behavior. MA plots were generated before and after $\log_2$FC shrinkage. The $\log_2$-scaled counts were subjected to *varianceStabilizingTransformation* and transformed as median-centered *z*-score using the function *scale*[268]. *ComplexHeatmap* package was used to visualize the DEGs[270,271]. Genes were considered to be DEGs if they were significant in at least one of the defined contrasts. DEG lists from all contrasts were combined to generate a unified set of SynCom- and timepoint-responsive genes, which served as the basis for downstream functional annotation, enrichment analysis, and candidate gene selection.

3.5.10 Gene Ontology annotation and integration

For Gene Ontology analysis, combined data from Phytozome[29] and extended GO data generated using the Gene Ontology Meta Annotator for Plants (GOMAP)[188] pipeline were used. To combine the two annotations, the Phytozome-derived annotation table corresponding to the reference transcriptome was imported into R (v4.4.0)[179] and only gene identifiers and associated GO terms were retained. GO annotations generated in the GOMAP[188] pipeline were imported from the aggregated Gene Annotation File (GAF), containing predictions from multiple annotation sources, including Argot2.5, FANN-GO, InterProScan5, PANNZER, UniProt, and Arabidopsis TAIR annotations implemented in GOMAP[188]. Gene identifiers, GO term accessions, and annotation sources were retained. When a gene–GO term association was supported by multiple annotation sources, each source was recorded as a separate entry. Phytozome and GOMAP annotations were combined into a single table, and duplicate gene–GO term

associations were removed. The final integrated annotation set represents the union of curated (Phytozome) and computationally inferred (GOMAP) GO annotations. For downstream analyses, GO terms associated with each gene were collapsed into space-delimited strings and exported as a gene-to-GO mapping file. The shared and unique gene–GO associations from both Phytozome and GOMAP annotations were visualized using UpSet plots.

Functional enrichment analysis was conducted using R package *topGO*[272] using the combined annotations. All genes included in the transcriptomics dataset were used as the background gene universe. GO enrichment analyses were conducted separately for the Biological Process (BP) and Molecular Function (MF) ontologies for the DEGs in each of the comparisons and in each expression cluster. Enrichment testing was performed using the classic Fisher's exact test implemented in topGO, and *P*-values obtained from Fisher's exact tests were corrected for multiple testing using the Benjamini–Hochberg false discovery rate (FDR) procedure[272]. GO terms with an adjusted FDR below 0.1 were considered significantly enriched. For each significantly enriched GO term, the list of contributing genes was extracted and saved for downstream interpretation and visualization.

To reduce redundancy among enriched GO terms and improve interpretability, we decided to perform semantic similarity-based GO term reduction. First, a custom organism annotation database (OrgDb) for *C. reinhardtii* was constructed using the *AnnotationForge*[273] package. The combined gene-to-GO annotation mapping generated in this study was converted into standard AnnotationForge input tables, including a gene–GO association table, gene identifier and symbol mappings and chromosome-level annotations inferred from gene identifiers. The resulting annotation package (org.Creinhardtii.eg.db) was built, installed locally and used as the reference annotation database for all semantic similarity calculations. Semantic similarity-based GO term reduction was performed using the *GOSemSim*[274] and *rrvgo*[275] packages. Semantic similarity matrices were computed separately for the BP and MF ontologies using the Relevance (Rel) semantic similarity measure[276] and the prepared org.Creinhardtii.eg.db. GO terms were clustered based on semantic similarity for each comparison and expression cluster, and representative parent terms were selected using a similarity threshold of 0.7. GO terms were scored using

the negative logarithm of the adjusted *P*-value (–log10 FDR), such that more statistically significant terms were preferentially retained as representatives. GO terms that could not be merged due to insufficient similarity or missing semantic information were retained as individual terms. Results of the GO BP and MF enrichment analyses were visualized using bubble plots in which bubble color corresponds to the minimum adjusted *P*-value (FDR) of the terms assigned to the reduced term, and bubble size represents the maximum number of annotated significant genes per each reduced GO term. For visualization purposes, the reduced GO terms were further manually grouped into higher-level functional categories (e.g. gene expression, metabolic processes) to facilitate interpretation.

3.5.11 KEGG pathway analysis

For analysis of the KEGG pathways enrichment, KEGG Orthology (KO) annotations from Phytozome[29] for *C. reinhardtii* were combined with extended data obtained by performing annotations with available tools: BlastKOALA, GhostKOALA[194], KofamKOALA[196], and eggNOG[195]. The annotations were combined in the following way: if a gene had Phytozome annotation, then it was retained; if not, the KO term appearing most frequently among the annotations from the different sources was chosen. If different KO terms were present for different tools and appeared only once, then the final KO term was chosen with the following priority: KofamKOALA[196], BlastKOALA, GhostKOALA[194] and eggNOG[195]. The resulting merged KO annotation table was used for all downstream KEGG analyses. Gene Set Enrichment Analysis (GSEA) was conducted in R (v4.4.0)[179] using packages *GAGE*[277] and *clusterProfiler*[278,279], but only *clusterProfiler* GSEA results were used for downstream analyses and visualizations, since GSEA consistently identified a larger number of enriched pathways across experimental contrasts and produced enrichment patterns that were more coherent and interpretable in the context of coordinated transcriptional regulation, while GAGE yielded fewer significantly enriched pathways. Given that KEGG pathways often comprise genes with mixed regulatory directions, the ranked, permutation-based framework of GSEA was better suited to capture biologically meaningful pathway-level trends in this dataset. Therefore, GSEA results computed using *clusterProfiler*[278,279] were used for final interpretation, visualization, and reporting.

To improve biological interpretability, KEGG results were curated to exclude pathways not relevant to *C. reinhardtii* biology, such as human– or fly–related (e.g. "Digestive system", "Phototransduction – fly") (Supplementary Table S3.16). For each comparison, GSEA results were visualized using bubble plots at two hierarchical KEGG levels: pathway level, representing individual KEGG pathways, and level B (functional category) level, representing broader functional groupings. In bubble plots, the x-axis represents the normalized enrichment score, the y-axis represents KEGG pathway or functional category names, bubble size represents gene set size, and bubble color represents enrichment score magnitude and direction. Plots were faceted by experimental comparison to enable direct visual comparison across conditions.

3.5.12 KEGG pathway visualization

Identified enriched pathways and several additional pathways identified as meaningful based on the analysis of the transcriptomics data were visualized using the *pathview*[280] package. Gene-level $\log_2$FC values were projected onto the curated KEGG Orthology pathways. If multiple genes mapped to the same KO identifier, expression values were averaged by KO, as implemented in *pathview*[280], to represent a pathway element for visualization. For each experimental comparison, differential expression values mapped to KO identifiers were overlaid onto KEGG native pathway diagrams using the global KEGG reference. Pathway visualization was performed using a fixed $\log_2$FC scale (−1 to +1) to ensure comparability across conditions. Mapped pathway elements were colored using a diverging color scale to indicate up- (orange) and downregulation (teal), while pathway elements without expression data were displayed in grey. Selected pathways were visualized for supplementary figures to illustrate transcriptional regulation patterns within biologically relevant pathways.

3.5.13 Identification of Reciprocal Best Hits between *C. reinhardtii*, *A. thaliana* and *L. japonicus* genes

Protein and transcript sequence data were obtained from Phytozome for *C. reinhardtii*[29] and *A. thaliana* (Araport11)[281], and from *Lotus* Base for *L. japonicus* (Gifu 1.3)[282]. The analysis was performed at the protein level using primary (representative) isoforms. Because a curated primary proteome was not available for *L. japonicus*, a representative primary proteome was constructed. For each gene, the

".1" isoform was selected when present; if absent, the first available isoform was used as a fallback. Protein headers were parsed to identify isoform structure, and a gene-to-primary-isoform mapping table was generated. Corresponding protein sequences were extracted using *seqkit* (v0.10.1)[175]. The resulting FASTA file constituted the *L. japonicus* primary proteome used for the analysis. Next, protein BLAST databases were constructed for each species using makeblastdb (BLAST v2.12.0, q-dbtype prot)[169,170]. Databases were generated from the following primary proteomes: *C. reinhardtii*, *A. thaliana* and *L. japonicus*. Pairwise BLASTP searches were performed between species to identify candidate orthologs. Searches were conducted in both directions for each species pair to enable Reciprocal Best Hit (RBH) inference (e.g., *C. reinhardtii*, → *A. thaliana* and *A. thaliana* → *C. reinhardtii*). BLASTP (v2.12.0)[169,170] was executed on a high-performance computing cluster using SLURM job arrays, allowing parallel execution of all query–database combinations. Each BLASTP run used the following parameters: maximum of one target sequence per query (-max_target_seqs 1), e-value cutoff of 10, with tabular output format (outfmt 6) including query and subject IDs, alignment length, percent identity, bitscore, and e-value. All protein sequences were queried (not only DEGs) to ensure unbiased identification of best hits. The best hits lists were imported into R (v4.4.0)[179] as tab-delimited tables and gene identifiers were standardized across species by removing isoform suffixes and species-specific annotations to ensure consistent gene-level comparisons. For each BLAST hit, query coverage and subject coverage were calculated as query coverage = alignment length / query length and subject coverage = alignment length / subject length to assess the extent of sequence overlap and to reduce short local alignments. To assess the robustness of RBH inference, RBH identification was repeated across a grid of different filtering thresholds of the minimum query coverage, minimum subject coverage, minimum percent identity, maximum e-value and minimum bitscore. For each threshold combination, BLAST results were filtered, RBHs were identified, and the numbers of pairwise and three-way RBHs were assessed, allowing evaluation of how stringency affected RBH identification. Based on threshold exploration to balance specificity and sensitivity, the following criteria were selected: minimum query coverage 0.5, minimum subject coverage 0.5, minimum percent identity 30%, maximum e-value 1×10^{-5}, and minimum bitscore 80. RBHs were defined as gene pairs that were mutually identified

as best hits in both BLAST directions after filtering, and were computed independently for each species pair: *C. reinhardtii – A. thaliana*, *C. reinhardtii – L. japonicus*, and *A. thaliana – L. japonicus*. Genes were classified into conservation categories based on RBH patterns across the three species: conserved in all three organisms, conserved between two organisms only, RBHs present but not identical across species, or not conserved. For the analysis of conservation of SynCom-responsive DEGs among species, DEGs lists were imported. SynCom-responsive DEGs for *C. reinhardtii* were derived from this study's RNA-seq analysis and included genes significantly regulated in response to SynCom presence or time-dependent SynCom effects. Lists of SynCom-responsive DEGs for *A. thaliana* and *L. japonicus* were obtained from Wippel *et al.* (2021)[84]. Gene conservation patterns were summarized and visualized for all genes and for DEGs separately using three-way presence–absence tables based on RBH relationships. *C. reinhardtii* DEGs were further divided into upregulated and downregulated gene sets, and each set was compared with *A. thaliana* and *L. japonicus* data to determine the consistency of transcriptional regulation direction. Genes were classified as consistently regulated (same direction in all species with RBHs), inconsistently regulated, or not differentially expressed in one or more species. Barplots were generated to quantify the proportion of upregulated and downregulated *C. reinhardtii* genes exhibiting any degree of conservation, as well as those showing conserved regulation direction across species. To identify candidate genes, RBHs between *C. reinhardtii*, *A. thaliana,* and *L. japonicus* were annotated using previously defined plant gene expression clusters[84]. These clusters summarize transcriptional responses in the corresponding plant species and were assigned to RBH genes according to the gene. Cluster descriptions were simplified into categorical labels ("Upregulated in native SynCom presence", "Upregulated in non-native SynCom presence", "Upregulated in native and non-native SynCom presence", "Downregulated in native SynCom presence", "Downregulated in non-native SynCom presence", "Downregulated in native and non-native SynCom presence") to facilitate visualization. Binary matrices were constructed to represent the presence or absence of RBHs between *C. reinhardtii* genes and each plant species, with rows corresponding to *C. reinhardtii* genes, and columns corresponding to RBH relationships with *A. thaliana* or *L. japonicus*. The matrices were visualized using the *ComplexHeatmap*[270] package. Rows were hierarchically clustered based on RBH

patterns, while columns were not clustered. Heatmap color scales encoded RBH presence or absence. Row annotations were added to indicate the functional cluster assignments of the corresponding *Arabidopsis* and *Lotus* RBHs.

3.5.14 Candidate gene selection

C. reinhardtii DEGs were functionally annotated by integrating Phytozome annotation[29], combined Phytozome and GOMAP[188] GO annotations, extended KEGG annotations, results of *k*-means clustering, GO BP and MF enrichment analyses, KEGG GSEA analysis, and evolutionary conservation (RBH) analysis. All annotation layers were integrated into a single comprehensive DEGs annotation table. Candidate genes were selected using a multi-criteria scoring framework that prioritizes DEGs that are transcriptionally responsive to SynCom conditions, evolutionarily conserved, and functionally supported by enrichment analyses. Analyses focused on DEGs upregulated in *C. reinhardtii* in the co-culture with the SynCom, defined as genes significantly upregulated in either the exponential or the stationary phase when comparing the co-cultures with the SynCom and axenic cultures. To facilitate narrowing down the list of candidate genes, each DEG was assigned a score based on the following criteria (each worth one point): presence of an RBH in *A. thaliana*, presence of an RBH in *L. japonicus*, upregulation of the RBH gene in *A. thaliana*, upregulation of the RBH gene in *L. japonicus*, induction of the RBH gene in response to the native SynComs in *A. thaliana*, induction of the RBH gene in response to the native SynComs in *L. japonicus*, membership in enriched KEGG pathways in *C. reinhardtii*, membership in enriched GO BP terms in *C. reinhardtii*, membership in enriched GO MF terms in *C. reinhardtii*. From the scored gene set, candidates conserved across all three organisms (*C. reinhardtii*, *A. thaliana,* and *L. japonicus*) and upregulated in response to the SynComs in all species were selected for further refinement. Finally, a second filtering step was applied to define a narrow set of high-confidence candidates. DEGs were retained if they met at least one of the following criteria: (1) $\log_2$FC ≥ 1.25 in exponential or stationary phase in *C. reinhardtii* and upregulated in response to bacteria in both *A. thaliana* and *L. japonicus*, or (2) coordinated induction in response to the native SynComs in both *A. thaliana* and *L. japonicus* species and $\log_2$FC ≥ 0.5 in *C. reinhardtii* in either exponential or stationary phase of growth. The $\log_2$FC ≥ 1.25 was chosen to balance stringency with

sufficient gene recovery. Additionally, one gene (Cre12.g551352, LAO1) was chosen based on its very high upregulation in the SynCom presence in the stationary phase (log_2FC = 10.2) and prior evidence for its role in *C. reinhardtii*–bacteria interactions[116].

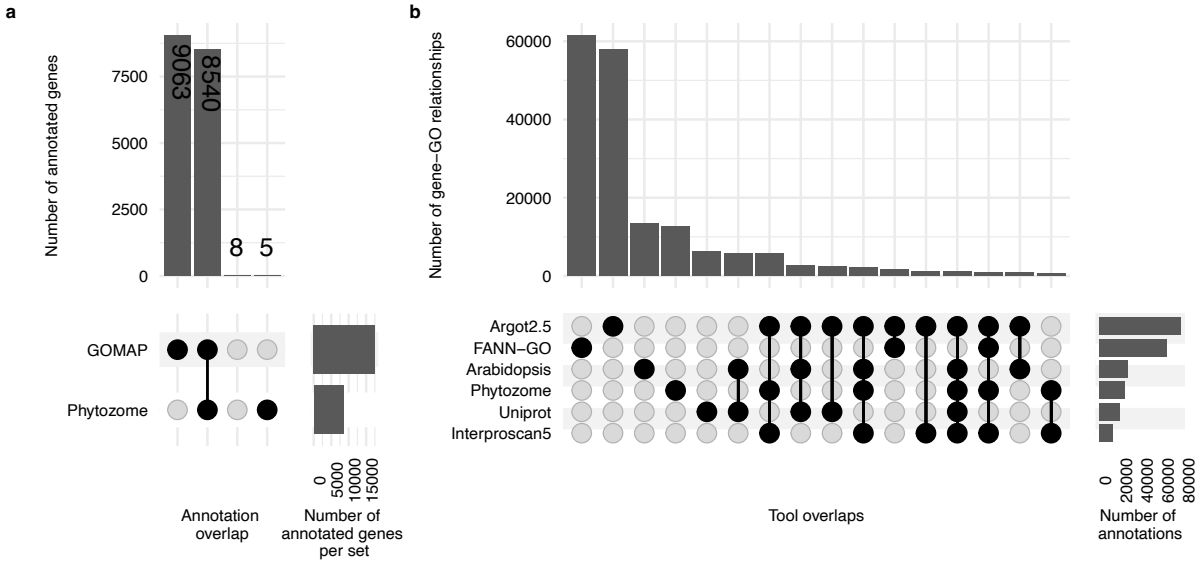
3.5.15 Transparency and AI use

Generative AI models (ChatGPT by OpenAI and Claude by Anthropic) were used to assist in drafting code templates, refining analytical workflows, language refinement, and improving textual explanations. The models did not independently generate results or draw conclusions. All AI-generated content was reviewed, verified, and edited by the author. All AI-generated content was reviewed, verified, and edited by the author. All analyses were designed, executed, and interpreted by the authors. The author retains full responsibility for the accuracy, originality, and integrity of the work.

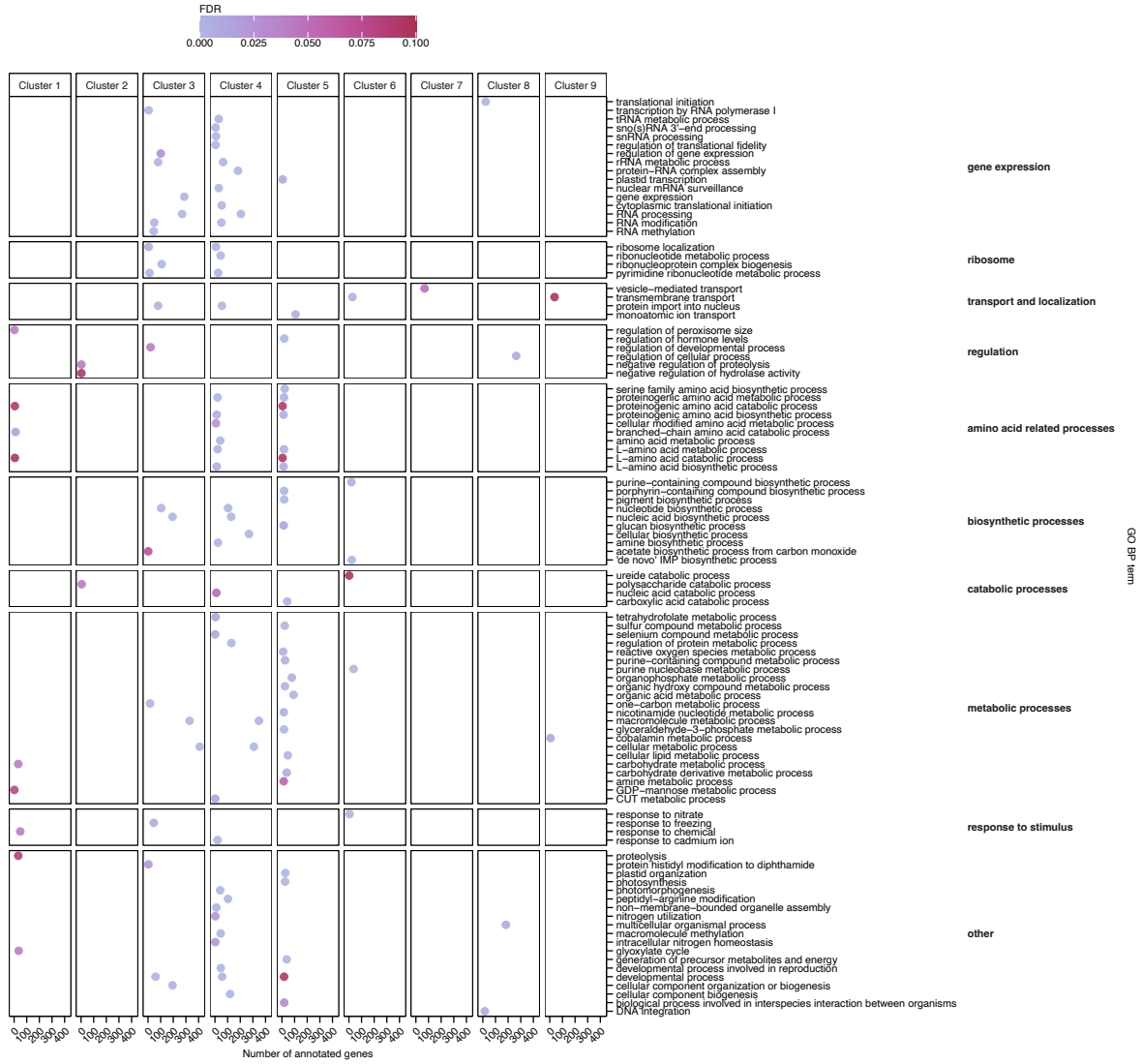
3.5.16 Figures and illustrations

Data plots were generated using R (v4.4.0)[179] with the packages *ggplot2*[283], *ComplexHeatmap*[270], and *pathview*[280]. Illustrations were created using BioRender.

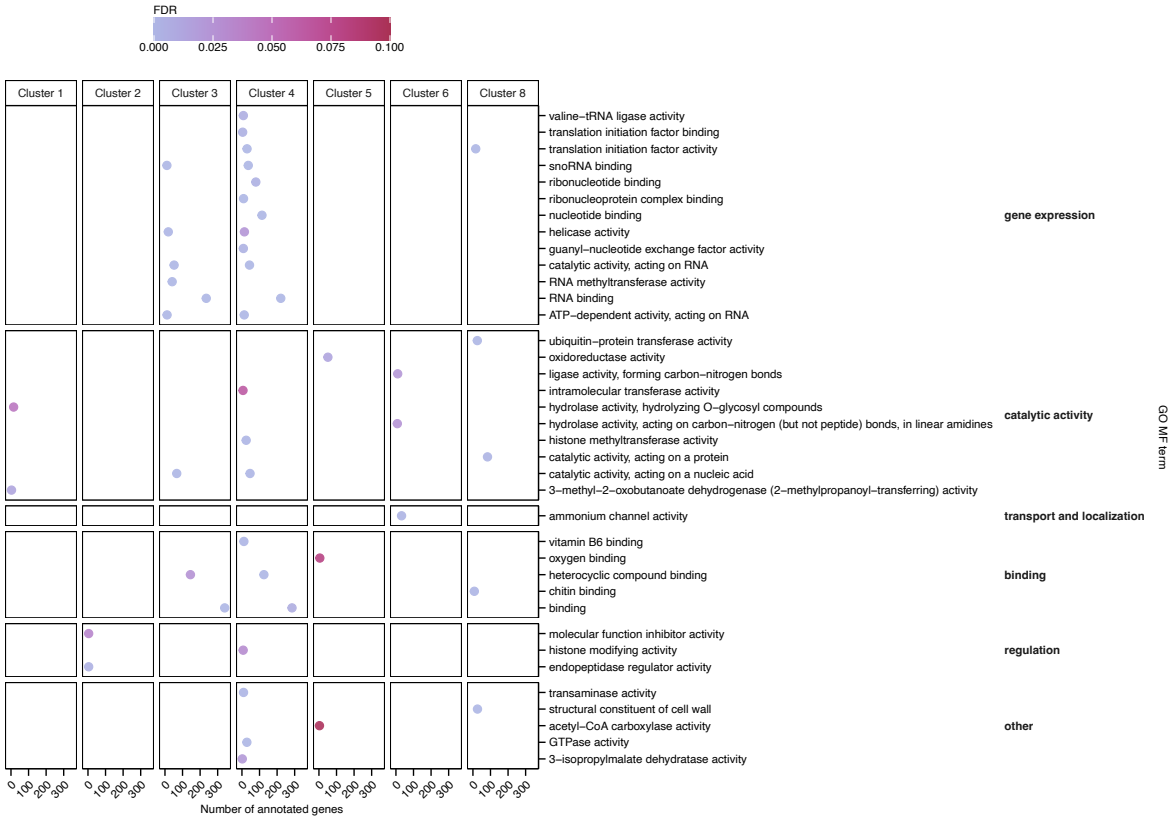
3.6 Supplementary Figures



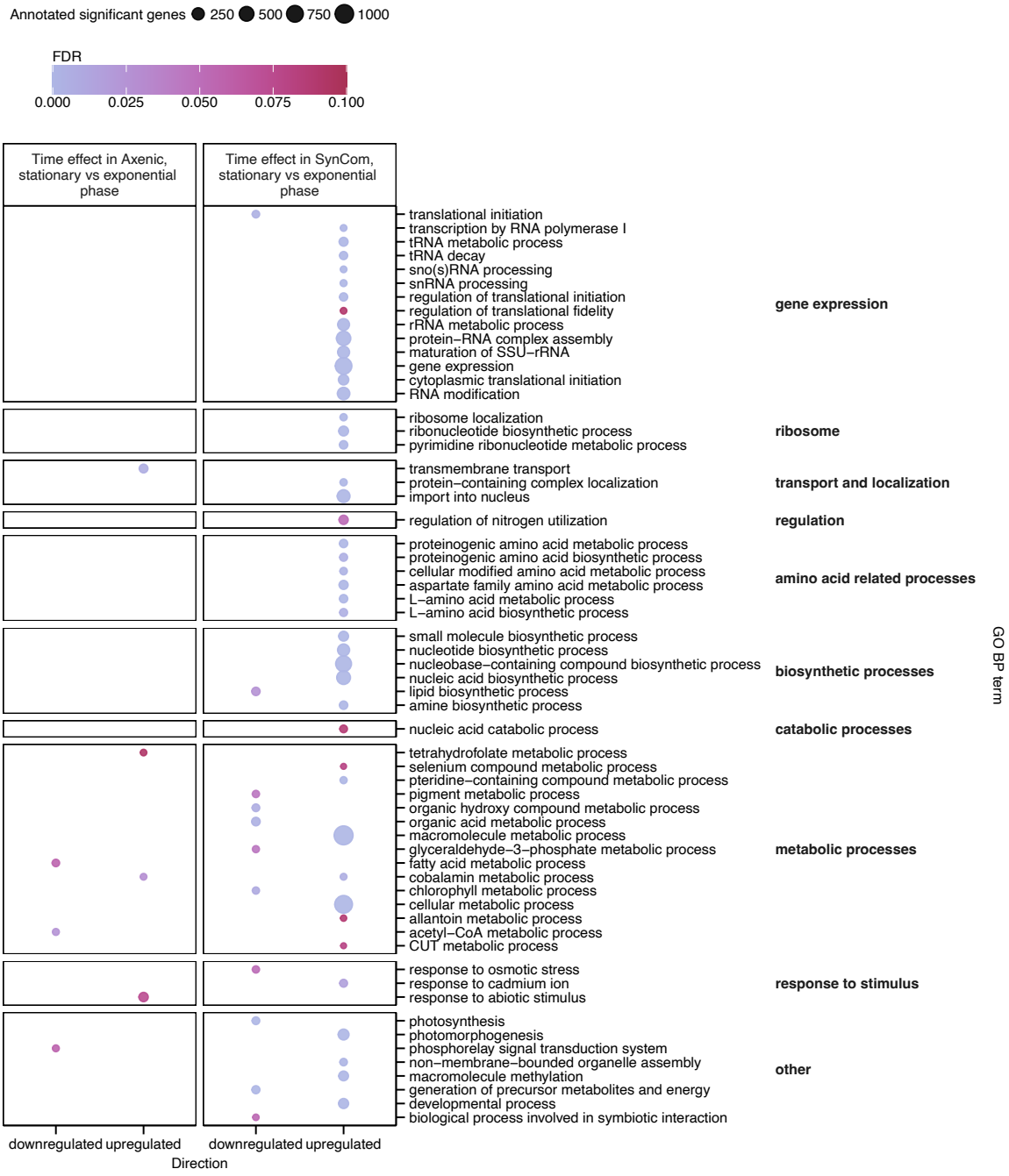
Supplementary Figure S3.1 Extended Gene Ontology annotation of *C. reinhardtii* genes. a – UpSet plot showing the overlap of GO annotations from Phytozome and GOMAP. Vertical bars represent the number of genes with annotations unique to either source or shared between them and the horizontal bars show the number of annotated genes per source. b – UpSet plot showing overlap of GO annotations across GOMAP tools. Vertical bars represent the number of gene-GO term relationships annotated by each combination of tools, and the horizontal bars on the right indicate the total number of annotations by each tool.



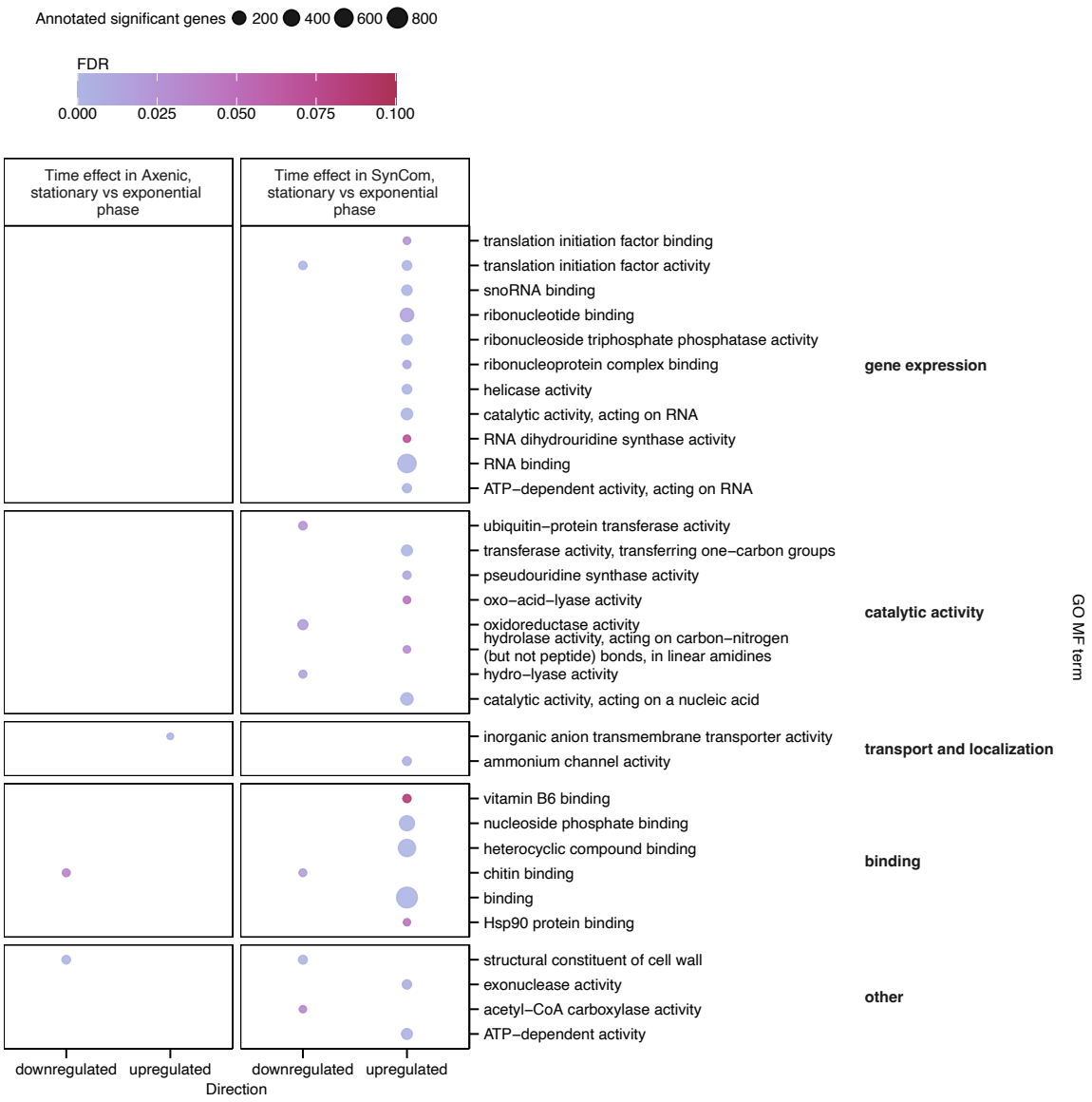
Supplementary Figure S3.2 Gene Ontology Biological Process enrichment for *C. reinhardtii* in different expression clusters. Bubble plot showing GO BP terms enriched among DEGs belonging to different expression clusters. Terms were semantically clustered and reduced, and representative parent terms are displayed. Bubble color indicates adjusted *P*-values (FDR).



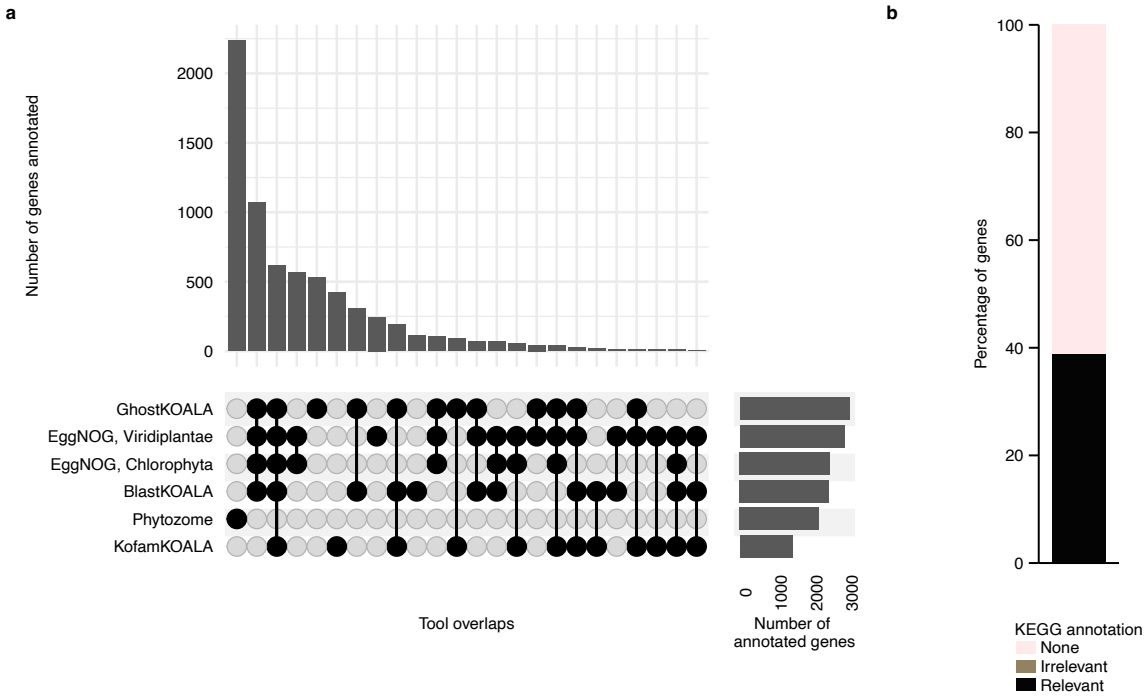
Supplementary Figure S3.3 Gene Ontology Molecular Function enrichment for *C. reinhardtii* in different expression clusters. Bubble plot showing GO MF terms enriched among DEGs belonging to different expression clusters. Terms were semantically clustered and reduced, and representative parent terms are displayed. Bubble color indicates adjusted *P*-values (FDR).



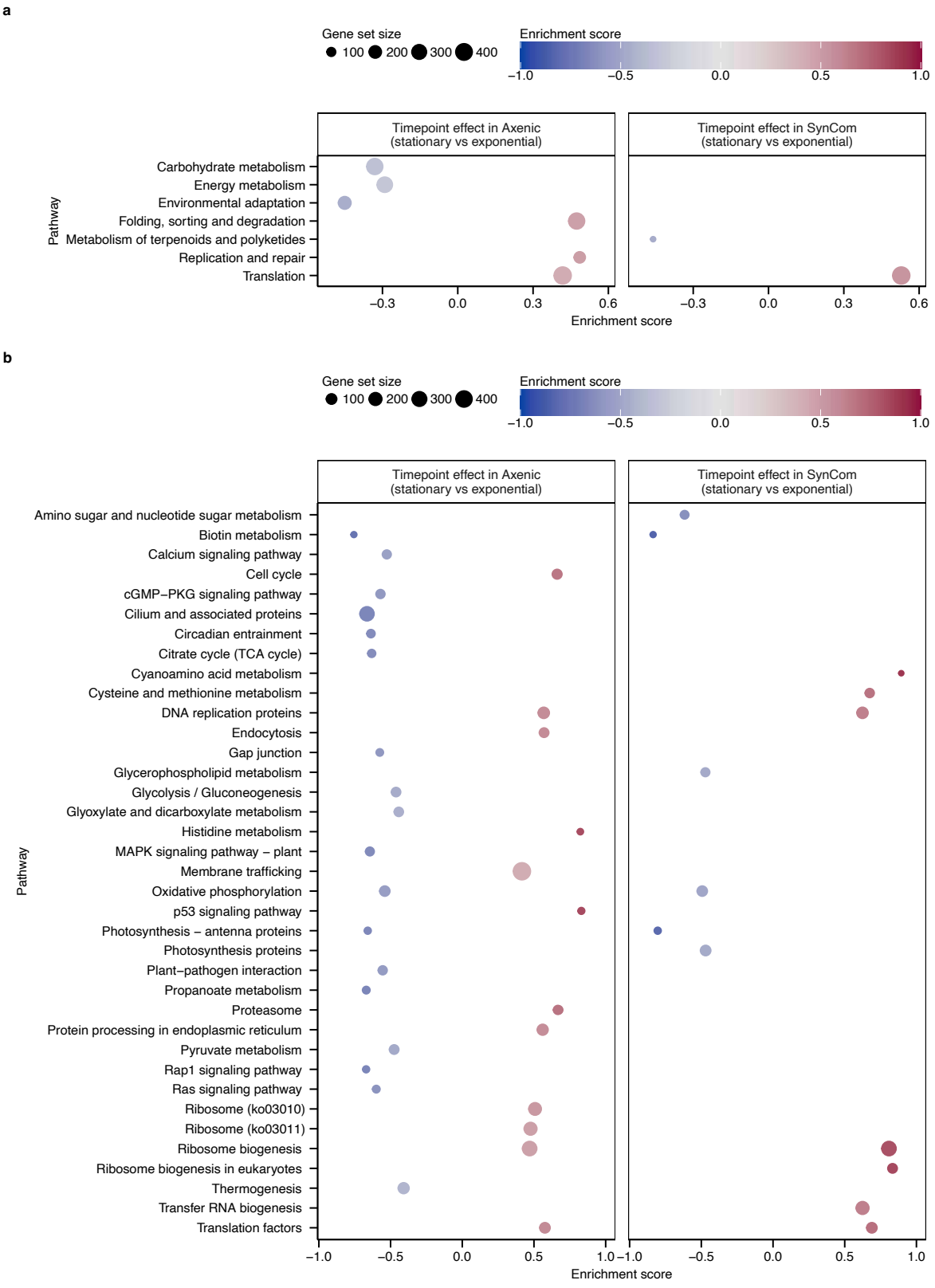
Supplementary Figure S3.4 Gene Ontology Biological Process enrichment for *C. reinhardtii* in the stationary vs exponential phase in axenic cultures and in the co-cultures with SynCom. Bubble plot showing GO BP terms enriched among *C. reinhardtii* down- and upregulated DEGs, comparing the stationary growth phase to the exponential phase in the axenic control and in the co-culture with SynCom. Terms were semantically clustered and reduced, and representative parent terms are displayed. Bubble size denotes the number of significant genes annotated to each term, and color indicates adjusted *P*-values (FDR).



Supplementary Figure S3.5 Gene Ontology Molecular Function enrichment for *C. reinhardtii* in the stationary vs exponential phase in axenic cultures and in the co-cultures with SynCom. Bubble plot showing GO MF terms enriched among *C. reinhardtii* down- and upregulated DEGs, comparing the stationary growth phase to the exponential phase in the axenic control and in the co-culture with SynCom. Terms were semantically clustered and reduced, and representative parent terms are displayed. Bubble size denotes the number of significant genes annotated to each term, and color indicates adjusted *P*-values (FDR).

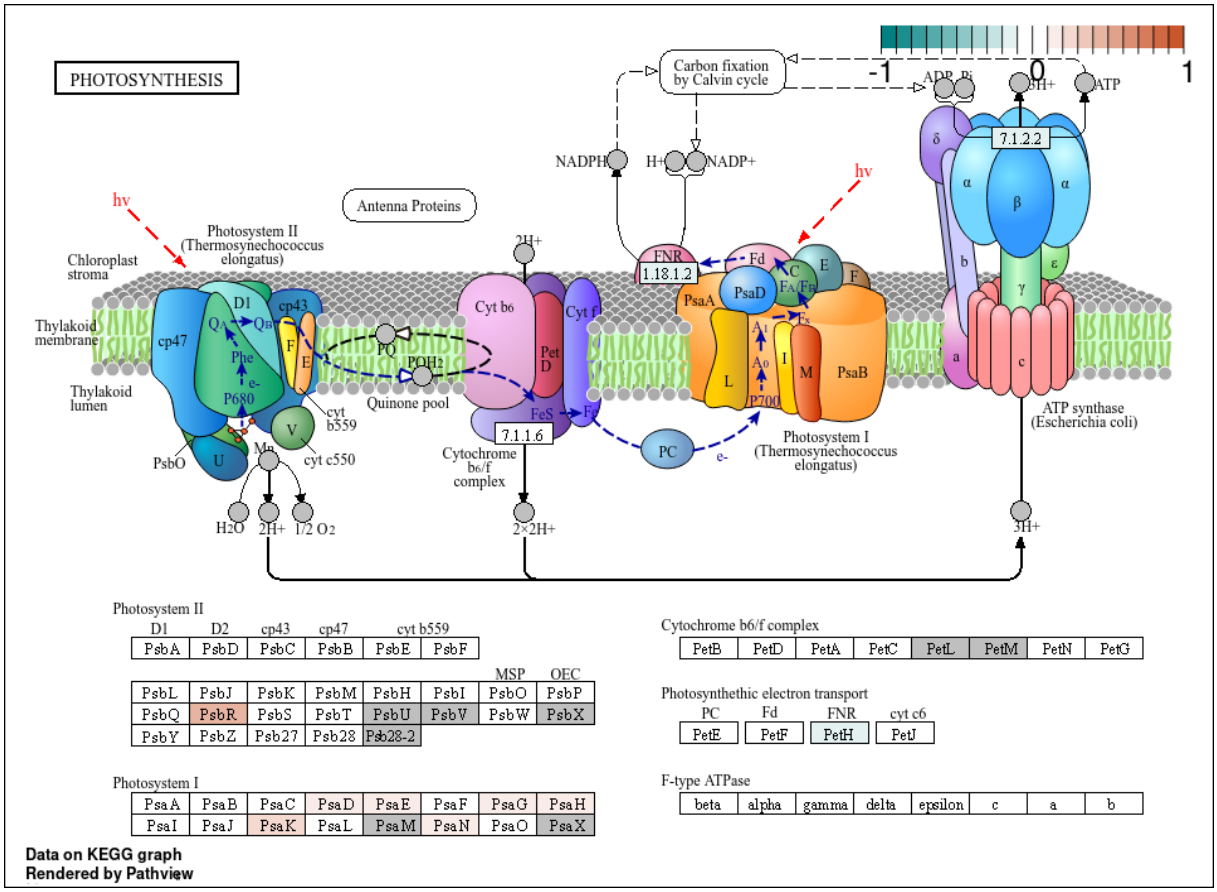


Supplementary Figure S3.6 Extended KEGG annotation of *C. reinhardtii* genes. a – UpSet plot showing the number of genes with KEGG annotations from Phytozome (not considered for annotation extension) and the overlap of KEGG annotations from different tools. Vertical bars represent the number of genes annotated by different combinations of tools or Phytozome alone, and the horizontal bars on the right indicate the total number of annotations by each tool. b – Barplot showing the percentage of genes with relevant KEGG annotations, irrelevant KEGG annotations (e.g. animal-specific, below 1% of cases), or no KEGG annotations.

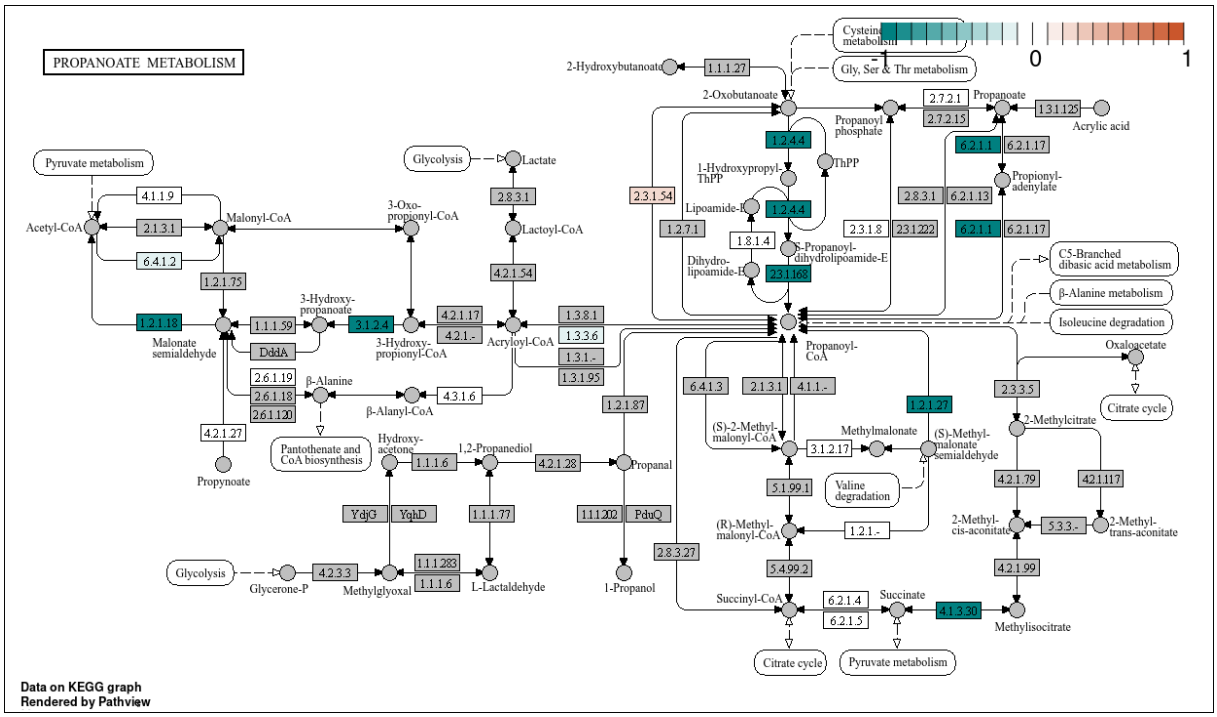


Supplementary Figure S3.7 Curated KEGG pathway enrichment for *C. reinhardtii* in the stationary vs exponential phase in axenic cultures and in the co-cultures with SynCom. a – GSEA at the KEGG level B category for *C. reinhardtii* grown in the co-culture with SynCom relative to axenic controls in exponential and stationary phases. b – GSEA at the KEGG pathway level for the same contrasts. Both

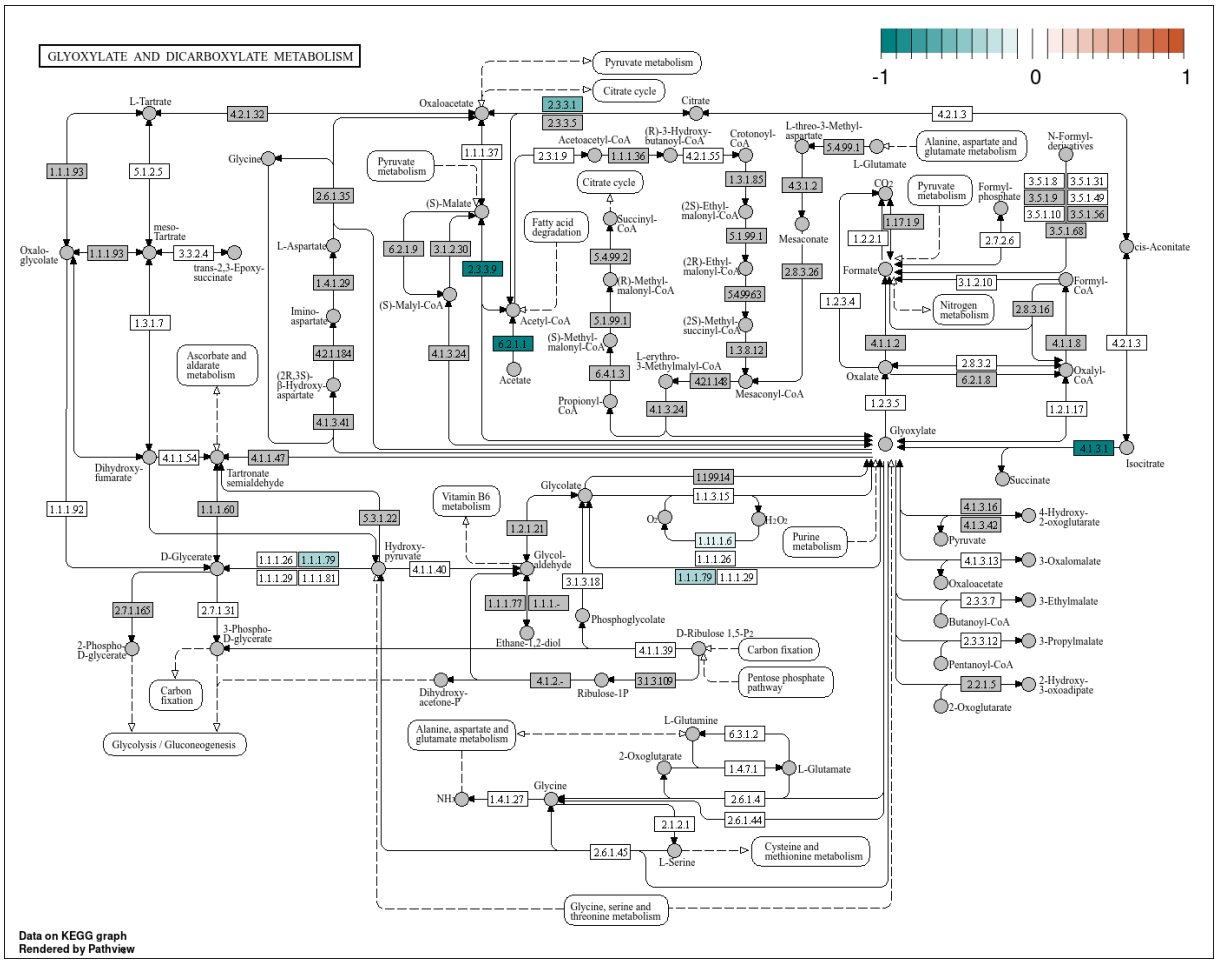
panels are faceted by contrasts: time effect in axenic controls and in the co-cultures with SynCom. Pathways were curated to exclude i.e. animal-specific processes. Points are sized according to gene set sizes and colored by enrichment scores, with shades of blue indicating downregulation of a process or pathway, and red – upregulation.



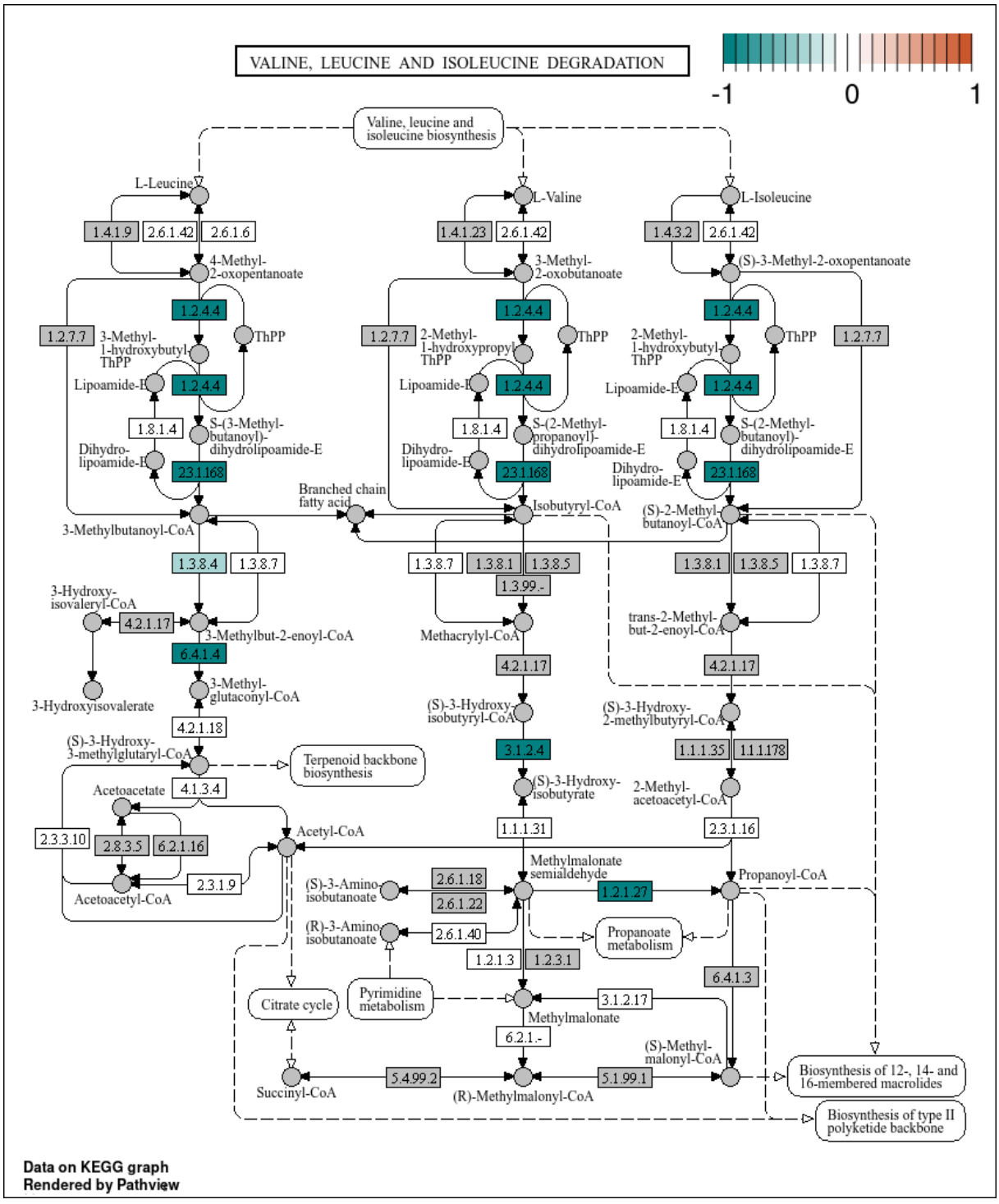
Supplementary Figure S3.8 KEGG pathway visualization: Photosynthesis. Differential expression (log_2FC) in *C. reinhardtii* co-cultured with SynCom versus cultured axenically in the exponential phase, mapped to KEGG Orthology (KO) identifiers. When multiple *C. reinhardtii* genes share the same KO annotation, their log_2FC values are averaged. Pathways were visualized using a fixed log_2FC scale (−1 to +1); orange indicates upregulation and teal indicates downregulation. Grey indicates pathway elements without mapped expression data.



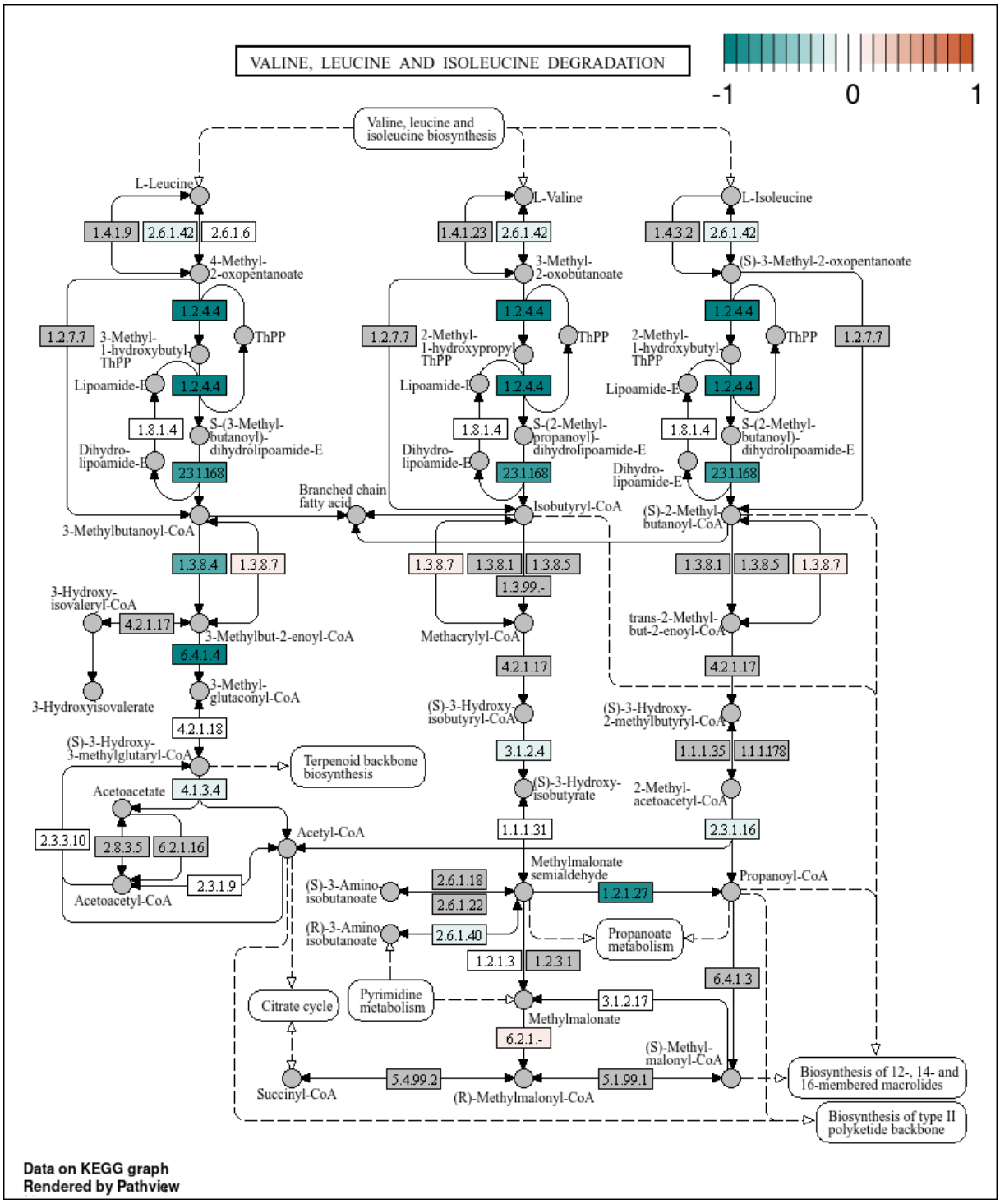
Supplementary Figure S3.9 KEGG pathway visualization: Propanoate metabolism. Differential expression (log_2FC) in *C. reinhardtii* co-cultured with SynCom versus cultured axenically in the exponential phase, mapped to KEGG Orthology KO identifiers. When multiple *C. reinhardtii* genes share the same KO annotation, their log_2FC values are averaged. Pathways were visualized using a fixed log_2FC scale (−1 to +1); orange indicates upregulation and teal indicates downregulation. Grey indicates pathway elements without mapped expression data.



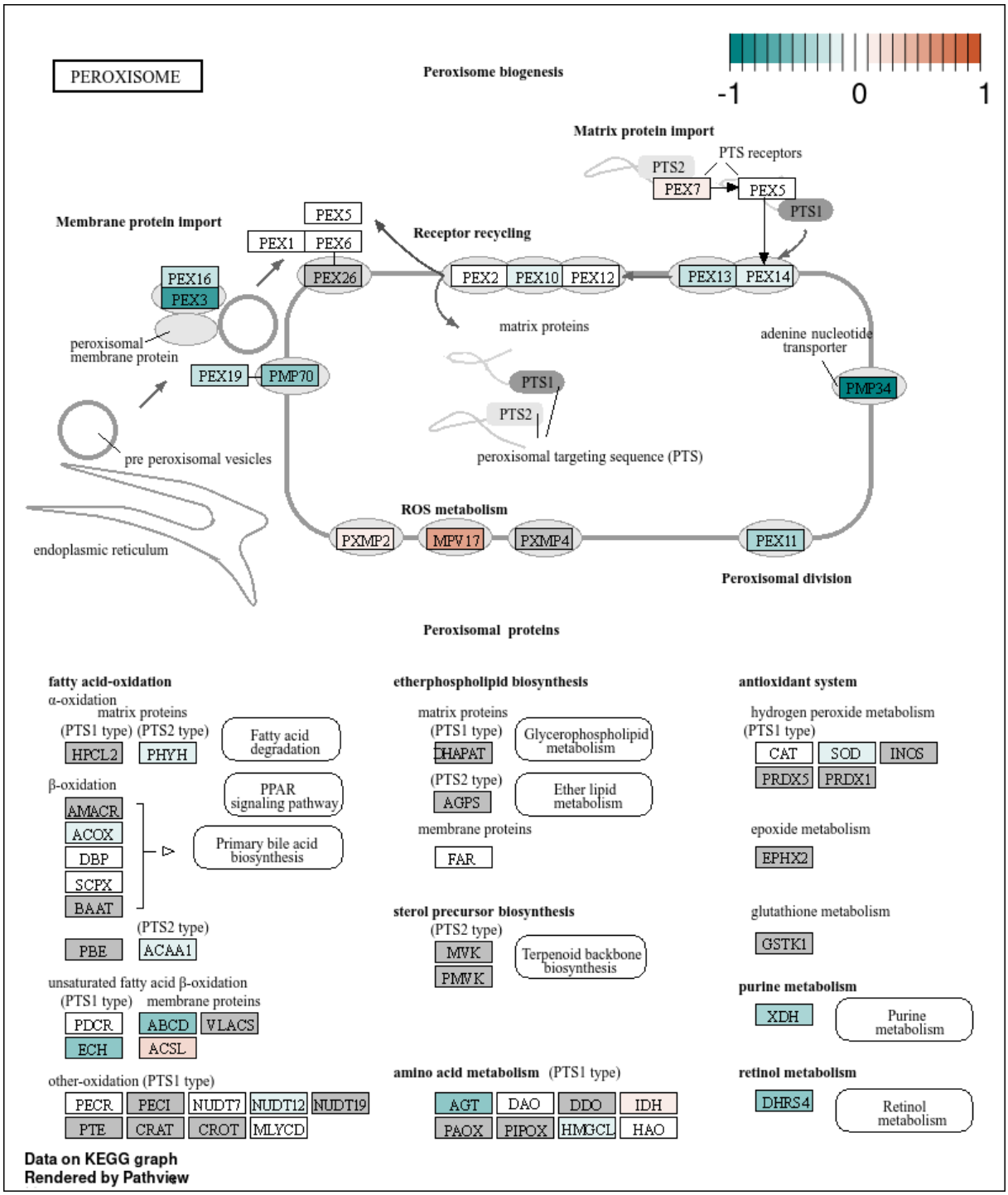
Supplementary Figure S3.10 KEGG pathway visualization: Glyoxylate and dicarboxylate metabolism. Differential expression (log_2FC) in *C. reinhardtii* co-cultured with SynCom versus cultured axenically in the exponential phase, mapped to KEGG Orthology KO identifiers. When multiple *C. reinhardtii* genes share the same KO annotation, their log_2FC values are averaged. Pathways were visualized using a fixed log_2FC scale (−1 to +1); orange indicates upregulation and teal indicates downregulation. Grey indicates pathway elements without mapped expression data.



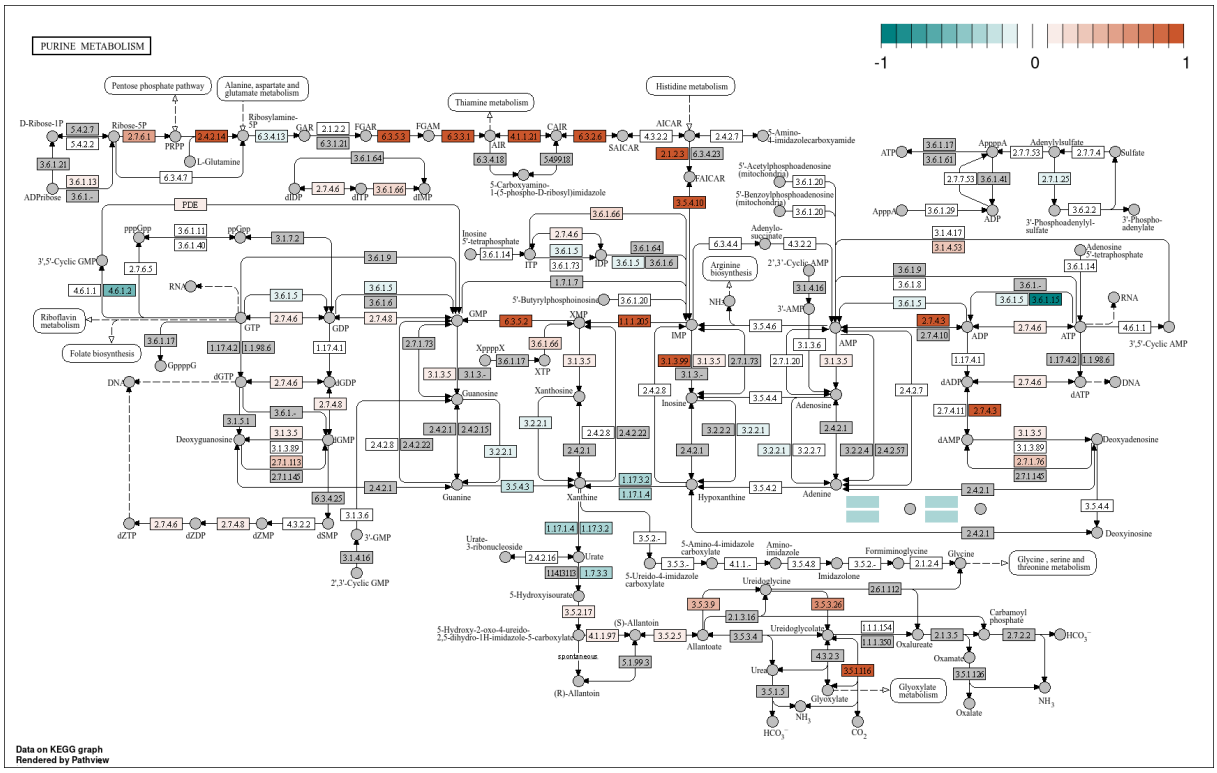
Supplementary Figure S3.11 KEGG pathway visualization: Valine, leucine and isoleucine degradation. Differential expression (log_2FC) in *C. reinhardtii* co-cultured with SynCom versus cultured axenically in the exponential phase, mapped to KO identifiers. When multiple *C. reinhardtii* genes share the same KO annotation, their log_2FC values are averaged. Pathways were visualized using a fixed log_2FC scale (−1 to +1); orange indicates upregulation and teal indicates downregulation. Grey indicates pathway elements without mapped expression data.



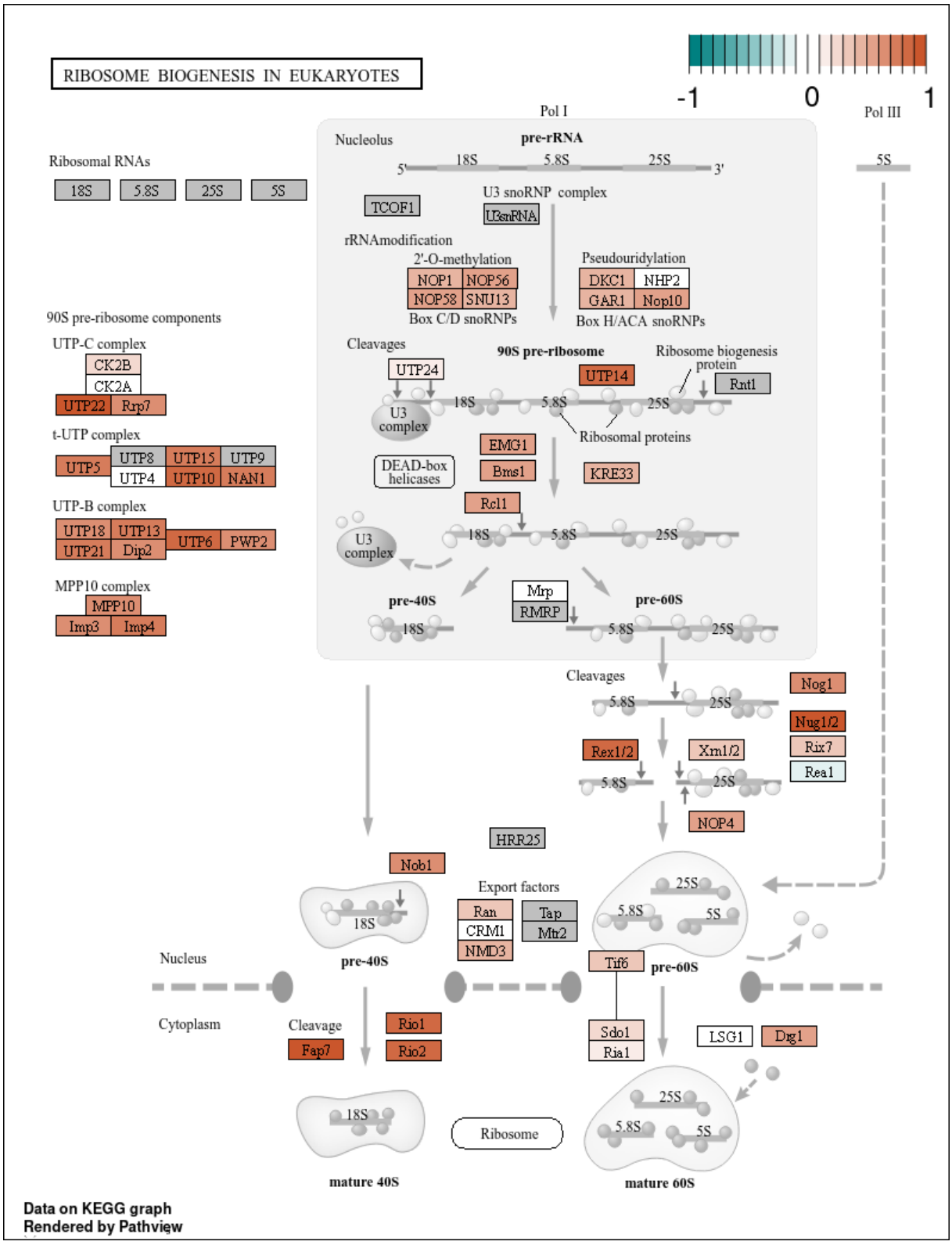
Supplementary Figure S3.12 KEGG pathway visualization: Valine, leucine and isoleucine degradation. Differential expression (log_2FC) in *C. reinhardtii* co-cultured with SynCom versus cultured axenically in the stationary phase, mapped to KO identifiers. When multiple *C. reinhardtii* genes share the same KO annotation, their log_2FC values are averaged. Pathways were visualized using a fixed log_2FC scale (−1 to +1); orange indicates upregulation and teal indicates downregulation. Grey indicates pathway elements without mapped expression data.



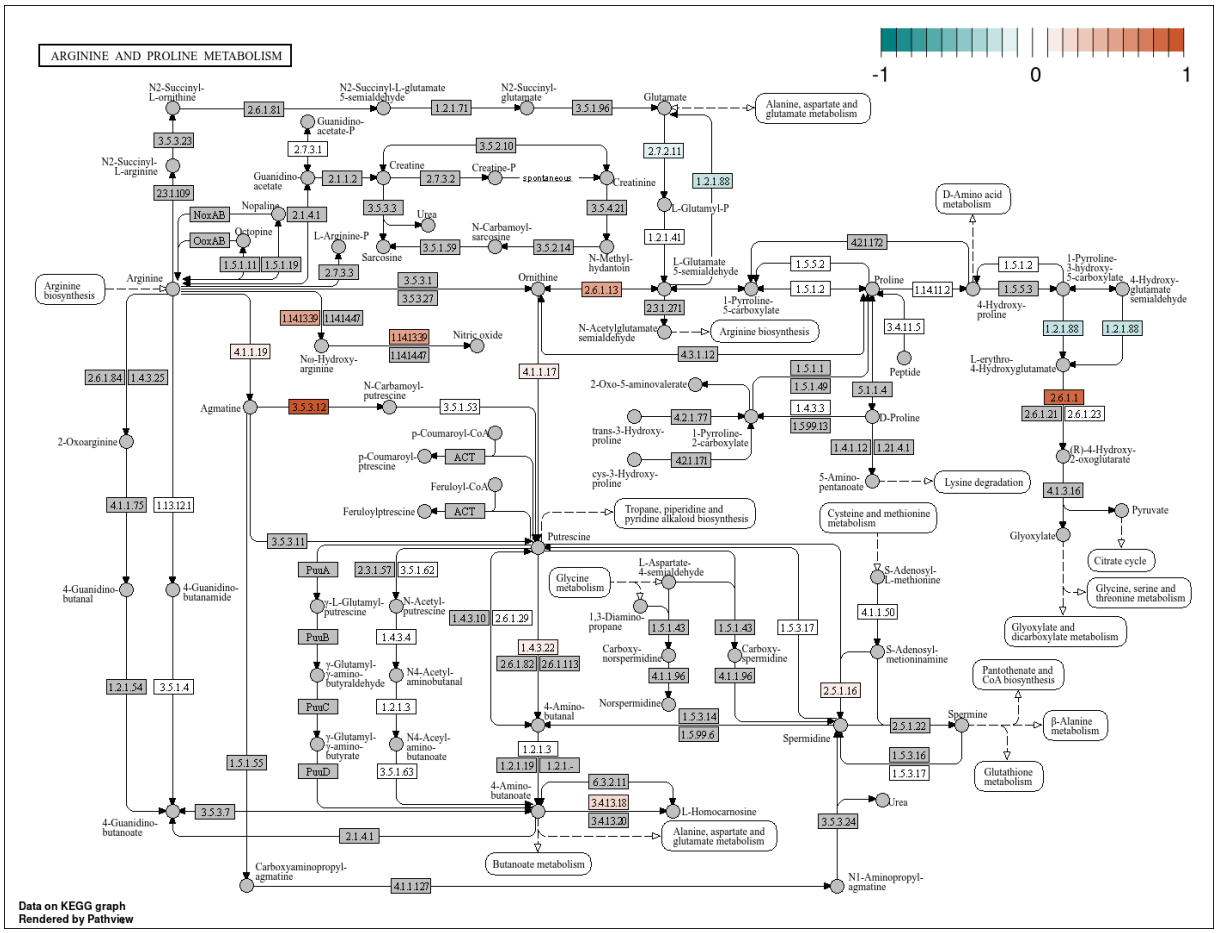
Supplementary Figure S3.13 KEGG pathway visualization: Peroxisome. Differential expression (log_2FC) in *C. reinhardtii* co-cultured with SynCom versus cultured axenically in the stationary phase, mapped to KO identifiers. When multiple *C. reinhardtii* genes share the same KO annotation, their log_2FC values are averaged. Pathways were visualized using a fixed log_2FC scale (−1 to +1); orange indicates upregulation and teal indicates downregulation. Grey indicates pathway elements without mapped expression data.



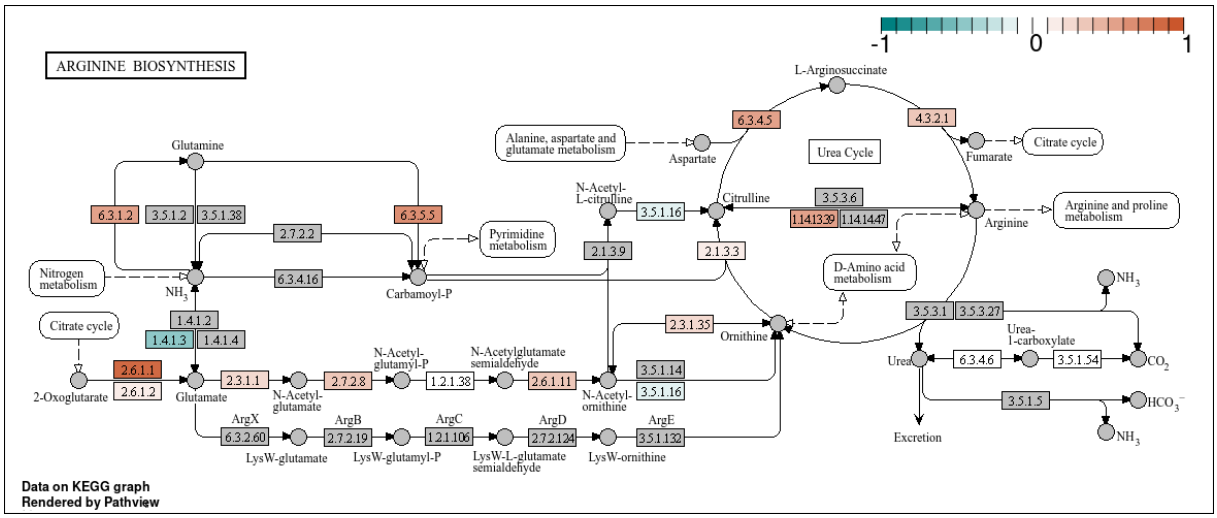
Supplementary Figure S3.14 KEGG pathway visualization: Purine metabolism. Differential expression (log_2FC) in *C. reinhardtii* co-cultured with SynCom versus cultured axenically in the stationary phase, mapped to KO identifiers. When multiple *C. reinhardtii* genes share the same KO annotation, their log_2FC values are averaged. Pathways were visualized using a fixed log_2FC scale (–1 to +1); orange indicates upregulation and teal indicates downregulation. Grey indicates pathway elements without mapped expression data.



Supplementary Figure S3.15 KEGG pathway visualization: Ribosome biogenesis in eukaryotes. Differential expression (log_2FC) in *C. reinhardtii* co-cultured with SynCom versus cultured axenically in the stationary phase, mapped to KO identifiers. When multiple *C. reinhardtii* genes share the same KO annotation, their log_2FC values are averaged. Pathways were visualized using a fixed log_2FC scale (−1 to +1); orange indicates upregulation and teal indicates downregulation. Grey indicates pathway elements without mapped expression data.



Supplementary Figure S3.16 KEGG pathway visualization: Arginine and proline metabolism. Differential expression (log_2FC) in *C. reinhardtii* co-cultured with SynCom versus cultured axenically in the stationary phase, mapped to KO identifiers. When multiple *C. reinhardtii* genes share the same KO annotation, their log_2FC values are averaged. Pathways were visualized using a fixed log_2FC scale (−1 to +1); orange indicates upregulation and teal indicates downregulation. Grey indicates pathway elements without mapped expression data.



Supplementary Figure S3.17 KEGG pathway visualization: Arginine biosynthesis. Differential expression (log_2FC) in *C. reinhardtii* co-cultured with SynCom versus cultured axenically in the stationary phase, mapped to KO identifiers. When multiple *C. reinhardtii* genes share the same KO annotation, their log_2FC values are averaged. Pathways were visualized using a fixed log_2FC scale (–1 to +1); orange indicates upregulation and teal indicates downregulation. Grey indicates pathway elements without mapped expression data.

3.7 Supplementary Tables

Supplementary Table S3.1 Complete list of the identified DEGs with log2FC values, adjusted *P*–values, functional annotations and information about conservation with *A. thaliana* and *L. japonicus*. Table in the additional file: Chapter 3 Supplementary Tables.

Supplementary Table S3.2 Gene Ontology Biological Process terms significantly enriched in the DEGs expression clusters. Table in the additional file: Chapter 3 Supplementary Tables.

Supplementary Table S3.3 Gene Ontology Molecular Function terms significantly enriched in the DEGs expression clusters. Table in the additional file: Chapter 3 Supplementary Tables.

Supplementary Table S3.4 Gene Ontology Biological Process terms significantly enriched in the different experimental comparisons. Table in the additional file: Chapter 3 Supplementary Tables.

Supplementary Table S3.5 Gene Ontology Molecular Function terms significantly enriched in the different experimental comparisons. Table in the additional file: Chapter 3 Supplementary Tables.

Supplementary Table S3.6 KEGG processes (level B) significantly enriched in the different experimental comparisons.

Comparison	**Regulation direction**	**Process ID**	**Gene Set Size**	**Enrichment score**	***P*–value**	**Adjusted *P*–value**	**Relevance**
SynCom vs axenic (exponential phase)	DOWN	09105 Amino acid metabolism	303	-0.61	6.82×10^{-4}	2.76×10^{-2}	relevant, kept
SynCom vs axenic (exponential phase)	DOWN	09101 Carbohydrate metabolism	372	-0.57	1.15×10^{-3}	2.76×10^{-2}	relevant, kept
SynCom vs axenic (stationary phase)	UP	09122 Translation	436	0.51	1.09×10^{-8}	5.22×10^{-7}	relevant, kept
SynCom vs axenic (stationary phase)	UP	09104 Nucleotide metabolism	167	0.56	1.10×10^{-5}	1.32×10^{-4}	relevant, kept
SynCom vs axenic (stationary phase)	UP	09121 Transcription	198	0.48	1.51×10^{-3}	1.21×10^{-2}	relevant, kept
SynCom vs axenic (stationary phase)	DOWN	09103 Lipid metabolism	219	-0.50	3.51×10^{-7}	6.35×10^{-6}	relevant, kept
SynCom vs axenic (stationary phase)	DOWN	09141 Transport and catabolism	315	-0.44	3.97×10^{-7}	6.35×10^{-6}	relevant, kept
SynCom vs axenic (stationary phase)	DOWN	09143 Cell growth and death	275	-0.38	8.36×10^{-4}	8.03×10^{-3}	relevant, kept
SynCom vs axenic (stationary phase)	DOWN	09162 Cancer: specific types	101	-0.45	3.29×10^{-3}	2.26×10^{-2}	not relevant, removed
SynCom vs axenic (stationary phase)	DOWN	09107 Glycan biosynthesis and metabolism	151	-0.42	3.79×10^{-3}	2.27×10^{-2}	relevant, kept

Timepoint effect in Axenic (stationary vs exponential)	UP	09172 Infectious disease: viral	350	0.48	2.14×10^{-4}	1.71×10^{-3}	not relevant, removed
Timepoint effect in Axenic (stationary vs exponential)	UP	09123 Folding, sorting and degradation	369	0.47	3.86×10^{-4}	2.64×10^{-3}	relevant, kept
Timepoint effect in Axenic (stationary vs exponential)	UP	09124 Replication and repair	162	0.49	9.23×10^{-3}	4.44×10^{-2}	relevant, kept
Timepoint effect in Axenic (stationary vs exponential)	UP	09122 Translation	436	0.42	9.25×10^{-3}	4.44×10^{-2}	relevant, kept
Timepoint effect in Axenic (stationary vs exponential)	DOWN	09159 Environmental adaptation	211	-0.45	5.61×10^{-8}	2.69×10^{-6}	relevant, kept
Timepoint effect in Axenic (stationary vs exponential)	DOWN	09166 Cardiovascular disease	214	-0.42	1.66×10^{-6}	3.99×10^{-5}	not relevant, removed
Timepoint effect in Axenic (stationary vs exponential)	DOWN	09101 Carbohydrate metabolism	372	-0.33	4.97×10^{-5}	7.94×10^{-4}	relevant, kept
Timepoint effect in Axenic (stationary vs exponential)	DOWN	09153 Circulatory system	78	-0.51	8.19×10^{-5}	9.83×10^{-4}	not relevant, removed
Timepoint effect in Axenic (stationary vs exponential)	DOWN	09157 Sensory system	45	-0.59	1.55×10^{-4}	1.49×10^{-3}	not relevant, removed
Timepoint effect in Axenic (stationary vs exponential)	DOWN	09102 Energy metabolism	333	-0.29	8.29×10^{-3}	4.44×10^{-2}	relevant, kept
Differential timepoint response (SynCom and Axenic cultures)	UP	09122 Translation	436	0.59	1.36×10^{-8}	6.51×10^{-7}	relevant, kept
Differential timepoint response (SynCom and Axenic cultures)	DOWN	09141 Transport and catabolism	315	-0.46	3.64×10^{-5}	8.74×10^{-4}	relevant, kept
Differential timepoint response (SynCom and Axenic cultures)	DOWN	09107 Glycan biosynthesis and metabolism	151	-0.54	2.99×10^{-4}	4.78×10^{-3}	relevant, kept
Timepoint effect in SynCom (stationary vs exponential)	UP	09122 Translation	436	0.53	1.15×10^{-4}	5.38×10^{-3}	relevant, kept
Timepoint effect in SynCom (stationary vs exponential)	DOWN	09109 Metabolism of terpenoids and polyketides	65	-0.46	1.67×10^{-3}	3.92×10^{-2}	relevant, kept

Supplementary Table S3.7 KEGG pathways significantly enriched in the different experimental comparisons.

Supplementary Table S3.8 Summary of gene conservation in *C. reinhardtii*, *A. thaliana*, *L. japonicus* with the conservation categories, number and percentage of genes.

Category	**Organism**	**Number of genes**	**Percentage of genes**
Conserved in all 3 organisms	*C. reinhardtii*	1879	10.62
Conserved, *Chlamydomonas-Arabidopsis*	*C. reinhardtii*	219	1.24
Conserved, *Chlamydomonas-Lotus*	*C. reinhardtii*	156	0.88
Not conserved	*C. reinhardtii*	14883	84.12
RBHs in 2 organisms, not identical	*C. reinhardtii*	556	3.14
Conserved in all 3 organisms	*A. thaliana*	1879	6.79
Conserved, *Arabidopsis-Lotus*	*A. thaliana*	9082	32.83
Conserved, *Chlamydomonas-Arabidopsis*	*A. thaliana*	394	1.42
Not conserved	*A. thaliana*	15924	57.57
RBHs in 2 organisms, not identical	*A. thaliana*	381	1.38
Conserved in all 3 organisms	*L. japonicus*	1879	6.26
Conserved, *Arabidopsis-Lotus*	*L. japonicus*	9056	30.15
Conserved, *Chlamydomonas-Lotus*	*L. japonicus*	305	1.02
Not conserved	*L. japonicus*	18390	61.22
RBHs in 2 organisms, not identical	*L. japonicus*	407	1.35

Supplementary Table S3.9 Summary of DEGs conservation in *C. reinhardtii*, *A. thaliana*, *L. japonicus* with the conservation categories, number and percentage of DEGs.

Category	**Organism**	**Number of DEGs**	**Percentage of DEGs**
Conserved in all 3 organisms	*C. reinhardtii*	172	15.40
Conserved, *Chlamydomonas-Arabidopsis*	*C. reinhardtii*	16	1.43
Conserved, *Chlamydomonas-Lotus*	*C. reinhardtii*	10	0.90
Not conserved	*C. reinhardtii*	870	77.89
RBHs in 2 organisms, not identical	*C. reinhardtii*	49	4.39
Conserved in all 3 organisms	*A. thaliana*	1831	9.48
Conserved, *Arabidopsis-Lotus*	*A. thaliana*	8082	41.84
Conserved, *Chlamydomonas-Arabidopsis*	*A. thaliana*	357	1.85
Not conserved	*A. thaliana*	8687	44.97
RBHs in 2 organisms, not identical	*A. thaliana*	361	1.87
Conserved in all 3 organisms	*L. japonicus*	1701	9.00
Conserved, *Arabidopsis-Lotus*	*L. japonicus*	7377	39.04
Conserved, *Chlamydomonas-Lotus*	*L. japonicus*	253	1.34
Not conserved	*L. japonicus*	9189	48.63
RBHs in 2 organisms, not identical	*L. japonicus*	377	2.00

Supplementary Table S3.10 Summary of *C. reinhardtii* downregulated DEGs with the conservation categories, number and percentage of DEGs.

Category	**Organism**	**Number of DEGs**	**Percentage of DEGs**
Conserved in all 3 organisms, consistently downregulated	*C. reinhardtii*	6	5.94
Conserved in all 3 organisms, inconsistent regulation	*C. reinhardtii*	42	41.58
Conserved, *Chlamydomonas-Arabidopsis*, inconsistent regulation	*C. reinhardtii*	6	5.94
Conserved, *Chlamydomonas-Lotus*, consistently downregulated	*C. reinhardtii*	2	1.98
Conserved, *Chlamydomonas-Lotus*, inconsistent regulation	*C. reinhardtii*	1	0.99
Conserved, *Chlamydomonas-Lotus*, not differentially expressed in both	*C. reinhardtii*	2	1.98
RBHs in 2 organisms, not identical	*C. reinhardtii*	42	41.58

Supplementary Table S3.11 Summary of *C. reinhardtii* upregulated DEGs with the conservation categories, number and percentage of DEGs.

Category	**Organism**	**Number of DEGs**	**Percentage of DEGs**
Conserved in all 3 organisms, consistently upregulated	*C. reinhardtii*	48	24.62
Conserved in all 3 organisms, inconsistent regulation	*C. reinhardtii*	75	38.46
Conserved in all 3 organisms, not differentially expressed in all	*C. reinhardtii*	1	0.51
Conserved, *Chlamydomonas-Arabidopsis*, consistently upregulated	*C. reinhardtii*	7	3.59
Conserved, *Chlamydomonas-Arabidopsis*, inconsistent regulation	*C. reinhardtii*	3	1.54
Conserved, *Chlamydomonas-Lotus*, consistently upregulated	*C. reinhardtii*	4	2.05
Conserved, *Chlamydomonas-Lotus*, inconsistent regulation	*C. reinhardtii*	1	0.51
RBHs in 2 organisms, not identical	*C. reinhardtii*	56	28.72

Supplementary Table S3.12 Candidate genes differentially expressed in response to SynCom presence, with full functional annotation, expression changes, KEGG pathway assignment, and conservation in *A. thaliana* and *L. japonicus*. Table in the additional file: Chapter 3 Supplementary Tables.

Supplementary Table S3.13 Chemical composition of TP10 medium used in the experiment (pH 7.0, adjusted with HCl).

Component	Chemical formula	Concentration in 1 L TP10
EDTA-Na_2	$C_{10}H_{14}N_2Na_2O_8 \times 2\ H_2O$	0.025 mM
$(NH_4)_6MO_7$		0.0000285 µM
Na_2SeO_3		0.0001 mM
Zn × EDTA	$ZnSO_4 \times 7H_2O$	0.0025 mM
	$C_{10}H_{14}N_2Na_2O_8 \times 2\ H_2O$	0.00275 mM
Mn × EDTA	$MnCl_2 \times 4\ H_2O$	0.006 mM
	$C_{10}H_{14}N_2Na_2O_8 \times 2\ H_2O$	0.006 mM
Fe × EDTA	$FeCl3 \times 6H_2O$	0.02 mM
	$C_{10}H_{14}N_2Na_2O_8 \times 2\ H_2O$	0.022 mM
	Na_2CO_3	0.022 mM
Cu × EDTA	$CuCl_2 \times 2\ H_2O$	0.002 mM
	$C_{10}H_{14}N_2Na_2O_8 \times 2\ H_2O$	0.002 mM
Tris	2-Amino-2-(hydroxymethyl)-propan-1,3-diol	20 mM
Phosphate buffer	K_2HPO_4	0.620 mM
	KH_2PO_4	0.411 mM
Nutrient stock	$CaCl_2 \times H_2O$	0.340 mM
	$MgSO_4*7H_2O$	0.406 mM
	KNO_3	6.750 mM
	NH_4Cl	0.748 mM

Supplementary Table S3.14 DEGs related to nitrogen metabolism and assimilation. Comparisons in which a gene was significantly differentially regulated are colored in green.

Gene	Description	Log2FC, SynCom vs axenic, exponential phase	Log2FC, SynCom vs axenic, stationary phase	Log2FC, stationary vs exponential phase (SynCom)	Log2FC, stationary vs exponential (axenic)
Ammonium transporters					
Cre03.g159254	Ammonium transporter, Amt family (amt, AMT, MEP)	0.000	0.043	4.810	3.086
Cre06.g284150	Ammonium transporter Rh (rhesus protein)	-0.008	0.009	2.729	2.213
Cre02.g111050	Ammonium transporter	-0.076	0.521	0.642	-0.098
Cre12.g531000	Ammonium transporter, Amt family (amt, AMT, MEP)	0.043	0.527	0.619	0.181
Cre13.g569850	Ammonium transporter	0.007	0.074	1.141	0.805
Cre14.g629920	Ammonium transporter, Amt family (amt, AMT, MEP)	-0.007	1.378	3.070	0.136
Cre09.g410800	High affinity nitrate transporter	0.007	0.052	1.546	1.350
Cre09.g410850	Nitrate transporter	0.004	0.073	6.991	4.872
Nitrate/nitrite transporters					
Cre09.g410900	Nitrite transporter accessory protein	0.003	0.047	3.095	2.034
Cre03.g150101	Nitrate/nitrite transporter	-0.304	0.157	1.119	0.264
Cre03.g150151	Nitrate/nitrite transporter	-0.132	0.068	0.527	0.041
Nitrite reductase					
Cre08.g360550	nitrite reductase (NO-forming) activity	0.030	-0.010	-0.562	-0.425
Cre16.g693202	Sulfite reductase (ferredoxin) (sir), putative ferredoxin-nitrite reductase	-0.016	0.300	4.039	2.659
Cre08.g365692	Ferredoxin-sulfite reductase	-0.040	0.097	0.474	0.219

Cre09.g410750	Nitrite reductase	0.008	0.150	1.672	0.127
Nitrate reductase					
Cre09.g410950		0.009	-0.016	0.216	1.606
GS/GOGAT (glutamine synthetase)					
Cre03.g153400	Glutamine synthetase activity	-1.151	-0.434	0.240	0.007
Cre02.g113200	Glutamine synthetase	0.011	0.591	-0.089	-0.590
Cre03.g146607	Glutamine synthetase activity	-0.068	-0.096	0.452	0.321
Cre12.g530600	Glutamine synthetase activity; ATP binding	-0.019	0.630	0.809	0.022
GS/GOGAT (glutamate synthase)					
Cre16.g694850	N-acetylglutamate synthase	-0.011	0.252	-0.212	-0.633
Cre13.g592200	Glutamate synthase, NADH-dependent	0.147	0.714	0.381	0.040

Supplementary Table S3.15 Availability of CLiP mutants for the candidate genes[32–34,284].

Gene	**Description**	**Program**	**Mutant availability**
Cre02.g092250	FKBP42, Peptidylprolyl isomerase	protein homeostasis	CLiP1
Cre01.g003376	DNAJC2; DnaJ homolog subfamily C member 2	protein homeostasis	CLiP1, CliP2
Cre16.g683500	DEAD/DEAH box helicase, ATP-dependent RNA helicase DDX49/DBP8 (*A. thaliana*: RH36)	translation machinery	CLiP1, CliP2
Cre01.g009350	agmatine iminohydrolase (aguA)	metabolic exchange	CLiP1, CliP2
Cre01.g063267	bystin-like protein	translation machinery	CLiP1, CliP2
Cre02.g092900	GMP synthase (glutamine-hydrolyzing)	growth preparation	CLiP1, CliP2
Cre02.g095400	ribosome biogensis protein (BRX1, BRIX1)	translation machinery	CLiP1, CliP2
Cre03.g159900	RIO kinase 2 (RIOK2)	translation machinery	CLiP1, CliP2
Cre03.g175000	3'-5' exoribonuclease PH component of the exosome (RRP42, EXOSC7)	translation machinery	CLiP2
Cre05.g234638	amidophosphoribosyltransferase (purF, PPAT)	growth preparation	CLiP1, CliP2
Cre06.g254150	Nucleolar protein, component of the U3 processome; periodic tryptophan protein 2 (PWP2, UTP1)	translation machinery	no
Cre06.g277050	protein SDA1 (SDA1, SDAD1); ARM repeat superfamily protein	translation machinery	CLiP1, CliP2
Cre08.g364800	phosphoribosylformylglycinamidine synthase (*A. thaliana* PUR4)	growth preparation	no
Cre08.g378150	glucose-6-phosphate 1-dehydrogenase (G6PD)	growth preparation	CLiP1, CliP2
Cre10.g433100	GTP-binding protein (obgE)	translation machinery	CLiP1, CliP2
Cre10.g438700	U3 small nucleolar RNA-associated protein 12 (DIP2, UTP12, WDR3)	translation machinery	CLiP1, CliP2
Cre10.g442650	ATP-dependent RNA/DNA helicase	translation machinery	CLiP1, CliP2
Cre12.g487200	ribosome assembly protein (SQT1)	translation machinery	CLiP1, CliP2
Cre12.g525200	nucleolar protein 56 (NOP56)	translation machinery	CLiP1, CliP2
Cre12.g551352	periplasmic L-amino acid oxidase 1 (LAO1)	metabolic exchange	CLiP1, CliP2
Cre16.g685900	protein arginine methyltransferase (PRMT3)	translation machinery	CLiP1, CliP2

Supplementary Table S3.16 List of curated KEGG terms relevant for *C. reinhardtii*. Table in the additional file: Chapter 3 Supplementary Tables.

3.8 Appendix: Pilot experiments for experiments with CLiP mutants

To enable functional validation of candidate genes using CLiP mutant libraries[32–34], we first conducted preliminary experiments with the background strains of the CLiP libraries[32–34].

3.8.1 Results

The background strain used in the first CLiP mutant library (CLiP1), CC-5325 cw15 mt- (CC-5325, identical to CC-4533 (CMJ030))[32,33,123], is different from the wild-type strain CC-1690 used in our experiments[105]. Both CC-5325 and CLiP1 mutants have been grown in both TAP and TP media[33]. However, to grow CC-1690 with the SynCom, we use a modified carbon-free TP10 medium, containing 10% of NH_4Cl included in the carbon-free medium TP[105]. In the CLiP experiments, the cultures were grown under constant light at 22 °C, aerated, and mixed with a magnetic stirrer at 200 rpm[33], while we cultivate CC-1690 under constant light at 25 °C in 200-mL Erlenmeyer flasks on a rotary shaker[105]. To determine if CC-5325 can grow similarly to CC-1690 in our experimental setup, and if optimization of growth conditions is needed, we investigated its growth at different temperatures, shaking speeds, CO_2 concentrations, and investigated the influence of regular passaging of the cultures on growth.

3.8.2 *C. reinhardtii* CC-5325 growth conditions optimization

We investigated the influence of shaking and temperature on CC-5325 photoautotrophic growth in TP10 by comparing three different speeds of shaking (rotary shaker): 0 rpm, the usual 100 rpm currently used in experiments with CC-1690, and 175 rpm, a speed similar to the mixing speed in CLiP experiments (200 rpm could not be used, as it causes the flasks to fall off from the shaker)[33]. We maintained the cultures at two temperatures: 22 °C and 25 °C and monitored the growth by measuring *C. reinhardtii* culture fluorescence at 680 nm for 18 days (Fig. A3.1a).

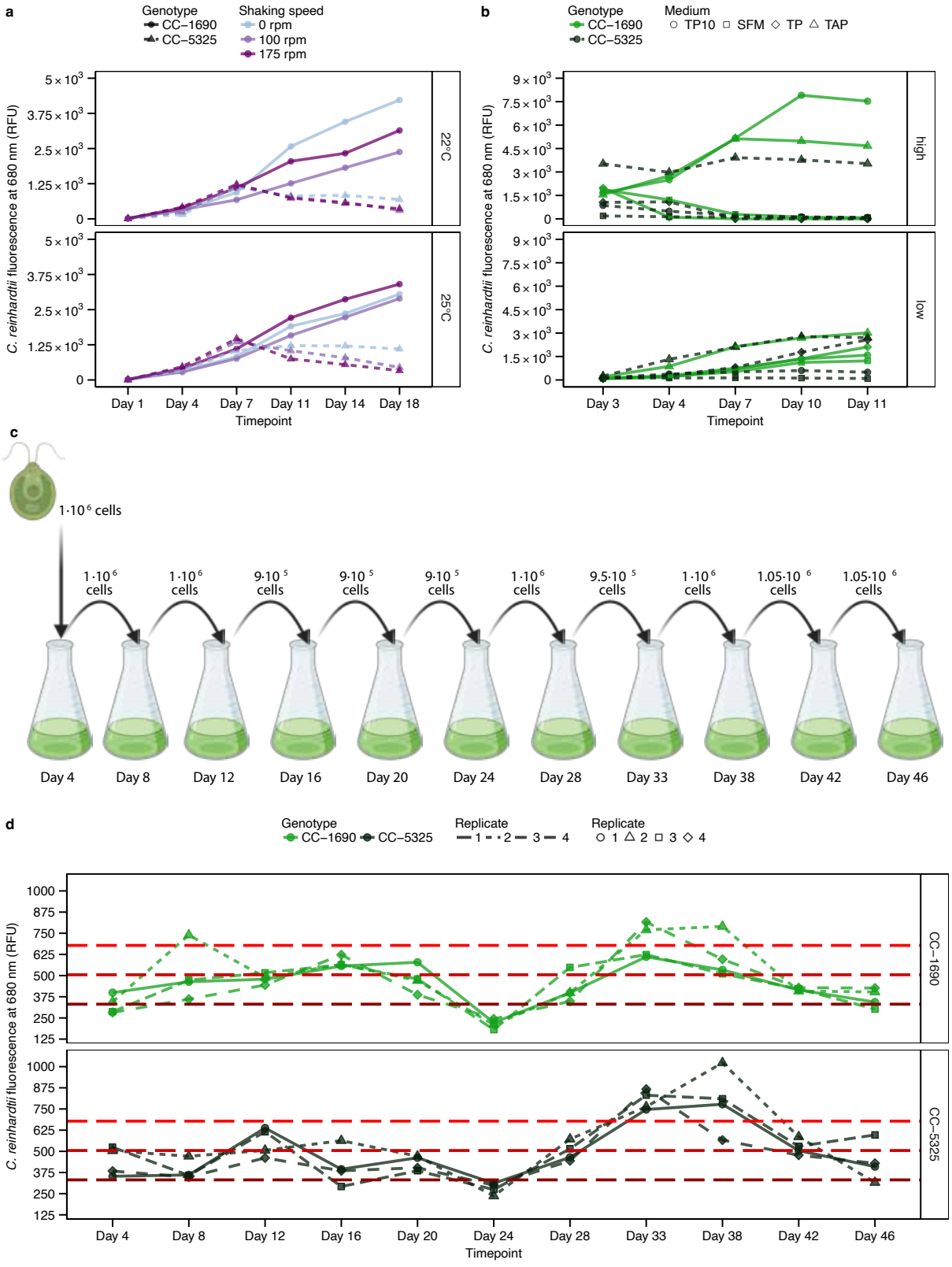


Figure A3.1 CC-5325 growth optimization experiments. a – Comparison of CC-1690 and CC-5325 cultures fluorescence at 680 nm at 22 °C and 25 °C, at 3 different shaking speeds: 0, 100, and 175 rpm. b – Comparison of CC-1690 and CC-5325 cultures fluorescence at 680 nm in high CO_2 and low CO_2 conditions in different culture media. c – A scheme showing the experimental setup of the experiment

investigating the influence of regular passaging on CC-5325 growth; numbers of cells used for initial inoculation (exact, measured with a cell counter device) and each passage (approximate, based on the fluorescence of the culture at 680 nm) are shown. d – CC-1690 and CC-5325 fluorescence at 680 nm over time, when the cultures were regularly passaged. The measurements were taken before each passage. The red lines show fluorescence values corresponding to 1×10^6 cells (dark red, bottom), 1.5×10^6 cells, (red, middle), and 2×10^6 cells (light red, top).

We observed minimal impact of temperature on algal growth. Until day 7, CC-1690 and CC-5325 grew similarly and achieved comparable fluorescence levels on days 1, 4, and 7 both at 22 °C and 25 °C for all shaking speeds. From day 7 onward, the CC-1690 and CC-5325 growth patterns differed. While fluorescence intensities of CC-1690 cultures increased over time, fluorescence intensities of CC-5325 cultures decreased, indicating a lack of growth and cell death. Apart from the decrease in fluorescence levels, we observed yellowing of the CC-5325 cultures with time, while CC-1690 cultures remained green (Fig. A3.2). We concluded that modifying the shaking and temperature is not beneficial to CC-5325 growth. Thus, we decided to keep the parameters at 100 rpm and 25 °C in the following experiments, allowing the cultivation of both strains in the same chambers and comparison with previous CC-1690 experiments.

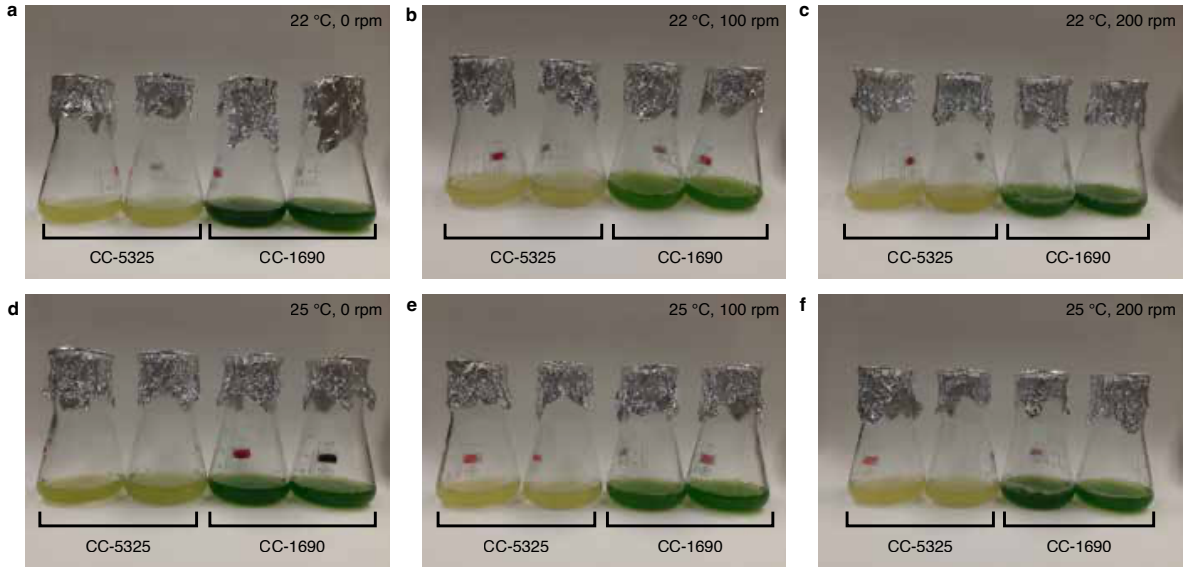


Figure A3.2 CC-1690 and CC-5325 cultures on day 18 after inoculation. Cultures cultivated at 22 °C with shaking speed set to 0 rpm (a), 100 rpm (b) and 175 rpm (c). Cultures cultivated at 25 °C with shaking speed set to 0 rpm (d), 100 rpm (e) and 175 rpm (f).

We hypothesized that CC-5325 grew worse in TP10 than CC-1690 due to a defective carbon concentrating mechanism (CCM). Since CC-5325 is a progeny of a cell wall-deficient strain (CC-4425 cw15 nit2-203 mt+ (D66)[285,286]), and in a *C. reinhardtii*

mutant lacking a cell wall, an enzyme required for normal photoautotrophic growth in atmospheric levels of CO_2, carbonic anhydrase, is released into the growth medium[287–289], we hypothesized that CC-5325 might have a less effective CCM, preventing prolonged photoautotrophic growth. To test whether higher CO_2 levels have a positive impact on CC-5325 growth, we compared cultures grown in low (atmospheric) and high (10,000 ppm) CO_2 levels (Fig. A3.1b). In this experiment, we additionally investigated the impact of growth media, including TP10[105], TP[205], SFM[105], and TAP[290] media. CC-1690 and CC-5325 grew similarly in TP in both CO_2 conditions. High CO_2 level accelerated the processes happening in the cultures: in TAP, both strains grew dynamically until day 7 and then achieved a stable fluorescence level, while under the low CO_2 condition, they grew more slowly until day 11. Under the high CO_2 condition, CC-1690 grew best in TP10, and CC-5325 – in TAP. Both strains did not grow well in SFM and TP media. While poor algal growth in TP can be explained by acidification of the medium, poor CC-5325 growth in TP10 cannot be attributed to acidification, as the medium pH was around 7 for the whole duration of the experiment (Table A3.1).

Table A3.1 pH values of the CC-1690 and CC-5325 cultures in the experiment investigating the effect of CO_2 concentration on CC-5325 growth. pH was measured with indicator test strips.

Timepoint	CO2 condition	Genotype	Medium	pH
Day 4	high	CC-1690	TAP	8
			TP	4.5
			SFM	7
			TP10	8
		CC-5325	TAP	8
			TP	4
			SFM	5.5
			TP10	7
	low	CC-1690	TAP	7.5
			TP	7
			SFM	7
			TP10	7
		CC-5325	TAP	7.5
			TP	5.5
			SFM	6
			TP10	7
Day 7	high	CC-1690	TAP	8
			TP	4.5
			SFM	7
			TP10	8
		CC-5325	TAP	8
			TP	4
			SFM	5.5
			TP10	7
	low	CC-1690	TAP	8
			TP	6.5
			SFM	7
			TP10	7
		CC-5325	TAP	8
			TP	7
			SFM	6

			TP10	8
day 11	high	CC-1690	TAP	8
			TP	4.5
			SFM	7
			TP10	7.5
		CC-5325	TAP	8.5
			TP	4
			SFM	6
			TP10	7
	low	CC-1690	TAP	8
			TP	4.5
			SFM	7
			TP10	7.5
		CC-5325	TAP	8.5
			TP	5.5
			SFM	6
			TP10	7

Under high CO_2, only the cultures grown in TAP and CC-1690 culture grown in TP10 remained green for 7 days, turning yellowish on day 11, probably due to the fast accumulation of dying cells. Cultures grown in TP, SFM, and CC-5325 culture grown in TP10 turned yellow on day 4 (Fig. A3.3a, b). Under low CO_2, CC-5325 cultures in SFM and TP10 media started turning yellow on day 11, with the TP culture remaining green (Fig. A3.3c, d). We decided to stop the experiment at this point, as we observed poor CC-5325 growth in TP10 and SFM and slightly better performance in TP medium, provided there was no acidification. In the CC-5325 culture grown in the high CO_2 level, already on day 4, pH was 4, whereas in the low CO_2 level, pH dropped to 5.5 only on day 11 (Table A3.1). These observations suggest that higher CO_2 levels accelerate the growth of both CC-1690 and CC-5325 and that the previously observed poor growth of CC-5325 in TP10 is not due to a defect in CCM. We decided not to attempt CC-5325 cultivation with the SynCom in TP because, in the past, CO_2 produced in co-cultures of *C. reinhardtii* and the SynCom by bacterial respiration caused acidification of the growth medium, making prolonged cultivation impossible.

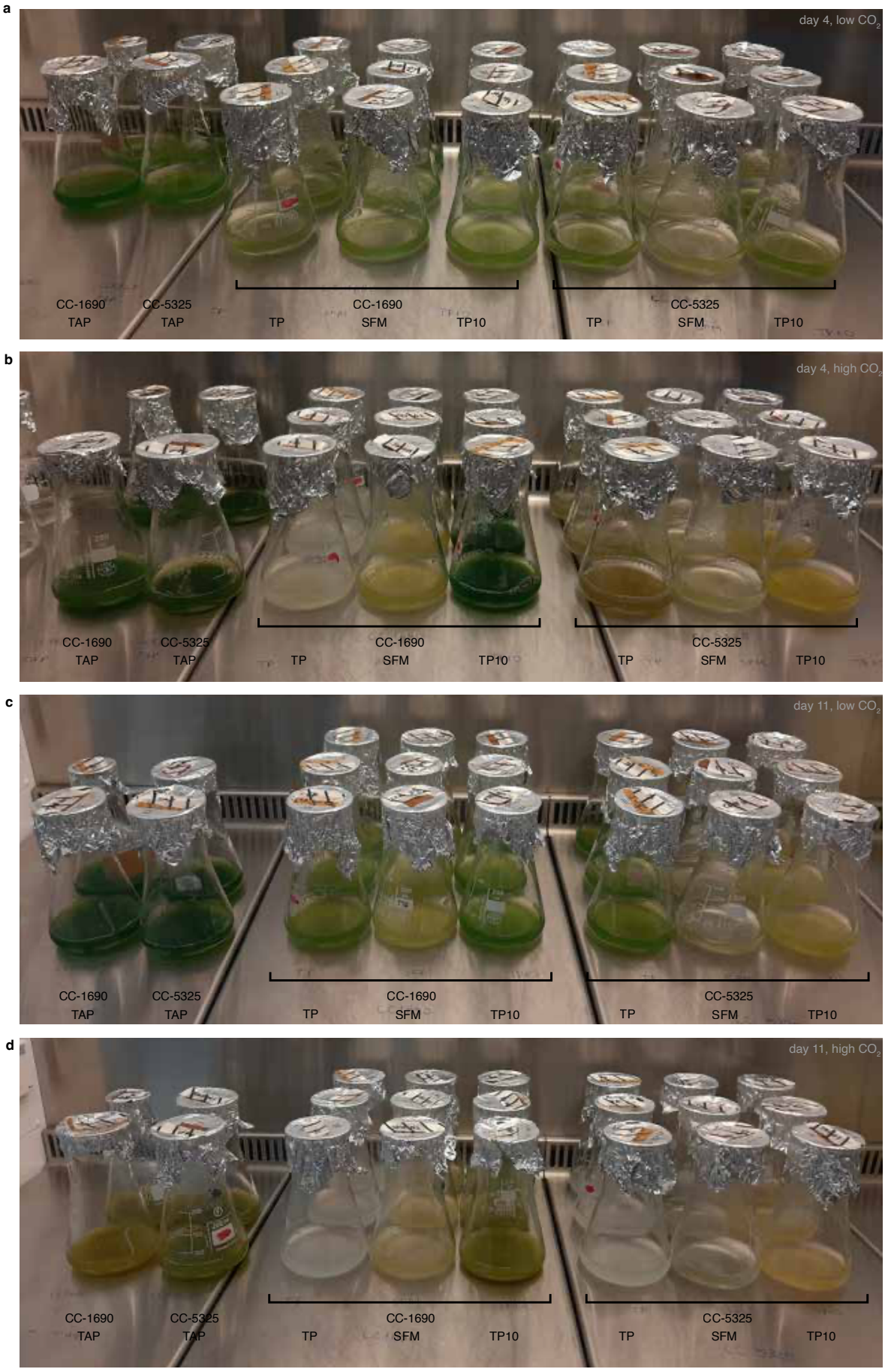

a
day 4, low CO_2
CC-1690 TAP
CC-5325 TAP
TP
CC-1690 SFM
TP10
TP
CC-5325 SFM
TP10
b
day 4, high CO_2
CC-1690 TAP
CC-5325 TAP
TP
CC-1690 SFM
TP10
TP
CC-5325 SFM
TP10
c
day 11, low CO_2
CC-1690 TAP
CC-5325 TAP
TP
CC-1690 SFM
TP10
TP
CC-5325 SFM
TP10
d
day 11, high CO_2
CC-1690 TAP
CC-5325 TAP
TP
CC-1690 SFM
TP10
TP
CC-5325 SFM
TP10

Figure A3.3 CC-1690 and CC-5325 cultures in the experiment investigating the effect of CO_2 concentration on CC-5325 growth. a – cultures cultivated under low CO_2 condition on day 4 after inoculation, b – cultures cultivated under high CO_2 condition on day 4 after inoculation, c – cultures cultivated under low CO_2 condition on day 11 after inoculation, d – cultures cultivated in high CO_2 condition on day 11 after inoculation.

Then, we hypothesized that the poor photoautotrophic growth of CC-5325 could be due to a higher sensitivity to some metabolic by-products, which could be either produced only by CC-5325, and not CC-1690, or which affect CC-5325 more, possibly due to its thinner cell wall. Therefore, as CC-5325 was previously successfully grown both in TAP and TP[33], we decided to grow it in TP10 but regularly passaged the cells to keep the cultures in the cell density range used in the CLiP experiments[33]: from 2×10^4 cells mL^{-1} up to 2×10^6 cells mL^{-1} (Fig. A3.1c). We continued the experiment for 46 days and observed similar growth for CC-1690 and CC-5325, with fluorescence levels corresponding to cell densities ranging approximately from 2×10^4 cells mL^{-1} up to almost 3×10^6 cells mL^{-1} (Fig. A3.1d). These results suggest that CC-5325 can be cultivated for a prolonged time if the cultures are regularly supplied with fresh medium, and cell densities do not exceed 3×10^6 cells mL^{-1}.

3.8.3 Determination of experimental conditions for the experiments with the CLiP2 mutant library

Initially, we planned to use the pooled CLiP2 mutant library in the *C. reinhardtii* CC-5415 background to perform large-scale phenotyping experiments[34]. Therefore, we performed several experiments to determine the optimal conditions for the large-scale phenotyping experiments. These results can serve as a foundation for validating individual candidate genes.

For large-scale phenotyping with the CLiP2 mutant library, the target response of the wild-type *C. reinhardtii* CC-5415 was defined as 50% change in growth compared to axenic controls, assessed when the control cultures reached a cell density of 2×10^6 cells mL^{-1} (Fig. A3.4a)[34]. While for chemical treatments and SynCom with low numbers of SynCom cells, it is possible to approximate growth from culture fluorescence at 680 nm (excitation 440/9 nm, emission 680/20 nm), for SynCom treatments with high numbers of SynCom cells, presence of bacteria influences the fluorescence intensity of the culture samples, causing the fluorescence of the samples containing the same amount of *C. reinhardtii* to differ between axenic samples and

those containing different bacteria. Thus, for those, chlorophyll extraction was performed before growth assessment, and growth was estimated based on chlorophyll fluorescence at 685 nm.

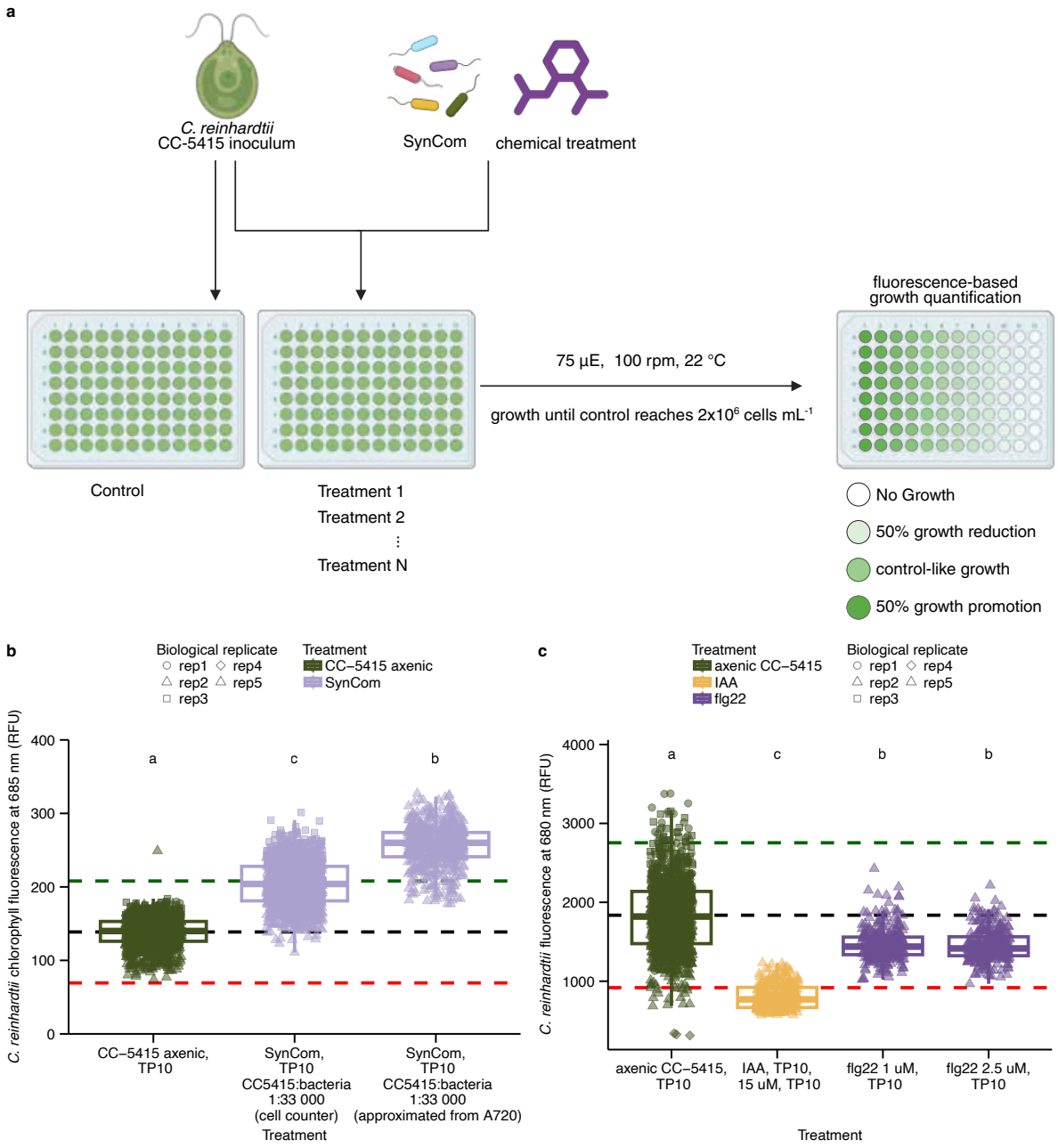


Figure A3.4 Preliminary experiments with the CLiP2 mutant library background strain, *C. reinhardtii* CC-5415**.** a – Experimental setup scheme for the determination of large-scale phenotyping conditions. b – Boxplots showing the results of the SynCom experiment at optimal *C. reinhardtii*:SynCom cell ratio (1:33,000). c – Boxplots showing the effect of IAA and flg22 treatments on *C. reinhardtii* growth. In panels b and c**,** the black horizontal line represents the average growth of axenic *C. reinhardtii,* the green line represents 50% growth promotion, and the red line represents 50% growth reduction. n = 1–5 (independent biological replicates).

Here, we present only a subset of the tested treatment conditions, selecting the most appropriate from a broader range of treatments. For the SynCom experiments, we observed 50% growth promotion effect when *C. reinhardtii* CC-5415 was mixed with the SynCom in 1:33,000 cell:cell ratio (Fig. A3.4b). Initially, heat-killed SynCom was included as a control; however, unlike the treatments with live SynCom, results were inconsistent across experimental rounds, and further optimization is required before heat-killed controls can be used reliably. The 1:33,000 cell:cell ratio is much higher than the 1:100 we have used in our experiments; however, it can be used if growth-promoting effects are to be observed within a short timeframe. Our results suggests that *C. reinhardtii* CC-1690 (our usually used genotype) and CC-5415 (CLiP2 mutant background strain) both benefit from SynCom presence.

Additionally, we identified two chemical treatments that could be potentially used in the large-scale phenotyping: 15 μM IAA, causing 50 % growth inhibition, and 1 μM flg22, which caused *C. reinhardtii* growth reduction, but not by 50 % (Fig. A3.4c). However, because 2.5 μM flg22 did not cause stronger growth reduction, it is possibly the strongest effect we could obtain for this treatment. We included flg22 in our preliminary experiments, expecting to see no effect on *C. reinhardtii* growth, as flg22 is a peptide representing the most conserved domain of bacterial flagellin perceived in *A. thaliana*[291]. In *A. thalian*a, flg22 is perceived by an LRR-type receptor kinase flagellin sensing 2 (FLS2), which activates an immune response[291]. Since the *C. reinhardtii* immune system is much simpler than that of plants, and its RLK homologs do not encode canonical domain architectures, we did not anticipate that *C. reinhardtii* would respond to flg22 treatment[46]. Therefore, the observed reduction in *C. reinhardtii* growth following flg22 treatment was surprising and requires additional research to confirm the effect and elucidate the underlying mechanism. Our determined IAA concentration is much lower than the one that was shown to be inhibitory in the N-free medium[116], but this might be connected to the other medium used, as we used N-rich TP10 medium. Nevertheless, it would be interesting to determine whether IAA-induced growth reduction is rescued by the SynCom presence and whether this effect is SynCom-dependent.

3.8.4 Conclusion

In our preliminary experiments, we determined appropriate growth conditions for the CLiP1 background strain[32,33], and experimental conditions suitable for large-scale phenotyping experiments using the pooled CLiP2 mutant library[34]. Both of the wild-type background strains can be successfully grown and the results can serve as a foundation for validating individual candidate genes, using mutants from both CLiP1 and CLiP2 mutant libraries.

3.8.5 Materials and methods

3.8.5.1 Determination of the influence of temperature and shaking speed on *C. reinhardtii* CC-5325 growth

C. reinhardtii CC-1690 and CC-5325 cultures were inoculated to a density of 10^5 cells mL^{-1} in four replicates per condition into 50 mL TP10 medium (Supplementary Table S3.1), pH 7.0, in 200 mL Erlenmeyer flasks. The cultures were grown axenically under continuous light conditions (125 µmol m^{-2} s^{-1}), at 22 °C or 25 °C on rotary shakers with speeds set to 0 rpm, 100 rpm, or 175 rpm. The growth was monitored by measuring *C. reinhardtii* culture fluorescence at 680 nm (excitation 440/9 nm, emission 680/20 nm)[105] in an Infinite M2000Pro (TECAN Austria GmbH, Grödig, Austria) plate reader with the TECAN i-control software (v2.0.10.0) on days 1, 4, 7, 11, 14 and 18.

3.8.5.2 Determination of the influence of CO_2 concentration on *C. reinhardtii* CC-5325 growth

C. reinhardtii CC-1690 and CC-5325 cultures were inoculated to a density of 10^5 cells mL^{-1} in triplicates into 50 mL medium (TP[205], TP10, or Synthetic Freshwater Medium (SFM)[105]) in 200 mL Erlenmeyer flasks and in duplicates for the controls cultured in TAP[290]. The cultures were grown axenically under continuous light conditions (125 µmol m^{-2} s^{-1}), at 25 °C on rotary shakers (100 rpm) in either atmospheric CO_2 level (ca. 421 ppm[292]) or in high CO_2 (10,000 ppm) level. The growth was monitored by measuring *C. reinhardtii culture* fluorescence at 680 nm (excitation 440/9 nm, emission 680/20 nm)[105] in the Plate reader on days 3, 4, 7 and 11. The pH of the cultures was determined using paper pH indicators.

3.8.5.3 Determination of the influence of regular passaging on *C. reinhardtii* CC-5325 growth

C. reinhardtii CC-1690 and CC-5325 cultures were inoculated to a cell density of 2×10^4 cells mL^{-1} . The cultures were grown axenically in 50 mL TP10 medium[105] under continuous light conditions (125 µmol m^{-2} s^{-1}), at 25 °C in 200-mL Erlenmeyer flasks on rotary shakers set to 100 rpm in an atmospheric CO_2 level. The growth was monitored by measuring *C. reinhardtii* culture fluorescence at 680 nm (excitation 440/9 nm, emission 680/20 nm)[105] in the plate reader. The cultures were passaged every four days by transferring approximately 1×10^6, 9×10^5 or 1.05×10^6 cells (depending on the trend in the culture fluorescence levels) into 50 mL fresh TP10 medium. The approximate cell density was calculated as $Fluo_{680} - (-15.75)/3.47 \times 10^{-4}$ (this equation was determined previously by correlating cell densities of different dilutions of 7-day-old CC-1690 cultures with their fluorescence levels). The experiment was continued for 46 days.

3.8.5.4 *C.* reinhardtii CC-5415 inocula preparation

The following sections (3.8.1.4–3.8.1.7) describe methods for experiments with the CLiP2 background strain CC-5415. 7 days before starting the experiments, *C. reinhardtii* CC-5415 was pre-grown in liquid TAP medium, pH 7.5[290]. Three days prior to the start, cultures were refreshed by transferring 10 mL of the culture into 40 mL of fresh medium. The cultures were grown axenically under continuous light (125 µmol m^{-2} s^{-1}), at 25 °C on rotary shakers at 100 rpm. On the day of the inoculation, *C. reinhardtii* cells were washed three times by centrifugation (1,000 × g, 5-7 min, RT) using the experimental medium (TP10, pH 7.0[105]) and resuspended in the same medium after the final wash. Afterward, cell density was determined using a Multisizer 4e Coulter particle counter, and the cultures were diluted accordingly for the experiment.

3.8.5.5 SynCom inocula preparation

Bacterial cultures from 50% TSA agar plates incubated at 25 °C for 3-4 days (depending on growth speed) were used to inoculate sterile cultivation tubes (number of tubes was dependent on the bacterial strain, based on previous experiments) containing 5 mL of

50% TSB medium per bacterial strain, 8 days before the experiment started. The tubes were incubated at 25 °C and 180 rpm. After bacterial growth in the pre-culture was observed (1-2 days), each pre-culture was transferred to fresh 50% TSB in a flask (45 mL fresh 50 % TSB per 5 mL pre-culture), and incubated at 25 °C and 180 rpm. One day before starting the experiment, the cultures were pelleted at 3,000 × g for 30 min, RT, and washed twice with 25 mL experimental medium (TP10, pH 7.0) at 3,000 × g for 20 min (RT). After the second washing, the bacterial pellets were resuspended in 25 mL experimental medium (TP10, pH 7.0) and incubated at 25 °C and 180 rpm for an additional day. On the inoculation day, bacterial concentration in the washed cultures was determined by measuring absorbance at 720 nm using the plate reader, and the bacteria were then pooled in equal ratios based on absorbance. Subsequently, the cell density of the pooled SynCom was measured using the Multisizer 4e cell counter. If used, heat-killed SynCom was prepared by incubating the SynCom at 96 °C for 1 h.

3.8.5.6 Determination of experimental conditions for the experiments with the CLiP2 mutant library in the CC-5415 background

In all experiments, axenic *C. reinhardtii* cultures with a cell density of 2×10^4 cells mL^{-1} were used as controls. In all treatment conditions, the starting *C. reinhardtii* cell density was equal *to* 2×10^4 cells mL^{-1}. For chemical treatment conditions, *C. reinhardtii* cultures were mixed with a treatment (15 μM auxin prepared in TP10, pH 7.0, 1 μM flg22 prepared in sterile water or 2.5 μM flg22 prepared in sterile water), and for SynCom treatments, with the SynCom inocula in a desired *C. reinhardtii* cells:SynCom cells ratio (1:33 000). SynCom inocula cell density was determined either with the Multisizer 4e particle counter or using an approximation (based on earlier experiments) from absorbance at 720 nm using the plate reader. The control and treatment cultures were dispensed into 96-well round-bottom plates using an electronic pipette VIAFLO96 (VIAFLO 96, INTEGRA Biosciences GmbH). Plates were incubated under continuous light conditions (75 μmol m^{-2} s^{-1}), at 22 °C on rotary shakers at 100 rpm, with relative humidity of 65-75 %. After 4 days, the cell density of the controls was assessed using the Multisizer 4e Coulter particle counter. After the controls reached the cell density of 2×10^6 cells mL^{-1}, the cultures were resuspended using VIAFLO96. For the chemical treatments and corresponding controls,

C. reinhardtii growth was assessed in the Plate reader by measuring *C. reinhardtii* culture fluorescence at 680 nm (excitation 440/9 nm, emission 680/20 nm). For the SynCom treatments and corresponding controls, chlorophyll extraction was performed before growth assessment. Chlorophyll was extracted as follows: 50 µL of the culture was added to 200 µL of 100% ethanol using a VIAFLO96, incubated at room temperature for 2 min, and then centrifuged at 17,000 × g for 7 min[261]. 100 µL of the supernatant was recovered using VIAFLO96, and chlorophyll fluorescence at 685 nm (excitation 470/9 nm, emission 685/20 nm) was measured in the plate reader. Culture fluorescence was measured at 680 nm (excitation 440/9 nm, emission 680/20 nm) for axenic samples and chemical treatments. For SynCom treatments, chlorophyll was extracted and fluorescence measured at 685 nm (excitation 470/9 nm, emission 685/20 nm) to avoid interference from bacterial cells.

3.8.5.7 Data analysis and visualization

Data were analyzed using R (v4.4.0)[179]. To account for systematic differences between experiments (biological replicates) in the chemical treatment experiments with CC-5415, fluorescence measurements were corrected for batch effects using the *removeBatchEffect* function from the *limma*[293] package in R (v4.4.0)[179]. Experimental batch was defined by experiment identifier, and batch-corrected values were used for downstream analyses where indicated. For experiments with the SynCom and CC-5415, batch correction was not applied because the inter-experiment differences were minimal. Statistical differences among treatments were assessed using a Kruskal–Wallis test, followed by Dunn's post hoc test for pairwise comparisons when the overall test was significant. Results were summarized using compact letter displays (CLDs), where treatments sharing the same letter were not significantly different at $P < 0.05$. CLDs were added to figures as annotations to facilitate visual interpretation of statistically distinct groups. Data plots were generated with the package *ggplot2*[283]. Illustrations were created using BioRender.

4 Outlook

This thesis provides the first comprehensive analysis of *C. reinhardtii* transcriptomic response to its phycosphere microbiota, and extends our previous findings by demonstrating that diverse subaerial algae, spanning Chlorophyta and Streptophyta, assemble distinct phycosphere communities from soil-borne bacteria and fungi, with host identity as the primary driver of community composition. These results deepen our understanding of principles governing the interaction of terrestrial algae with their phycosphere microbiota and of the evolutionary origins of interactions between photosynthetic organisms and bacteria.

In Chapter 2, we demonstrate that the capacity to assemble distinct phycosphere communities from soil-borne bacteria and fungi is a feature shared between subaerial green algae, including the model alga *Chlamydomonas reinhardtii* and microscopic algae originating from the same environment. Our panel encompassed Chlorophyta (Chlorophyceae, Ulvophyceae, Trebouxiophyceae) and Streptophyta algae (Klebsormidiophyceae, Zygnematophyceae), with *Spirogloea muscicola* representing the algal lineage believed to be the closest extant relative to the most recent common ancestor of streptophyte algae and land plants. We found the host identity to be the main driver of the community composition, with bacterial communities showing stronger and more consistent responses to the host than fungal communities. In Chapter 3, we provide the first transcriptomic analysis of the *C. reinhardtii* response to a complex SynCom representative of the phycosphere microbiota, identifying genes and processes possibly important for the interaction of *C. reinhardtii* with its associated microbiota. We show that parts of the transcriptional response of *C. reinhardtii* to its phycosphere microbiota are conserved with evolutionarily distant land plants, suggesting their possible ancestral origin.

Despite these advances, several questions remain regarding the phycosphere microbiota assembly. First, although we showed that subaerial green algae assemble their phycosphere microbiota from soil-borne bacteria and fungi, the influence of phycosphere communities on algal fitness needs to be tested. This could be investigated in cross-inoculation experiments and could help distinguish

host-specific from generalist microbial effects. The experiments could be conducted either by using the soil-borne microbes as inoculum, cultivating them with a given algal host and then cross-inoculating the algae with the microbial communities originally associated with different algae, or by creating specific SynComs for each alga. While the experiments using the first approach can be readily performed, this system does not provide high level of control, as the initial inoculum depends on the microbes present in a given soil batch. The second approach provides a higher level of control and reproducibility, but would require establishing culture collections of microbes associated with each alga, similar to those established for *C. reinhardtii*[105] or land plants[84,147], and using them to construct representative SynComs for each algal host, similarly to what was done in the past for *A. thaliana* and *L. japonicus*[84]. Although it is labor-intensive, it would provide a systematic, controlled workflow allowing characterization of the phycosphere influence on the host and dissection of the observed interactions. Therefore, the first approach can be readily used for preliminary experiments, and then SynComs could be used to provide a frame for systematic investigation of the phycosphere communities associated with each host. Second, although we observed host-driven phycosphere microbiota assembly, we do not know which factors determine the observed differences. This could be investigated using metabolomic approaches, which would also be useful to determine which metabolites are exchanged between *C. reinhardtii* and its phycosphere microbiota. The main technical challenge for metabolomic flux analyses for algae is the lack of spatial and temporal separation, as in our experiments, algae and their associated phycosphere microbiota share one compartment, where metabolites are continuously produced and consumed, making it difficult to determine which organism is the producer and which the consumer. This challenge might be addressed by using systems that enable spatial separation, such as the recently designed MetaFlowTrain[260]. However, previous research has shown that the assembly and growth of *C. reinhardtii* phycosphere microbiota beneficial to its host may require both physical proximity and bidirectional exchange of metabolites or signals, making future metabolic profiling challenging. Moreover, because we used only transcriptomics, which reflects mRNA abundance, but not protein levels, metabolic fluxes, or enzyme activities, we cannot determine which metabolites are exchanged between *C. reinhardtii* and the

SynCom, nor can we directly establish causality without genetic perturbation. To this end, we prioritize the functional validation of the candidate genes using available *C. reinhardtii* mutants[32–34]. If causality is found and the expression of a given gene in response to bacteria was conserved among *C. reinhardtii*, *A. thaliana*, and *L. japonicus*, the next step could be perturbing a given gene in *A. thaliana* and/or *L. japonicus* to determine whether the response is also functionally conserved. Additionally, in our research, we focused on the algal side of the interaction. However, our understanding of the interaction can be extended by analyzing the bacterial side. This might be initiated by analyzing genomes of the SynCom bacteria to identify genes that could be important for the interaction with *C. reinhardtii*.

In this thesis, we show that the evolutionary origins of plant-microbe interactions might trace back to the common ancestor of green algae and land plants, possibly predating the divergence of Chlorophyta and Streptophyta. Since some microbial taxa can associate both with algae and plant roots, this suggests shared mechanisms of establishing the interactions, on both the algal and microbial sides. Understanding interactions between photosynthetic organisms and their associated microbiota has implications beyond basic science. Conserved mechanisms between algae and land plants suggest ancient origins of phototroph-microbe interactions, and understanding them can help us harness them for preserving ecosystem health, boosting agricultural and biotechnological yields. Using SynComs to investigate these interactions provides an efficient, controllable system, and the findings can be extrapolated to natural communities. However, increasing the system complexity while maintaining a high control level can facilitate studies of ecological principles governing the interactions. Thus, future research should also focus on developing and employing more complex synthetic "fabricated ecosystems"[146], allowing the control of microbes, environmental variables, and spatiotemporal analysis. The observed similarities between algal phycosphere and root microbiota indicate that algae can serve as convenient model systems informing design of SynComs for various applications. Phycosphere research is an exciting field with implications extending from basic science to agriculture, biotechnology, and ecosystem health.

5 References

1. Holland, H. D. The oxygenation of the atmosphere and oceans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 903–915 (2006).

2. Allen, J. F. & Martin, W. Evolutionary biology: out of thin air: Evolutionary biology. *Nature* **445**, 610–612 (2007).

3. Tomitani, A., Knoll, A. H., Cavanaugh, C. M. & Ohno, T. The evolutionary diversification of cyanobacteria: molecular-phylogenetic and paleontological perspectives. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5442–5447 (2006).

4. De Clerck, O., Bogaert, K. A. & Leliaert, F. Diversity and evolution of algae. in *Advances in Botanical Research* 55–86 (Elsevier, 2012).

5. Gould, S. B., Waller, R. F. & McFadden, G. I. Plastid evolution. *Annu. Rev. Plant Biol.* **59**, 491–517 (2008).

6. Keeling, P. J. The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 729–748 (2010).

7. Leliaert, F., Verbruggen, H. & Zechman, F. W. Into the deep: new discoveries at the base of the green plant phylogeny. *Bioessays* **33**, 683–692 (2011).

8. Archibald, J. M. The evolution of algae by secondary and tertiary endosymbiosis. in *Advances in Botanical Research* vol. 64 87–118 (Elsevier, 2012).

9. Kim, E. & Archibald, J. M. Diversity and evolution of plastids and their genomes. in *Plant Cell Monographs* 1–39 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008).

10. Andersen, R. A. Biology and systematics of heterokont and haptophyte algae. *Am. J. Bot.* **91**, 1508–1522 (2004).

11. Greenwood, A. D. Chloroplasts and cell compartments in Cryptophyceae. *Br. Phycol. J.* **12**, 119 (1977).

12. Tengs, T. *et al.* Phylogenetic analyses indicate that the 19'Hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol. Biol. Evol.* **17**, 718–729 (2000).

13. Leliaert, F. *et al.* Phylogeny and molecular evolution of the Green algae. *CRC Crit. Rev. Plant Sci.* **31**, 1–46 (2012).

14. Melkonian, M. Flagellae apparatus ultrastructure in relation to green algal classification. *Systematics of the Green Algae* 73–120 (1984).

15. de Vries, J. & Archibald, J. M. Plant evolution: landmarks on the path to terrestrial life. *New Phytol.* **217**, 1428–1434 (2018).

16. Marin, B. & Melkonian, M. Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-encoded rRNA operons. *Protist* **161**, 304–336 (2010).

17. Latasa, M., Scharek, R., Gall, F. L. & Guillou, L. PIGMENT SUITES AND TAXONOMIC GROUPS IN PRASINOPHYCEAE[1]. *J. Phycol.* **40**, 1149–1155 (2004).

18. Darienko, T. & Pröschold, T. Toward a monograph of non-marine Ulvophyceae using an integrative approach (Molecular phylogeny and systematics of terrestrial Ulvophyceae II.). *Phytotaxa* **324**, 1 (2017).

19. Bierenbroodspot, M. J. *et al.* Phylogeny and evolution of streptophyte algae. *Ann. Bot.* **134**, 385–400 (2024).

20. Watanabe, S. & Nakayama, T. Ultrastructure and phylogenetic relationships of the unicellular green algae *Ignatius tetrasporus* and *Pseudocharacium americanum* (Chlorophyta). *Phycological Res.* **55**, 1–16 (2007).

21. Mine, I., Menzel, D. & Okuda, K. Morphogenesis in giant-celled algae. *Int. Rev. Cell Mol. Biol.* **266**, 37–83 (2008).

22. Cocquyt, E., Verbruggen, H., Leliaert, F. & De Clerck, O. Evolution and cytological diversification of the green seaweeds (Ulvophyceae). *Mol. Biol. Evol.* **27**, 2052–2061 (2010).

23. McNaughton, E. E. & Goff, L. J. The role of microtubules in establishing nuclear spatial patterns in multinucleate green algae. *Protoplasma* **157**, 19–37 (1990).

24. Motomura, T. Cell cycle analysis in a multinucleate green alga, *Boergesenia forbesii* (Siphonocladales, Chlorophyta). *Phycological Res.* **44**, 11–17 (1996).

25. Friedl, T. & Bhattacharya, D. Origin and evolution of green lichen algae: Evolution of green lichen algae. in *Cellular Origin, Life in Extreme Habitats and Astrobiology* 341–357 (Springer Netherlands, Dordrecht, 2001).

26. de Koning, A. P. & Keeling, P. J. The complete plastid genome sequence of the parasitic green alga Helicosporidium sp. is highly reduced and structured. *BMC Biol.* **4**, 12 (2006).

27. Salomé, P. A. & Merchant, S. S. A series of fortunate events: Introducing Chlamydomonas as a reference organism. *Plant Cell* **31**, 1682–1707 (2019).

28. Sager, R. & Granick, S. Nutritional control of sexuality in Chlamydomonas reinhardi. *J. Gen. Physiol.* **37**, 729–742 (1954).

29. Craig, R. J. *et al.* The Chlamydomonas Genome Project, version 6: Reference assemblies for mating-type plus and minus strains reveal extensive structural mutation in the laboratory. *Plant Cell* **35**, 644–672 (2023).

30. Shimogawara, K., Fujiwara, S., Grossman, A. & Usuda, H. High-efficiency transformation of Chlamydomonas reinhardtii by electroporation. *Genetics* **148**, 1821–1828 (1998).

31. Neupert, J., Shao, N., Lu, Y. & Bock, R. Genetic transformation of the model green alga Chlamydomonas reinhardtii. *Methods Mol. Biol.* **847**, 35–47 (2012).

32. Li, X. *et al.* A genome-wide algal mutant library and functional screen identifies genes required for eukaryotic photosynthesis. *Nat. Genet.* **51**, 627–635 (2019).

33. Fauser, F. *et al.* Systematic characterization of gene function in the photosynthetic alga Chlamydomonas reinhardtii. *Nat. Genet.* 1–10 (2022).

34. Lunardon, A. *et al.* The*Chlamydomonas reinhardtii*CLiP2 mutant collection expands genome coverage with high-confidence disrupting alleles. *bioRxiv* 2024.12.16.626622 (2024) doi:10.1101/2024.12.16.626622.

35. Marin, B. & Melkonian, M. Mesostigmatophyceae, a new class of streptophyte green algae revealed by SSU rRNA sequence comparisons. *Protist* **150**, 399–417 (1999).

36. Bierenbroodspot, M. J. *et al.* Phylogenomic insights into the first multicellular streptophyte. *Curr. Biol.* **34**, 670-681.e7 (2024).

37. Graham, L. E., Cook, M. E. & Busse, J. S. The origin of plants: body plan changes contributing to a major evolutionary radiation. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 4535–4540 (2000).

38. Graham, L. E. Coleochaete and the origin of land plants. *Am. J. Bot.* **71**, 603 (1984).

39. Amaral Zettler, L. A. *et al.* Microbiology: eukaryotic diversity in Spain's River of Fire. *Nature* **417**, 137 (2002).

40. Procházková, L., Řezanka, T., Nedbalová, L. & Remias, D. Unicellular versus Filamentous: The Glacial Alga Ancylonema alaskana comb. et stat. nov. and Its Ecophysiological Relatedness to Ancylonema nordenskioeldii (Zygnematophyceae, Streptophyta). *Microorganisms* **9**, 1103 (2021).

41. Lewis, L. A. & Lewis, P. O. Unearthing the molecular phylodiversity of desert soil green algae (Chlorophyta). *Syst. Biol.* **54**, 936–947 (2005).

42. Hess, S. *et al.* A phylogenomically informed five-order system for the closest relatives of land plants. *Curr. Biol.* **32**, 4473-4482.e7 (2022).

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

44. Wodniok, S. *et al.* Origin of land plants: do conjugating green algae hold the key? *BMC Evol. Biol.* **11**, 104 (2011).

45. Ramanan, R., Kim, B.-H., Cho, D.-H., Oh, H.-M. & Kim, H.-S. Algae-bacteria interactions: Evolution, ecology and emerging applications. *Biotechnol. Adv.* **34**, 14–29 (2016).

46. Han, G.-Z. Origin and evolution of the plant immune system. *New Phytol.* **222**, 70–83 (2019).

47. Zilber-Rosenberg, I. & Rosenberg, E. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol. Rev.* **32**, 723–735 (2008).

48. Vandenkoornhuyse, P., Quaiser, A., Duhamel, M., Le Van, A. & Dufresne, A. The importance of the microbiome of the plant holobiont. *New Phytol.* **206**, 1196–1206 (2015).

49. Delaux, P.-M. & Schornack, S. Plant evolution driven by interactions with symbiotic and pathogenic microbes. *Science* **371**, eaba6605 (2021).

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

51. Besserer, A. *et al.* Strigolactones stimulate arbuscular mycorrhizal fungi by activating mitochondria. *PLoS Biol.* **4**, e226 (2006).

52. Genre, A. *et al.* Short-chain chitin oligomers from arbuscular mycorrhizal fungi trigger nuclear Ca2+ spiking in Medicago truncatula roots and their production is enhanced by strigolactone. *New Phytol.* **198**, 190–202 (2013).

53. Maillet, F. *et al.* Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature* **469**, 58–63 (2011).

54. Ané, J.-M. *et al.* Medicago truncatula DMI1 required for bacterial and fungal symbioses in legumes. *Science* **303**, 1364–1367 (2004).

55. Op den Camp, R. *et al.* LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume Parasponia. *Science* **331**, 909–912 (2011).

56. Lévy, J. *et al.* A putative Ca2+ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science* **303**, 1361–1364 (2004).

57. Kaló, P. *et al.* Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. *Science* **308**, 1786–1789 (2005).

58. Gobbato, E. *et al.* A GRAS-type transcription factor with a specific function in mycorrhizal signaling. *Curr. Biol.* **22**, 2236–2241 (2012).

59. Singh, S., Katzer, K., Lambert, J., Cerri, M. & Parniske, M. CYCLOPS, a DNA-binding transcriptional activator, orchestrates symbiotic root nodule development. *Cell Host Microbe* **15**, 139–152 (2014).

60. MacLean, A. M., Bravo, A. & Harrison, M. J. Plant signaling and metabolic pathways enabling arbuscular mycorrhizal symbiosis. *Plant Cell* **29**, 2319–2335 (2017).

61. Feng, F. *et al.* A combination of chitooligosaccharide and lipochitooligosaccharide recognition promotes arbuscular mycorrhizal associations in Medicago truncatula. *Nat. Commun.* **10**, 5047 (2019).

62. Jones, J. D. G. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–329 (2006).

63. Couto, D. & Zipfel, C. Regulation of pattern recognition receptor signalling in plants. *Nat. Rev. Immunol.* **16**, 537–552 (2016).

64. Zipfel, C. & Oldroyd, G. E. D. Plant signalling in symbiosis and immunity. *Nature* **543**, 328–336 (2017).

65. Yu, X., Feng, B., He, P. & Shan, L. From chaos to Harmony: Responses and signaling upon microbial pattern recognition. *Annu. Rev. Phytopathol.* **55**, 109–137 (2017).

66. Spoel, S. H. & Dong, X. How do plants achieve immunity? Defence without specialized immune cells. *Nat. Rev. Immunol.* **12**, 89–100 (2012).

67. Fu, Z. Q. & Dong, X. Systemic acquired resistance: turning local infection into global defense. *Annu. Rev. Plant Biol.* **64**, 839–863 (2013).

68. Ding, Y. *et al.* Opposite roles of salicylic acid receptors NPR1 and NPR3/NPR4 in transcriptional regulation of plant immunity. *Cell* **173**, 1454-1467.e15 (2018).

69. Bürger, M. & Chory, J. Stressed out about hormones: How plants orchestrate immunity. *Cell Host Microbe* **26**, 163–172 (2019).

70. Feng, X. *et al.* Genomes of multicellular algal sisters to land plants illuminate signaling network evolution. *Nat. Genet.* **56**, 1018–1031 (2024).

71. Jimenez Aleman, G. H., Thirumalaikumar, V. P., Jander, G., Fernie, A. R. & Skirycz, A. OPDA, more than just a jasmonate precursor. *Phytochemistry* **204**, 113432 (2022).

72. Ding, S.-W. RNA-based antiviral immunity. *Nat. Rev. Immunol.* **10**, 632–644 (2010).

73. Tiepo, A. N. *et al.* Enhanced drought tolerance in seedlings of Neotropical tree species inoculated with plant growth-promoting bacteria. *Plant Physiol. Biochem.* **130**, 277–288 (2018).

74. Xu, J. *et al.* The structure and function of the global citrus rhizosphere microbiome. *Nat. Commun.* **9**, 4894 (2018).

75. Harbort, C. J. *et al.* Root-Secreted Coumarins and the Microbiota Interact to Improve Iron Nutrition in Arabidopsis. *Cell Host Microbe* **28**, 825-837.e6 (2020).

76. Castrillo, G. *et al.* Root microbiota drive direct integration of phosphate stress and immunity. *Nature* **543**, 513–518 (2017).

77. Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T. & Singh, B. K. Plant-microbiome interactions: from community assembly to plant health. *Nat. Rev. Microbiol.* **18**, 607–621 (2020).

78. Durán, P. *et al.* Microbial Interkingdom Interactions in Roots Promote Arabidopsis Survival. *Cell* **175**, 973-983.e14 (2018).

79. Carrión, V. J. *et al.* Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. *Science* **366**, 606–612 (2019).

80. Simmons, T. *et al.* Drought Drives Spatial Variation in the Millet Root Microbiome. *Front. Plant Sci.* **11**, 599 (2020).

81. Xu, L. *et al.* Drought delays development of the sorghum root microbiome and enriches for monoderm bacteria. *Proceedings of the National Academy of Sciences* **115**, E4284–E4293 (2018).

82. Zhalnina, K. *et al.* Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat Microbiol* **3**, 470–480 (2018).

83. Bulgarelli, D., Schlaeppi, K., Spaepen, S., Ver Loren van Themaat, E. & Schulze-Lefert, P. Structure and functions of the bacterial microbiota of plants. *Annu. Rev. Plant Biol.* **64**, 807–838 (2013).

84. Wippel, K. *et al.* Host preference and invasiveness of commensal bacteria in the Lotus and Arabidopsis root microbiota. *Nat Microbiol* **6**, 1150–1162 (2021).

85. Schlaeppi, K., Dombrowski, N., Oter, R. G., Ver Loren van Themaat, E. & Schulze-Lefert, P. Quantitative divergence of the bacterial root microbiota in Arabidopsis thaliana relatives. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 585–592 (2014).

86. Bulgarelli, D. *et al.* Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature* **488**, 91–95 (2012).

87. Durán, P. The core microbiota across the green lineage. *Curr. Opin. Plant Biol.* **77**, 102487 (2024).

88. Oldroyd, G. E. D. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* **11**, 252–263 (2013).

89. Yeoh, Y. K. *et al.* Evolutionary conservation of a core root microbiome across plant phyla along a tropical soil chronosequence. *Nat. Commun.* **8**, 1–9 (2017).

90. Bonito, G. *et al.* Plant host and soil origin influence fungal and bacterial assemblages in the roots of woody plants. *Mol. Ecol.* **23**, 3356–3370 (2014).

91. Thiergart, T. *et al.* Root microbiota assembly and adaptive differentiation among European Arabidopsis populations. *Nat. Ecol. Evol.* **4**, 122–131 (2020).

92. Ma, K.-W. *et al.* Coordination of microbe-host homeostasis by crosstalk with plant innate immunity. *Nat Plants* **7**, 814–825 (2021).

93. Hacquard, S., Spaepen, S., Garrido-Oter, R. & Schulze-Lefert, P. Interplay Between Innate Immunity and the Plant Microbiota. *Annu. Rev. Phytopathol.* **55**, 565–589 (2017).

94. Yang, L. *et al.* Mechanisms of rhizosphere plant-microbe interactions: molecular insights into microbial colonization. *Front. Plant Sci.* **15**, 1491495 (2024).

95. Kundu, A., Mishra, S., Kundu, P., Jogawat, A. & Vadassery, J. Piriformospora indica recruits host-derived putrescine for growth promotion in plants. *Plant Physiol.* **188**, 2289–2307 (2022).

96. Santoyo, G. How plants recruit their microbiome? New insights into beneficial interactions. *J. Adv. Res.* **40**, 45–58 (2022).

97. Zheng, Y. *et al.* Purines enrich root-associated Pseudomonas and improve wild soybean growth under salt stress. *Nat. Commun.* **15**, 3520 (2024).

98. Liu, X. *et al.* A dual role of amino acids from Sesbania rostrata seed exudates in the chemotaxis response of Azorhizobium caulinodans ORS571. *Mol. Plant. Microbe. Interact.* **32**, 1134–1147 (2019).

99. Yang, C.-X. *et al.* Plant exudates-driven microbiome recruitment and assembly facilitates plant health management. *FEMS Microbiol. Rev.* **49**, fuaf008 (2025).

100. Sasse, J., Martinoia, E. & Northen, T. Feed your friends: Do plant exudates shape the root microbiome? *Trends Plant Sci.* **23**, 25–41 (2017).

101. Solomon, W., Janda, T. & Molnár, Z. Unveiling the significance of rhizosphere: Implications for plant growth, stress response, and sustainable agriculture. *Plant Physiol. Biochem.* **206**, 108290 (2024).

102. Getzke, F., Amine Hassani, M., Crüsemann, M., Malisic, M. & Schulze-Lefert, P. Cofunctioning of bacterial exometabolites drives root microbiota establishment. *Proceedings of the National Academy of Sciences* **120**, e2221508120 (2023).

103. Liu, Z. *et al.* A genome-wide screen identifies genes in rhizosphere-associated Pseudomonas required to evade plant defenses. *MBio* **9**, (2018).

104. Seymour, J. R., Amin, S. A., Raina, J.-B. & Stocker, R. Zooming in on the phycosphere: the ecological interface for phytoplankton-bacteria relationships. *Nat Microbiol* **2**, 17065 (2017).

105. Durán, P. *et al.* Shared features and reciprocal complementation of the Chlamydomonas and Arabidopsis microbiota. *Nat. Commun.* **13**, 1–14 (2022).

106. Raina, J.-B. *et al.* Chemotaxis shapes the microscale organization of the ocean's microbiome. *Nature* **605**, 132–138 (2022).

107. Seymour, J. R., Ahmed, T. & Stocker, R. Bacterial chemotaxis towards the extracellular products of the toxic phytoplankton Heterosigma akashiwo. *J. Plankton Res.* **31**, 1557–1561 (2009).

108. Durham, B. P. *et al.* Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 453–457 (2015).

109. Fu, H., Uchimiya, M., Gore, J. & Moran, M. A. Ecological drivers of bacterial community assembly in synthetic phycospheres. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 3656–3662 (2020).

110. Moran, M. A. *et al.* Deciphering ocean carbon in a changing world. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 3143–3151 (2016).

111. Grant, M. A. A., Kazamia, E., Cicuta, P. & Smith, A. G. Direct exchange of vitamin B12 is demonstrated by modelling the growth dynamics of algal-bacterial cocultures. *ISME J.* **8**, 1418–1427 (2014).

112. Amin, S. A. *et al.* Photolysis of iron-siderophore chelates promotes bacterial-algal mutualism. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 17071–17076 (2009).

113. Paerl, R. W. *et al.* Use of plankton-derived vitamin B1 precursors, especially thiazole-related precursor, by key marine picoeukaryotic phytoplankton. *ISME J.* **11**, 753–765 (2016).

114. Fei, C. *et al.* Quorum sensing regulates "swim-or-stick" lifestyle in the phycosphere. *Environ. Microbiol.* **22**, 4761–4778 (2020).

115. Flores, D. C. *et al.* A mutualistic bacterium rescues a green alga from an antagonist. *Proceedings of the National Academy of Sciences* **121**, e2401632121 (2024).

116. Calatrava, V. *et al.* Genetic evidence for algal auxin production in Chlamydomonas and its role in algal-bacterial mutualism. *iScience* **27**, 108762 (2024).

117. Amin, S. A. *et al.* Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522**, 98–101 (2015).

118. Beiralas, R., Ozer, N. & Segev, E. Abundant Sulfitobacter marine bacteria protect Emiliania huxleyi algae from pathogenic bacteria. *ISME Commun.* **3**, 100 (2023).

119. Huang, A. C. *et al.* A specialized metabolic network selectively modulates *Arabidopsis* root microbiota. *Science* **364**, eaau6389 (2019).

120. Cirri, E. & Pohnert, G. Algae-bacteria interactions that balance the planktonic microbiome. *New Phytol.* **223**, 100–106 (2019).

121. Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J. & Smith, A. G. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* **438**, 90–93 (2005).

122. Sasso, S., Stibor, H., Mittag, M. & Grossman, A. R. From molecular manipulation of domesticated Chlamydomonas reinhardtii to survival in nature. *Elife* **7**, (2018).

123. CC-5325 cw15 mt- [Jonikas CMJ030-JR397]. *Chlamydomonas Resource Center* https://www.chlamycollection.org/product/cc-5325-cw15-mt-jonikas-cmj030-jr397/.

124. Calatrava, V., Hom, E. F. Y., Llamas, Á., Fernández, E. & Galván, A. Nitrogen scavenging from amino acids and peptides in the model alga Chlamydomonas reinhardtii. The role of extracellular l-amino oxidase. *Algal Research* **38**, 101395 (2019).

125. Park, W.-K. *et al.* Phytohormone supplementation significantly increases growth of Chlamydomonas reinhardtii cultivated for biodiesel production. *Appl. Biochem. Biotechnol.* **171**, 1128–1142 (2013).

126. Helliwell, K. E. *et al.* Unraveling vitamin B12-responsive gene regulation in algae. *Plant Physiol.* **165**, 388–397 (2014).

127. Aiyar, P. *et al.* Antagonistic bacteria disrupt calcium homeostasis and immobilize algal cells. *Nat. Commun.* **8**, 1756 (2017).

128. Hotter, V. *et al.* A polyyne toxin produced by an antagonistic bacterium blinds and lyses a Chlamydomonad alga. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

129. Ley, R. E. *et al.* Evolution of mammals and their gut microbes. *Science* **320**, 1647–1651 (2008).

130. Bouffaud, M.-L. *et al.* Is diversification history of maize influencing selection of soil bacteria by roots? *Mol. Ecol.* **21**, 195–206 (2012).

131. Colman, D. R., Toolson, E. C. & Takacs-Vesbach, C. D. Do diet and taxonomy influence insect gut bacterial communities? *Mol. Ecol.* **21**, 5124–5137 (2012).

132. Lim, S. J. & Bordenstein, S. R. An introduction to phylosymbiosis. *Proc. Biol. Sci.* **287**, 20192900 (2020).

133. Brucker, R. M. & Bordenstein, S. R. The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus Nasonia. *Science* **341**, 667–669 (2013).

134. Kohl, K. D. Ecological and evolutionary mechanisms underlying patterns of phylosymbiosis in host-associated microbial communities. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190251 (2020).

135. Erlandson, S., Wei, X., Savage, J., Cavender-Bares, J. & Peay, K. Soil abiotic variables are more important than Salicaceae phylogeny or habitat specialization in determining soil microbial community structure. *Mol. Ecol.* **27**, 2007–2024 (2018).

136. Rudman, S. M. *et al.* Microbiome composition shapes rapid genomic adaptation of Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 20025–20032 (2019).

137. Robinson, D. F. & Foulds, L. R. Comparison of phylogenetic trees. *Math. Biosci.* **53**, 131–147 (1981).

138. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220 (1967).

139. Li, J., Wei, X., Huang, D. & Xiao, J. The phylosymbiosis pattern between the fig wasps of the same genus and their associated Microbiota. *Front. Microbiol.* **12**, 800190 (2021).

140. Kwong, W. K. *et al.* Dynamic microbiome evolution in social bees. *Sci Adv* **3**, e1600513 (2017).

141. Osborne, O. G. *et al.* Phylosymbiosis shapes skin bacterial communities and pathogen-protective function in Appalachian salamanders. *ISME J.* **18**, (2024).

142. Han, G. H. *et al.* Phylosymbiosis in seven wild fish species collected off the southern coast of Korea: Skin microbiome most strongly reflects evolutionary pressures. *Microb. Ecol.* **87**, 153 (2024).

143. Pollock, F. J. *et al.* Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny. *Nat. Commun.* **9**, 4921 (2018).

144. Bouffaud, M.-L., Poirier, M.-A., Muller, D. & Moënne-Loccoz, Y. Root microbiome relates to plant host evolution in maize and other Poaceae. *Environ. Microbiol.* **16**, 2804–2814 (2014).

145. Fitzpatrick, C. R. *et al.* Assembly and ecological function of the root microbiome across angiosperm plant species. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E1157–E1165 (2018).

146. Northen, T. R. *et al.* Community standards and future opportunities for synthetic communities in plant-microbiota research. *Nat. Microbiol.* **9**, 2774–2784 (2024).

147. Bai, Y. *et al.* Functional overlap of the Arabidopsis leaf and root microbiota. *Nature* **528**, 364–369 (2015).

148. Kremer, J. M. *et al.* Peat-based gnotobiotic plant growth systems for Arabidopsis microbiome research. *Nat. Protoc.* **16**, 2450–2470 (2021).

149. Emmenegger, B. *et al.* Identifying microbiota community patterns important for plant protection using synthetic communities and machine learning. *Nat. Commun.* **14**, 7983 (2023).

150. Kimbrel, J. A. *et al.* Host selection and stochastic effects influence bacterial community assembly on the microalgal phycosphere. *Algal Res.* **40**, 101489 (2019).

151. Kim, H. *et al.* Bacterial response to spatial gradients of algal-derived nutrients in a porous microplate. *ISME J.* **16**, 1036–1045 (2022).

152. Friedlingstein, P. *et al.* Global carbon budget 2019. *Earth Syst. Sci. Data* **11**, 1783–1838 (2019).

153. Jassey, V. E. J. *et al.* Contribution of soil algae to the global carbon cycle. *New Phytol.* **234**, 64–76 (2022).

154. Knack, J. J. *et al.* Microbiomes of streptophyte algae and bryophytes suggest that a functional suite of Microbiota fostered plant colonization of land. *Int. J. Plant Sci.* **176**, 405–420 (2015).

155. Pershina, E. *et al.* Comparative analysis of prokaryotic communities associated with organic and conventional farming systems. *PLoS One* **10**, e0145072 (2015).

156. Bourceret, A. *et al.* Maize field study reveals covaried Microbiota and metabolic changes in roots over plant growth. *MBio* **13**, e0258421 (2022).

157. Tkacz, A., Bestion, E., Bo, Z., Hortala, M. & Poole, P. S. Influence of plant fraction, soil, and plant species on Microbiota: A multikingdom comparison. *MBio* **11**, (2020).

158. Chen, Y. *et al.* Soil fungal communities show more specificity than bacteria for plant species composition in a temperate forest in China. *BMC Microbiol.* **22**, 208 (2022).

159. Jackrel, S. L., Yang, J. W., Schmidt, K. C. & Denef, V. J. Host specificity of microbiome assembly and its fitness effects in phytoplankton. *ISME J.* **15**, 774–788 (2021).

160. Eigemann, F., Hilt, S., Salka, I. & Grossart, H.-P. Bacterial community composition associated with freshwater algae: species specificity vs. dependency on environmental conditions and source community. *FEMS Microbiol. Ecol.* **83**, 650–663 (2013).

161. Fracchia, F. *et al.* Colonization of naïve roots from Populus tremula x alba involves successive waves of fungi and bacteria with different trophic abilities. *Appl. Environ. Microbiol.* **87**, e02541-20 (2021).

162. Medlin, L., Elwood, H. J., Stickel, S. & Sogin, M. L. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* **71**, 491–499 (1988).

163. Marin, B., Klingberg, M. & Melkonian, M. Phylogenetic relationships among the Cryptophyta: Analyses of nuclear-encoded SSU rRNA sequences support the monophyly of extant Plastid-containing lineages. *Protist* **149**, 265–276 (1998).

164. Fay, M. F., Swensen, S. M. & Chase, M. W. Taxonomic Affinities of Medusagyne oppositifolia (Medusagynaceae). *Kew Bull.* **52**, 111 (1997).

165. Manhart, J. R. Phylogenetic analysis of green plant rbcL sequences. *Mol. Phylogenet. Evol.* **3**, 114–127 (1994).

166. McCourt, R. M. *et al.* PHYLOGENY OF THE CONJUGATING GREEN ALGAE (ZYGNEMOPHYCEAE) BASED ON rbc L SEQUENCES. *J. Phycol.* **36**, 747–758 (2000).

167. Gontcharov, A. A., Marin, B. & Melkonian, M. Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematophyceae (Streptophyta). *Mol. Biol. Evol.* **21**, 612–624 (2004).

168. Okonechnikov, K., Golosova, O., Fursov, M. & UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**, 1166–1167 (2012).

169. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

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

171. Caesar, J., Tamm, A., Ruckteschler, N., Leifke, A. L. & Weber, B. Revisiting chlorophyll extraction methods in biological soil crusts – methodology for determination of chlorophyll *a* and chlorophyll *a* + *b* as compared to previous methods. *Biogeosciences* **15**, 1415–1424 (2018).

172. Chelius, M. K. & Triplett, E. W. The Diversity of Archaea and Bacteria in Association with the Roots of Zea mays L. *Microb Ecol* **41**, 252–263 (2001).

173. Beckers, B. *et al.* Performance of 16s rDNA primer pairs in the study of rhizosphere and endosphere bacterial microbiomes in metabarcoding studies. *Front. Microbiol.* **7**, 650 (2016).

174. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).

175. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* **11**, e0163962 (2016).

176. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).

177. Yilmaz, P. *et al.* The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* **42**, D643-8 (2014).

178. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590-6 (2013).

179. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2024).

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

181. Abarenkov, K. *et al.* The UNITE database for molecular identification and taxonomic communication of fungi and other eukaryotes: sequences, taxa and classifications reconsidered. *Nucleic Acids Res.* **52**, D791–D797 (2024).

182. Robeson, M. S., 2nd *et al.* RESCRIPt: Reproducible sequence taxonomy reference database management. *PLoS Comput. Biol.* **17**, e1009581 (2021).

183. Community Ecology Package [R package vegan version 2.7-2]. *Comprehensive R Archive Network (CRAN)* https://CRAN.R-project.org/package=vegan (2025).

184. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).

185. Grafen, A. The phylogenetic regression. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **326**, 119–157 (1989).

186. Hadziavdic, K. *et al.* Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS One* **9**, e87624 (2014).

187. Hayden, H. S. & Waaland, J. R. PHYLOGENETIC SYSTEMATICS OF THE ULVACEAE (ULVALES, ULVOPHYCEAE) USING CHLOROPLAST AND NUCLEAR DNA SEQUENCES[1]. *J. Phycol.* **38**, 1200–1212 (2002).

188. Wimalanathan, K. & Lawrence-Dill, C. J. Gene Ontology Meta Annotator for Plants (GOMAP). *Plant Methods* **17**, 1–14 (2021).

189. Powers, T. Ribosome biogenesis: giant steps for a giant problem. *Cell* vol. 119 901–902 (2004).

190. Shore, D. & Albert, B. Ribosome biogenesis and the cellular energy economy. *Curr. Biol.* **32**, R611–R617 (2022).

191. Lee, J.-H., Munz, J., Dharmasiri, H. N., Jin, E. & Joo, S. Dissecting nitrogen starvation signaling in Chlamydomonas: Insights from arginine-fed transcriptome profiling. *Algal Res.* **85**, 103848 (2025).

192. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

193. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178-86 (2012).

194. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol.* **428**, 726–731 (2016).

195. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).

196. Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).

197. Bollig, K. *et al.* Structural analysis of linear hydroxyproline-bound O-glycans of Chlamydomonas reinhardtii--conservation of the inner core in Chlamydomonas and land plants. *Carbohydr. Res.* **342**, 2557–2566 (2007).

198. Poulhazan, A. *et al.* Molecular-level architecture of Chlamydomonas reinhardtii's glycoprotein-rich cell wall. *Nat. Commun.* **15**, 986 (2024).

199. Kong, F. *et al.* Interorganelle communication: Peroxisomal MALATE DEHYDROGENASE2 connects lipid catabolism to photosynthesis through redox coupling in Chlamydomonas. *Plant Cell* **30**, 1824–1847 (2018).

200. Kong, F. *et al.* Chlamydomonas carries out fatty acid β-oxidation in ancestral peroxisomes using a bona fide acyl-CoA oxidase. *Plant J.* **90**, 358–371 (2017).

201. Zhou, X. R. *et al.* Distribution dynamics and roles of starch in non-photosynthetic vegetative organs of Santalum album Linn., a hemiparasitic tree. *Front. Plant Sci.* **11**, 532537 (2020).

202. Koo, K. M. *et al.* The mechanism of starch over-accumulation in Chlamydomonas reinhardtii high-starch mutants identified by comparative transcriptome analysis. *Front. Microbiol.* **8**, 858 (2017).

203. Hildebrandt, T. M., Nunes Nesi, A., Araújo, W. L. & Braun, H.-P. Amino Acid Catabolism in Plants. *Mol. Plant* **8**, 1563–1579 (2015).

204. Liang, Y. *et al.* Branched-chain amino acid catabolism impacts triacylglycerol homeostasis in Chlamydomonas reinhardtii. *Plant Physiol.* **179**, 1502–1514 (2019).

205. Kropat, J. *et al.* A revised mineral nutrient supplement increases biomass and growth rate in Chlamydomonas reinhardtii. *Plant J.* **66**, 770–780 (2011).

206. Catalanotti, C., Yang, W., Posewitz, M. C. & Grossman, A. R. Fermentation metabolism and its evolution in algae. *Front. Plant Sci.* **4**, 150 (2013).

207. Plancke, C. *et al.* Lack of isocitrate lyase in Chlamydomonas leads to changes in carbon metabolism and in the response to oxidative stress under mixotrophic growth. *Plant J.* **77**, 404–417 (2014).

208. Miller, R. *et al.* Changes in transcript abundance in Chlamydomonas reinhardtii following nitrogen deprivation predict diversion of metabolism. *Plant Physiol.* **154**, 1737–1752 (2010).

209. Villa, E., Ali, E. S., Sahu, U. & Ben-Sahra, I. Cancer cells tune the signaling pathways to empower de Novo synthesis of nucleotides. *Cancers (Basel)* **11**, 688 (2019).

210. Ali, E. S. & Ben-Sahra, I. Regulation of nucleotide metabolism in cancers and immune disorders. *Trends Cell Biol.* **33**, 950–966 (2023).

211. Tran, D. H. *et al.* De novo and salvage purine synthesis pathways across tissues and tumors. *Cell* **187**, 3602-3618.e20 (2024).

212. Chua, S. M. & Fraser, J. A. Surveying purine biosynthesis across the domains of life unveils promising drug targets in pathogens. *Immunol. Cell Biol.* **98**, 819–831 (2020).

213. Buey, R. M. *et al.* Guanine nucleotide binding to the Bateman domain mediates the allosteric inhibition of eukaryotic IMP dehydrogenases. *Nat. Commun.* **6**, 8923 (2015).

214. Witte, C.-P. & Herde, M. Nucleotide metabolism in plants. *Plant Physiol.* **182**, 63–78 (2020).

215. Valvezan, A. J. *et al.* MTORC1 couples nucleotide synthesis to nucleotide demand resulting in a targetable metabolic vulnerability. *Cancer Cell* **32**, 624-638.e5 (2017).

216. Missbach, S. *et al.* 40S ribosome biogenesis co-factors are essential for gametophyte and embryo development. *PLoS One* **8**, e54084 (2013).

217. Huang, C.-K. *et al.* A DEAD-box protein, AtRH36, is essential for female gametophyte development and is involved in rRNA biogenesis in Arabidopsis. *Plant Cell Physiol.* **51**, 694–706 (2010).

218. Daugeron, M. C., Kressler, D. & Linder, P. Dbp9p, a putative ATP-dependent RNA helicase involved in 60S-ribosomal-subunit biogenesis, functionally interacts with Dbp6p. *RNA* **7**, 1317–1334 (2001).

219. Rouquette, J., Choesmel, V. & Gleizes, P.-E. Nuclear export and cytoplasmic processing of precursors to the 40S ribosomal subunits in mammalian cells. *EMBO J.* **24**, 2862–2872 (2005).

220. Geerlings, T. H., Faber, A. W., Bister, M. D., Vos, J. C. & Raué, H. A. Rio2p, an evolutionarily conserved, low abundant protein kinase essential for processing of 20 S Pre-rRNA in Saccharomyces cerevisiae. *J. Biol. Chem.* **278**, 22537–22545 (2003).

221. Zemp, I. *et al.* Distinct cytoplasmic maturation steps of 40S ribosomal subunit precursors require hRio2. *J. Cell Biol.* **185**, 1167–1180 (2009).

222. Strunk, B. S., Novak, M. N., Young, C. L. & Karbstein, K. A translation-like cycle is a quality control checkpoint for maturing 40S ribosome subunits. *Cell* **150**, 111–121 (2012).

223. Read, R. D. *et al.* A kinome-wide RNAi screen in Drosophila Glia reveals that the RIO kinases mediate cell proliferation and survival through TORC2-Akt signaling in glioblastoma. *PLoS Genet.* **9**, e1003253 (2013).

224. Liu, T. *et al.* Phosphorylation of right open reading frame 2 (Rio2) protein kinase by polo-like kinase 1 regulates mitotic progression. *J. Biol. Chem.* **286**, 36352–36360 (2011).

225. Babbio, F., Farinacci, M., Saracino, F., Carbone, M. L. A. & Privitera, E. Expression and localization studies of hSDA, the human ortholog of the yeast SDA1 gene. *Cell Cycle* **3**, 486–490 (2004).

226. Filipowicz, W. & Pogacić, V. Biogenesis of small nucleolar ribonucleoproteins. *Curr. Opin. Cell Biol.* **14**, 319–327 (2002).

227. Weinstein, L. B. & Steitz, J. A. Guided tours: from precursor snoRNA to functional snoRNP. *Curr. Opin. Cell Biol.* **11**, 378–384 (1999).

228. Eisinger, D. P., Dick, F. A., Denke, E. & Trumpower, B. L. SQT1, which encodes an essential WD domain protein of Saccharomyces cerevisiae, suppresses dominant-negative mutations of the ribosomal protein gene QSR1. *Mol. Cell. Biol.* **17**, 5146–5155 (1997).

229. Weis, B. L., Palm, D., Missbach, S., Bohnsack, M. T. & Schleiff, E. atBRX1-1 and atBRX1-2 are involved in an alternative rRNA processing pathway in Arabidopsis thaliana. *RNA* **21**, 415–425 (2015).

230. Sikorska, N., Zuber, H., Gobert, A., Lange, H. & Gagliardi, D. RNA degradation by the plant RNA exosome involves both phosphorolytic and hydrolytic activities. *Nat. Commun.* **8**, 2162 (2017).

231. Lange, H. & Gagliardi, D. Catalytic activities, molecular connections, and biological functions of plant RNA exosome complexes. *Plant Cell* **34**, 967–988 (2022).

232. Dragon, F. *et al.* A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* **417**, 967–970 (2002).

233. Bang, W. Y. *et al.* AtObgC, a plant ortholog of bacterial Obg, is a chloroplast-targeting GTPase essential for early embryogenesis. *Plant Mol. Biol.* **71**, 379–390 (2009).

234. CLiP. https://www.chlamylibrary.org/showGene?geneIdentifier=Cre10.g442650.

235. Locus: AT2G03270. https://www.arabidopsis.org/locus?name=AT2G03270#.

236. Hang, R. *et al.* Protein arginine methyltransferase 3 fine-tunes the assembly/disassembly of pre-ribosomes to repress nucleolar stress by interacting with RPS2B in arabidopsis. *Mol. Plant* **14**, 223–236 (2021).

237. Wang, Z. *et al.* AtPRMT3-RPS2B promotes ribosome biogenesis and coordinates growth and cold adaptation trade-off. *Nat. Commun.* **15**, 8693 (2024).

238. Tong, M. & Jiang, Y. FK506-binding proteins and their diverse functions. *Curr. Mol. Pharmacol.* **9**, 48–65 (2015).

239. Feng, J., Chen, D., Berr, A. & Shen, W.-H. ZRF1 chromatin regulators have Polycomb silencing and independent roles in development. *Plant Physiol.* **172**, 1746–1759 (2016).

240. Tsering, T. *et al.* HSP90 differentially stabilizes plant ABCB-type auxin transporters on the plasma membrane. *Nat. Commun.* **16**, 8643 (2025).

241. Guzmán-López, J. A., Abraham-Juárez, M. J., Lozano-Sotomayor, P., de Folter, S. & Simpson, J. Arabidopsis thaliana gonidialess A/Zuotin related factors (GlsA/ZRF) are essential for maintenance of meristem integrity. *Plant Mol. Biol.* **91**, 37–51 (2016).

242. Kim, J.-G. *et al.* Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 7888–7893 (2016).

243. Drechsel, H., Thieken, A., Reissbrodt, R., Jung, G. & Winkelmann, G. Alpha-keto acids are novel siderophores in the genera Proteus, Providencia, and Morganella and are produced by amino acid deaminases. *J. Bacteriol.* **175**, 2727–2733 (1993).

244. Florencio, F. J. & Vega, J. M. Utilization of nitrate, nitrite and ammonium by Chlamydomonas reinhardii : Photoproduction of ammonium: Photoproduction of ammonium. *Planta* **158**, 288–293 (1983).

245. Fernandez, E. & Galvan, A. Inorganic nitrogen assimilation in Chlamydomonas. *J. Exp. Bot.* **58**, 2279–2287 (2007).

246. Sanz-Luque, E., Chamizo-Ampudia, A., Llamas, A., Galvan, A. & Fernandez, E. Understanding nitrate assimilation and its regulation in microalgae. *Front. Plant Sci.* **6**, 899 (2015).

247. Joshi, K. *et al.* Spatial organization of putrescine synthesis in plants. *Plant Sci.* **349**, 112232 (2024).

248. Sobieszczuk-Nowicka, E. *et al.* Polyamines - A new metabolic switch: Crosstalk with networks involving senescence, crop improvement, and mammalian cancer therapy. *Front. Plant Sci.* **10**, 859 (2019).

249. Polyamines: double agents in disease and plant immunity. *Trends in Plant Science* **26**, 1061–1071 (2021).

250. Copeland, C. The feeling is mutual: Increased host putrescine biosynthesis promotes both plant and endophyte growth. *Plant Physiol.* **188**, 1939–1941 (2022).

251. de Souza, R. S. C., Armanhi, J. S. L., Damasceno, N. de B., Imperial, J. & Arruda, P. Genome sequences of a plant beneficial synthetic bacterial community reveal genetic features for successful plant colonization. *Front. Microbiol.* **10**, 1779 (2019).

252. Dündar, E. & Bush, D. R. BAT1, a bidirectional amino acid transporter in Arabidopsis. *Planta* **229**, 1047–1056 (2009).

253. Kirk, D. L. & Kirk, M. M. Carrier-mediated Uptake of Arginine and Urea by Chlamydomonas reinhardtii. *Plant Physiol.* **61**, 556–560 (1978).

254. Strijkert, P. J., Loppes, R. & Sussenbach, J. S. Arginine metabolism in Chlamydomonas reinhardi. Regulation of uptake and breakdown. *FEBS Lett.* **14**, 329–332 (1971).

255. Oku, S., Komatsu, A., Tajima, T., Nakashimada, Y. & Kato, J. Identification of chemotaxis sensory proteins for amino acids in Pseudomonas fluorescens Pf0-1 and their involvement in chemotaxis to tomato root exudate and root colonization. *Microbes Environ.* **27**, 462–469 (2012).

256. Patidar, S. K. Metabolic interactions between microalgae and bacteria: Multifunctional ecological interplay and environmental applications. *Algal Res.* **86**, 103904 (2025).

257. Phillips, D. A., Fox, T. C., King, M. D., Bhuvaneswari, T. V. & Teuber, L. R. Microbial products trigger amino acid exudation from plant roots. *Plant Physiol.* **136**, 2887–2894 (2004).

258. Muñoz-Blanco, J., Hidalgo-Martínez, J. & Cárdenas, J. Extracellular deamination of L-amino acids by Chlamydomonas reinhardtii cells. *Planta* **182**, 194–198 (1990).

259. Vallon, O., Bulté, L., Kuras, R., Olive, J. & Wollman, F. A. Extensive accumulation of an extracellular L-amino-acid oxidase during gametogenesis of Chlamydomonas reinhardtii. *Eur. J. Biochem.* **215**, 351–360 (1993).

260. Chesneau, G., Herpell, J., Wolf, S. M., Perin, S. & Hacquard, S. MetaFlowTrain: a highly parallelized and modular fluidic system for studying exometabolite-mediated inter-organismal interactions. *Nat. Commun.* **16**, 3310 (2025).

261. Torres, M. J. *et al.* Chlamydomonas-Methylobacterium oryzae cooperation leads to increased biomass, nitrogen removal and hydrogen production. *Bioresour. Technol.* **352**, 127088 (2022).

262. J. F. G. M. Wintermans, A. D. M. Spectrophotometric characteristics of chlorophylls a and b and their phenophytins in ethanol. *Biochimica et Biophysica Acta (BBA) - Biophysics including Photosynthesis* **109**, 448–453 (1965).

263. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).

264. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).

265. Zhang, P., Spaepen, S., Bai, Y., Hacquard, S. & Garrido-Oter, R. Rbec: a tool for analysis of amplicon sequencing data from synthetic microbial communities. *ISME Communications* **1**, 1–3 (2021).

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

267. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* **4**, 1521 (2015).

268. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

269. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2019).

270. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

271. Gu, Z. Complex heatmap visualization. *Imeta* **1**, (2022).

272. Alexa A, R. J. *TopGO: Enrichment Analysis for Gene Ontology*. (2023).

273. AnnotationForge. *Bioconductor* http://bioconductor.org/packages/AnnotationForge/.

274. Yu, G. Gene ontology semantic similarity analysis using GOSemSim. *Methods Mol. Biol.* **2117**, 207–215 (2020).

275. Sayols, S. rrvgo: a Bioconductor package for interpreting lists of Gene Ontology terms. *MicroPubl. Biol.* **2023**, (2023).

276. Schlicker, A., Domingues, F. S., Rahnenführer, J. & Lengauer, T. A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics* **7**, 302 (2006).

277. Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10**, 1–17 (2009).

278. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021).

279. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* **16**, 284–287 (2012).

280. Luo, W. & Brouwer, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831 (2013).

281. Cheng, C.-Y. *et al.* Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J.* **89**, 789–804 (2017).

282. Mun, T., Bachmann, A., Gupta, V., Stougaard, J. & Andersen, S. U. Lotus Base: An integrated information portal for the model legume Lotus japonicus. *Sci. Rep.* **6**, 39447 (2016).

283. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Preprint at https://ggplot2.tidyverse.org (2016).

284. CLiP. https://www.chlamylibrary.org/allMutants.

285. Schnell, R. A. & Lefebvre, P. A. Isolation of the Chlamydomonas regulatory gene NIT2 by transposon tagging. *Genetics* **134**, 737–747 (1993).

286. Pollock, S. V., Colombo, S. L., Prout, D. L., Jr, Godfrey, A. C. & Moroney, J. V. Rubisco activase is required for optimal photosynthesis in the green alga Chlamydomonas reinhardtii in a low-CO(2) atmosphere. *Plant Physiol.* **133**, 1854–1861 (2003).

287. Coleman, J. R., Berry, J. A., Togasaki, R. K. & Grossman, A. R. Identification of Extracellular Carbonic Anhydrase of Chlamydomonas reinhardtii. *Plant Physiol.* **76**, 472–477 (1984).

288. Coleman, J. R. & Grossman, A. R. Biosynthesis of carbonic anhydrase in Chlamydomonas reinhardtii during adaptation to low CO(2). *Proc. Natl. Acad. Sci. U. S. A.* **81**, 6049–6053 (1984).

289. Spalding, M. H., Spreitzer, R. J. & Ogren, W. L. Carbonic Anhydrase-Deficient Mutant of Chlamydomonas reinhardii Requires Elevated Carbon Dioxide Concentration for Photoautotrophic Growth. *Plant Physiol.* **73**, 268–272 (1983).

290. Gorman, D. S. & Levine, R. P. Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of Chlamydomonas reinhardi. *Proc. Natl. Acad. Sci. U. S. A.* **54**, 1665–1669 (1965).

291. Zipfel, C. *et al.* Bacterial disease resistance in Arabidopsis through flagellin perception. *Nature* **428**, 764–767 (2004).

292. Carbon dioxide now more than 50% higher than pre-industrial levels. https://www.noaa.gov/news-release/carbon-dioxide-now-more-than-50-higher-than-pre-industrial-levels.

293. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).