

A comparative analysis of Receptor-Like Kinases in *Chlorophyta* reveals the presence of putative Cell Wall Integrity sensors

Running title: Receptor-Like Kinases in *Chlorophyta*

Demetrio Marcianò^{1,2#}, Bastien G. Dauphin^{1#}, Fabian Basso³, Christiane Funk^{2*} and Laura Bacete^{1,4*}

¹ Umeå Plant Science Centre (UPSC), Department of Plant Physiology, Umeå University, 901 87 Umeå, Sweden

² Department of Chemistry, Umeå University, 901 87 Umeå, Sweden

³ School of Computation, Information and Technology, Technische Universität München, D-80333 München, Germany

⁴ Institute for Biology, Faculty of Natural Sciences, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway

[#]Demetrio Marcianò and Bastien G. Dauphin contributed equally to this article.

* Authors to whom correspondence should be addressed: laura.bacete@umu.se; christiane.funk@umu.se;

Abstract (247 words)

Receptor-like kinases (RLKs) perceive extracellular signals from other organisms as well as from the individual itself, triggering responses essential for growth and adaptation. Among these self-derived signals, a relevant role is sensing cell wall integrity (CWI), whereby changes in cell wall structure or composition activate cellular responses critical for development and defense. While RLKs are well characterized in vascular plants, their diversity and function remain poorly explored in green microalgae (*Chlorophyta*) photosynthetic organisms important for global oxygen production and carbon cycling. Their varied cell wall structures make them a useful system for studying the evolutionary origins of cell wall integrity sensing. In the present study, we employed advanced bioinformatics tools and AI-driven algorithms to map RLK distribution across 34 *Chlorophyta* species, characterize their structural and functional properties, and investigate their relationship with known plant counterparts. We identified 736 putative RLKs, significantly expanding the known inventory in microalgae. Structural analyses revealed a diverse array of extracellular domains, many resembling motifs associated with plant CWI sensors. These included domains linked to protein-protein interaction (e.g. LRR, PAN, ARM), cell wall remodeling (e.g. glycosyl hydrolases, lyases and carbohydrate-binding domains) and extracellular matrix mechanosensing (e.g. LPXTG, Fibronectin). This domain diversity indicates that sophisticated mechanisms for extracellular sensing and CWI monitoring may have been established early in evolution. Our findings provide valuable insights into the evolutionary diversification of RLKs and create a novel framework for future studies exploring their functional roles in algae and evolutionary links to conservation in vascular plant signaling pathways.

Keywords: Microalgae signal transduction; Receptor functional divergence; Algae-plants evolutionary conservation

Introduction

Cells respond to changing conditions by sensing extracellular signals that come from the environment, from other organisms, or from neighboring cells. Such cues report on structural integrity and guide rapid responses. Chemical or mechanical stimuli outside the plasma membrane are processed by membrane-bound kinase receptors (Tör et al., 2009; Lemmon and Schlessinger, 2010; Bhatla, 2018). To integrate this information, extracellular pattern-recognition receptors (PRRs) occur in large numbers (Lemmon and Schlessinger, 2010) in plants (Liu et al., 2024), whereas far fewer have been catalogued in other eukaryotes—even though those organisms share similar need to perceive extracellular cues. These proteins often share a tripartite layout: an extracellular domain (ectodomain), which perceives environmental stimuli, a single transmembrane domain (TM), and an intracellular (endodomain) domain having kinase features. In plants, the endodomain contains a generally conserved Ser/Thr/Tyr kinase domain, whereas the ectodomain is highly variable and can bind proteins, polysaccharides, or hormones (Dievart et al., 2020).

In vascular plants, this basic layout expanded into three major classes of extracellular PRRs: receptor-like kinases (RLKs; ectodomain–TM–endodomain), receptor-like proteins (RLPs; ectodomain–TM), and receptor proteins (RPs; ectodomain linked by a GPI anchor or secreted) (Del Hierro et al., 2021). *Arabidopsis thaliana* encodes over 610 PRRs divided into 14 ectodomain-defined families (Liu et al., 2024), and similar expansions occur in rice, maize (Shiu et al., 2004; Song et al., 2015; Yan et al., 2023) and *Fabaceae* (legumes) (Restrepo-Montoya et al., 2020). A recent computational study of *Arabidopsis* PRRs found that over 50% of the identified receptors have glycan-binding ectodomains, with many belonging to RLK subclasses linked to cell wall integrity domains; WI) surveillance such as wall-associated kinases (WAKs), *Catharanthus roseus* receptor-like kinases (CrRLK1Ls), malectin-containing RLKs and lysin motif (LysM) type RLKs (Del Hierro et al., 2021).

The cell wall—a dynamic matrix of polysaccharides and proteins—adds an extra layer of complexity to extracellular perception. Acting as the first barrier to abiotic and biotic stress, the cell wall is tightly integrated with signaling networks that safeguard its integrity (Bacete et al., 2018; Wolf, 2022). RLKs contribute to this integration either through sensing damage-derived fragments, or directly sensing force, although information on the latter is scarce (Bacete and

Hamann, 2020; Oelmüller et al., 2023). The cell wall is thought to have appeared before terrestrialization as the result of multiple events (Fangel et al., 2024) and diverse compositions and structures can already be observed in *Chlorophyta* (Popper and Tuohy, 2010; Domozych et al., 2012; Spain and Funk, 2022), suggesting that a varied set of CWI-related receptors arose early (Baez et al., 2022). Mechanical and chemical inputs are weighed differently in water and on land: aquatic algae resist hydrodynamic shear, whereas terrestrial plants contend chiefly with turgor-driven tension, a contrast that probably favored this early diversification of CWI sensors. Green algae therefore provide a tractable platform for examining how RLK structure and function relate to CWI perception, and we hypothesize that PRR diversification was indeed driven by changes in cell-wall architecture.

Evolutionary evidence supports this view: homologues of RLK genes are already present in chlorophyte and charophyte algae (Dievart et al., 2020), and the shift from unicellular to colonial forms in the Volvocales appears to have required new wall-sensing pathways. For example, in the colonial alga *Volvox carteri* a secreted glycoprotein pheromone, together with homologous extracellular-matrix glycoproteins, coordinates sexual development—linking cell wall remodeling in the unicellular-to-colonial transition to specialized wall-sensing pathways (Hallmann et al., 1998). Yet, first studies have only identified limited numbers of RLKs in species such as *Chlamydomonas reinhardtii* or *Ostreococcus tauri* (Derelle et al., 2006; Merchant et al., 2007; Lehti-Shiu et al., 2009) leaving open the question of how algae sense external cues (Vukašinović et al., 2023). Interestingly, (Yin et al., 2024) recently predicted RLKs across multiple *Viridiplantae* species and reported significantly higher number in few algae species. This improvement was partly possible due to the recent development of DeepTMHM, a machine-learning based tool, which predicts signal peptide and transmembrane domain with improved accuracy (Hallgren et al., 2022). Additionally, current advancements in genome annotation, protein domain recognition and structural prediction allow deeper insight into RLK identification and functional annotation.

Experimental characterization of algal RLKs remains slow because many species lack reliable transformation systems and functional redundancy can mask phenotypes; robust *in silico* pipelines therefore provide the principal route to discovery at present. Against this backdrop, we analyzed proteomes from 34 species and applied a similar bioinformatic pipeline to predict 736 RLKs – increasing fivefold the previous counts and exposing a rich, previously hidden receptor landscape.

Although some assemblies are still incomplete and may under-represent gene families, our data nevertheless show that chlorophytes harbor every major CWI-associated RLK subclass recognized in land plants. Moreover, nine relevant species were examined in detail, predicting the structural and functional features of each ectodomain and providing clues on their putative function. We highlight unpredicted members of key RLK families and discuss their putative role as extracellular sensors, thus presenting a new picture for RLK distribution in *Chlorophyta*. Our results link unicellular lifestyles, emerging multicellularity and wall monitoring, and they provide a curated resource that will streamline future functional investigations of algal and plant RLKs.

Results

1. From sequence to signal: A bioinformatic pipeline for enhanced RLK discovery and characterization in green algae

As described in the Introduction, RLKs constitute the largest PRR class, including every receptor subtype currently linked to CWI signaling, yet remain sparsely catalogued in *Chlorophyta*; we therefore centered our efforts on this division. Here we define an RLK as a protein that contains (i) a conserved intracellular Ser/Thr/Tyr kinase endodomain, (ii) a single transmembrane helix, and (iii) an extracellular ectodomain. Using the bioinformatic pipeline depicted in **Figure 1A**, we investigated the diversity and evolution of RLK across *Chlorophyta* phylum. Overall, all available proteomes in UniProt were selected, which covered 34 species. (see methods for further details) (**Table 1, Figure 1B**). Among these, we identified 12043 sequences containing at least one kinase domain corresponding to the InterPro entry IPR011009 (Protein kinase-like domain superfamily) (**Table 1**)

The prediction of signal peptides (SPs) and transmembrane domains (TMDs) was performed using DeepTMHMM. Signal peptides ranged from 10 to 60 amino acids (aa), with an average length of 27 aa. Additionally, the identified TMD spanned 10 to 27 residues, with an average length of 20. Alanine (24.05%), Valine (21.93%), Leucine (18.32%) and Glycine (12.52) were the most represented residues in TMDs. Combining these data with kinase domains representation, we identified 736 unique sequences matching the criteria for RLK classification, increasing the overall number of RLKs identified compared to previous studies, up from 137 (**Figure 1C; Supplemental**

table 1). Most notably, our data presents significant progress in some species for which past predictions only highlighted a few RLKs (**Figure 1C; Supplemental table 1**). The greatest improvements were in *Volvox africanus* (56 RLKs) and *Edaphochlamys debaryana* (73 RLKs), species previously reported to contain fewer than ten receptors.

Given the high diversity of RLKs ectodomains, we aimed at mapping the functional domains in putative RLKs using a subset of nine *Chlorophyta* species (**Table 1, in bold**). These were selected based on prior scientific studies, industrial relevance, lifestyle, and phylogenetic diversity in order to have a comprehensive picture of the *Chlorophyta* phylum. We used pLM-BLAST—a tool particularly sensitive to divergent sequences and therefore able to highlight previously undiscovered homologous relationships— (Kaminski et al., 2023) to uncover distant sequence homology, AlphaFold3 (Abramson et al., 2024) to produce full-length structural models, and Foldseek (van Kempen et al., 2024) to match these models with solved protein structures. The latter is particularly relevant, since structural features usually display a higher conservation than their aminoacid sequence may suggest (Illergård et al., 2009), a particularity already observed in *A. thaliana* RLK intraspecific variation (Wei et al., 2022).

2. Functional annotations highlight sequence and domain diversity with the presence of known RLKs domains

Using the aforementioned approaches, we successfully identified and mapped known motifs on 94% (333/ 353) of the predicted RLKs (**Supplemental table 2 and Supplemental material 1**). These were classified in seven functional groups according to their known biological role (i.e. (i) Carbohydrate active enzyme, (ii) Cell wall anchoring/adhesion, (iii) Signaling protein/Receptor like, (iv) Proteases, (v) Protein-protein/Protein-peptide interaction, (vi) Carbohydrate binding domains, (vii) Other domains) (**Figure 2A; Supplemental table 3; Supplemental table 4**). Most annotated regions matched plant motifs or structure, however regions matching with organisms outside *Viridiplantae* group were kept if relevant for the extracellular matrix sensing (such as “fibronectin” or “ankyrin”). Annotations having wrong or misleading relationship with intracellular processes and/or unrelated proteins were manually discarded (e.g. “blood clotting”, “ribosomal”, “DNA transcription”). Notably, a similar (~10%) proportion of discarded regions were found across species, thus showing consistency in our pipeline and in the manual refinement (**Supplemental table 5**). Overall, annotated regions belonging to each of the seven groups were

identified in most species. However, *Ostreococcus tauri* lacked protein-protein interaction motifs, probably due to the low number (two) of predicted RLKs in this species (**Figure 1C**).

The presence and distribution of annotated regions across the seven functional groups exhibited high variation in a species dependent manner (**Figure 2A**). For instance, we found significantly more motifs related to protein-protein interaction in *Chlorella vulgaris* than any other species (**Figure 2A**).

Overall, almost a third (32.6%) of the annotated regions presented motifs related to Carbohydrate-Active Enzymes (CAZymes) (**Figure 2A**). This is supported by the large number lyases related motifs, particularly active on pectins (**Figure 2B, Supplementary information 2**). Additionally, besides the protein-protein interaction group, the cell adhesion and signaling groups were the second and third main relevant, thus supporting their putative role as membrane bound sensors of the extracellular environment.

Despite lower abundance (7.3%), we further investigated the distribution of carbohydrate binding motifs as it is crucial for a cell wall sensor. We found several signaling related binding modules such as lectin, jacalin or LysM as well as general carbohydrate binding modules (CBM) and pectin binding domain (**Figure 2C**). Although lectin was the most represented group, the distribution of each binding domain differed greatly across species (**Figure 2C**).

In many cases, a single protein showed the presence of different domains belonging to multiple groups on the same annotated region (**Table 2**). To better highlight the functional crossovers between annotated regions we created a network plot (**Figure 3**). Notably, we observed a strong connection between the “cell adhesion/anchoring” group and the “carbohydrate active enzymes” one. Unsurprisingly, a similar connection was found between the latter and the “Protease” group. Moreover, we also identified an enrichment in the connection between “carbohydrate binding domains” and the “signaling protein/receptor-like” group, fitting with the model of already known RLK signaling pathways.

3. Structural regions involved in plant RLK signaling are relatively well conserved in *C. vulgaris*

Among the nine focal chlorophytes, *C. vulgaris* stands out: 36 RLKs were recovered, and 13 of them contain leucine-rich repeats (LRRs); Foldseek predicts that 12 ectodomains (~33 %) adopt

folds closely matching LRR-RLKs from *A. thaliana* (Figure 4A). This unusually high proportion of recognizable plant motifs, together with the availability of a high-quality genome and the species' biotechnological relevance, prompted us to use *C. vulgaris* as a detailed case study for RLKs and CWI receptors.

For instance, the protein A0A9D4TQY7 exhibited the highest similarity with the plant receptor TDR (PBD entry: 5gij), an LRR-RLK able to perceive CLE peptide during vascular development (Morita et al., 2016). Although the two sequences displayed only 21% homology in their ectodomain regions (28.16% if we consider the whole sequences), the structural alignment revealed a high similarity (Fig. 4A) that was also endorsed by the pLM-BLAST analysis identifying LRR motifs approximately spanning the same region (Fig. 4B). This convergence on a CLE-receptor fold indicates that peptide-binding architectures, central to CWI and developmental pathways in land plants, could already be present in green algae.

Beside the LRR domain, a total of 13 out of 36 RLKs (~36%) featured lectin-type ectodomains, including galactose- and fucose-binding lectins, LysM domains, and cellulose-binding modules. Additional notable domains included PAN (9 occurrences, 25%), MANEC (6, 17%), Kelch repeats (4, 11%) and serine-rich adhesion proteins (2, 5%). The combination of carbohydrate-binding, interaction and adhesion domains matches those in land-plant CWI receptors and may reflect a simple toolkit for the first steps toward coordinated, multicellular growth.

Although the largest number of proteins was characterized by structural similarities presenting LRR domains as major features, few displayed an interesting bipartite domain organization. For example, the residues 1-150 of the A0A9D4TVN9 protein ectodomain showed a good structural alignment with the LRR receptor BRASSINOSTEROID INSENSITIVE 1 from *A. thaliana* (PBD entry: 4m7e) (Fig. 4B). On the other hand, the pLM-BLAST analysis identified multiple matches with known lectin domains (i.e. LysM and galactose binding domains) in the region spanning residues 235-538, (Fig. 4B). Interestingly, in this ectodomain region (i.e. 283-589) the Foldseek analysis revealed good structural homology with the crystal structure of an F-lectin (fucolectin) from *Morone saxatilis* (*Animalia*, *Chordata*; PBD entry: 3cqo) (Fig. 4B). Such bipartite designs could couple peptide and glycan perception within a single receptor, expanding signal-integration capacity.

Discussion

1. Deep learning based bioinformatic tools greatly improves RLK prediction in non-model species

Considerable advances in high-throughput sequencing of novel species, combined with improved accuracy of bioinformatic predictions over the last decade, have greatly enhanced our ability to identify and characterize RLKs outside traditional model organisms. The incorporation of deep learning-based tools into annotation pipelines has been particularly transformative for non-model species. These AI-driven methods can more accurately predict protein features such as signal peptides and transmembrane domains, overcoming limitations of earlier motif-based or homology-only searches. By taking advantage of these improvements, we created a prediction and annotation pipeline aimed to uncover previously unrecognized RLKs across diverse Chlorophyta lineages (Fig. 1A, 1B). Moreover, we provided an overview of their main structural and sequence features at the ectodomain region, which is typically involved in extracellular sensing.

Notably, our predictions recovered the vast majority of RLKs previously reported in *Chlamydomonas reinhardtii* (66 out of 68, Fig 1C), while also adding another 12 candidates to that species' list (Table 1). Considering that proteome annotation remained similar, this dramatic leap in RLK numbers relative to prior studies likely reflects the enhanced predictive performance of the deep learning tools we employed – in particular, the accurate delineation of signal peptides and transmembrane helices by DeepTMHMM. Refining the detection of these features reduces false negatives (e.g. kinases that were missed because a short signal sequence was not recognized by older algorithms), thus capturing many receptors that went undetected by earlier methods.

Indeed, a recent broad study using a similar machine-learning pipeline (Yin et al., 2023; Yin et al., 2024) likewise reported significantly higher RLK counts in several algal species. This convergence underscores how the combination of improved algorithms and expanding genomic data enables a deeper exploration of RLK diversity in non-model lineages. Minor discrepancies between our catalog and other studies – for instance, a few RLKs present in one analysis but not the other – can be attributed to differences in input data (complete genomes vs. proteomes), in RLK inclusion criteria (whether a signal peptide was required or not), and in specific search methods (e.g. our use of the kinase superfamily motif IPR011009). Overall, the deployment of modern AI-driven tools allowed us to substantially broaden the inventory of Chlorophyta RLKs. We applied these tools on

34 species, 28 of whom were never scrutinized in such fashion and uncovered 736 RLKs, over 600 of which are new, establishing a more solid foundation for analyzing their evolutionary history and potential functions in these algae.

2. Overrepresentation of motifs associated with the extracellular matrix suggests a putative role in perception and signaling for *Chlorophyta* RLKs

CWI surveillance is a complex mechanism that remains largely understudied in microalgae. A broader investigation of this process begins with the identification of putative receptors, which is essential for understanding how and when CWI mechanisms evolved across the *Viridiplantae* lineage. Gaining insight into this evolution can have broad implications, from serving as a proxy to fundamental plant biology to informing improvement strategies for various species of importance to humans, including industrially relevant algal strains and crop plants. In this context, investigating the landscape of extracellular domains present in *Chlorophyta* RLKs – especially domains whose composition and biological roles are established in other systems – represents an initial approach toward pinpointing putative CWI sensors.

In plants, the ability to detect small carbohydrates (e.g. oligosaccharides) deriving from pathogen CWs or self-damage has been linked to several RLKs mediating plant response to invasion or damage (Bacete et al., 2018). One striking result is the abundance of carbohydrate-binding domains in *Chlorophyta* RLK ectodomains (Fig. 2). Many of these domains are well-known from plant immunity and symbiosis contexts. For example, several of the algal RLKs contain LysM motifs, which in land plants bind N-acetylglucosamine-containing ligands (such as chitin fragments from fungal cell walls or Nod factors from bacteria), thereby triggering immune responses or symbiotic signaling (Chiu and Paszkowski, 2020). We also identified multiple lectin-like domains, such as jacalin (galactose/mannose-binding) modules, akin to those found in certain plant lectin-RLKs that are associated with pathogen recognition and stress responses (Weidenbach et al., 2016). The enrichment of these carbohydrate-binding ectodomains strongly suggests that green microalgae possess the capacity to sense extracellular polysaccharides or glycan-derived signals – whether released from their own cell wall during remodeling or originating from interacting organisms in their environment.

In addition to these canonical plant-associated motifs, our survey uncovered some unexpected carbohydrate-binding modules not previously described in *Viridiplantae* RLKs. Notably, we found

instances of an intimin-like domain, a bacterial adhesion motif normally used by pathogenic *E. coli* to attach intimately to host intestinal cells (binding host integrin-like receptors) (Batchelor et al., 2000). We also detected F-type lectin domains, which bind fucose or galactose and are usually associated with pathogen carbohydrate recognition in vertebrate and invertebrate innate immune receptors (Vasta et al., 2004). The presence of these atypical domains in *Chlorophyta* RLKs expands the spectrum of potential ligands and interactions available to algae. It hints that microalgae may detect a broad array of extracellular carbohydrate cues, related to their diverse cell wall composition or perhaps including signals from bacteria or other eukaryotes in aquatic ecosystems. It also raises intriguing evolutionary questions, as such domains could have been acquired via horizontal gene transfer or represent ancient metazoan/bacterial modules that have been retained in algae. In either case, the incorporation of intimin, F-lectin and other rare motifs into algal RLKs points to a mosaic domain architecture that likely reflects adaptation to unique ecological interactions.

Another notable category in chlorophyte RLKs is the array of protease and protease-associated domains found in their ectodomains. Given the prominence of glycoproteins in many algal cell walls, it makes sense that receptors might be equipped to perceive proteinaceous cell wall components or modifications thereof. On one hand, some of these domains (for instance, the Aqualysin I or Proteinase K domains) may primarily serve as binding modules that recognize specific peptide motifs or unfolded protein regions in the wall (Tanaka et al., 1998; Helland et al., 2006). On the other hand, it is conceivable that certain RLKs carry enzymatically active protease domains that could directly participate in signaling cascades by processing ligands or even by autolytically activating the receptor. In animal systems, a well-known paradigm involves protease-activated receptors that are triggered by proteolytic cleavage of their extracellular domain (Heuberger and Schuepbach, 2019). While plant RLKs typically do not self-cleave, proteolysis is a key regulatory factor in plant signaling as well – for example, cell wall proteases can release peptide fragments or modify wall integrity during stress, which then serve as danger signals to be perceived by receptors (Bacete et al., 2018). The discovery of protease-like motifs within algal RLKs suggests a potential mechanistic link between cell wall remodeling and receptor activation in algae. These RLKs might sense when cell wall proteins are cleaved or dynamically adjust their signaling in response to extracellular protease activity, thereby integrating cell wall maintenance with rapid cellular communication.

Finally, we observed a substantial representation of protein–protein interaction and peptide-binding domains in *Chlorophyta* RLKs, notably leucine-rich repeats (LRRs) and PAN/Apple domains. The pervasive occurrence of LRR-RLKs in our microalgal dataset mirrors their dominance in land plant receptor repertoires. In flowering plants, LRR-RLKs form the core of many developmental and defense signaling pathways, often binding small secreted peptides or protein ligands that regulate cell growth, differentiation, and immune responses (Furumizu and Aalen, 2023). The consistent presence of numerous LRR-containing RLKs across diverse *Chlorophyta* – with *Chlorella vulgaris* being especially enriched in LRR ectodomains – raises the possibility that analogous peptide-based signaling circuits exist even in unicellular or colonial algae. It is conceivable that these algae produce simple secreted peptide hormones or cues for processes like cell aggregation, mating, or stress adaptation, although such signals remain to be discovered. Meanwhile, the PAN domain, whose function is not yet fully elucidated in plants, appears in some of the algal RLKs as well. Recent studies link PAN domains to cell-surface receptors involved in pathogen defense in plants (De et al., 2023) and to ligand binding in mammalian systems (Pal et al., 2022), suggesting this domain contributes to protein–ligand or protein–carbohydrate interactions. The occurrence of PAN motifs alongside LRRs and other domains in algal RLKs could therefore broaden the binding spectrum or modulate ligand specificity of these receptors. Taken together, the assortment of peptide-binding and protein–interaction modules found in *Chlorophyta* RLKs implies that, despite their mostly single-celled nature, green algae may employ surprisingly complex extracellular signaling modalities.

Core components of the peptide–receptor signaling networks that we associate with land plants were likely present early in the green lineage. But in *Chlorophyta*, many of the identified RLKs carry combinations of distinct domain types (Fig. 3). This modularity may allow individual receptors to integrate multiple signal types or respond to compound ligands with both structural and biochemical features. In contrast, in *Arabidopsis*, many PRRs have lost their independent signaling capabilities and act like scaffolds, such as FERONIA (FER) or BAK1 (Chinchilla et al., 2009; Stegmann et al., 2017). In *Chlorophyta*, with a lower number of RLKs, complex co-receptor systems do not seem to be frequent, and thus, such multi-functional ectodomains may reflect a strategy of combining recognition and anchoring in a single molecule. For instance, a receptor could bind a carbohydrate structure through one domain while simultaneously interacting with a protein ligand or matrix component via another. This dual recognition might be particularly suited

to extracellular matrix/cell walls built from both polysaccharides and hydroxyproline-rich glycoproteins, as seen in many chlorophytes. Understanding how these modular receptors function—and whether their architectures reflect adaptation to specific cell wall compositions or ecological conditions—could provide a foundation for future functional studies in both algae and plants.

3. Investigating *Chlorophyta* RLKs provides new insights into their emergence and ecological adaptation

Green algae in the *Chlorophyta* phylum offer a useful model for studying the evolution of multicellularity, as they span from solitary unicells to simple multicellular forms. Even though predominantly unicellular, *Chlorophyta* often display tendencies towards aggregation (Sathe and Durand, 2016). Moreover, the volvocine group, range from single-celled *Chlamydomonas* to colony-forming *Volvox* species, representing a gradient for the unicellular-to-multicellular transition (Miller, 2010). In this context, multicellularity is usually exhibited by colonies having permanent or transient cytoplasmic connections and coordinated communication networks between cells. Therefore, the role of cell wall (and the self-perception of that wall) in maintaining the cell–cell adhesive network most likely requires dedicated signaling patterns to sustain these lifestyles.

Our analysis of algal RLKs lends support to this idea. We observed a relatively high representation of ectodomains related to cell wall anchoring and cell adhesion in *Chlorophyta* RLKs. Besides known and widespread domains involved in cell-to-cell adhesion and communication (e.g. FAS1/fasciclin domains and other adhesin motifs important for plant development; (Chae et al., 2010; Ma et al., 2023), we also identified several uncommon domains that are traditionally associated with animal or microbial proteins – including fibronectin type III repeats, cadherin-like domains, and the LPXTG cell-wall anchoring motif. The discovery of these unusual domains, along with others like $\beta\gamma$ -crystallin repeats, suggests that algae have incorporated a broader range of structural and adhesive modules than previously expected. Cadherin and fibronectin domains, for instance, are central to animal cell adhesion systems (Lefort et al., 2011); their presence in algae may reflect a convergent strategy or the retention of ancient eukaryotic modules for matrix sensing.

The current paradigm posits that RLK emerged in the *Chlorophyta* division before the divergence of major plant lineages underwent a dramatic expansion in vascular plants (Dievart et al., 2020). This is supported by the lack of diversity and overall representation as compared to other *Viridiplantae* taxa. However, our work challenges this paradigm and points to an increased diversity already present in microalgae. Interestingly, although we highlighted double digit number of RLKs in multiple proteomes, we could not detect similar amounts, or any, in several cases (**Table1**). This could be due to factors such as relatively poor genome/proteome annotation or ecological lifestyle that drove adaptation. For example, we did not detect any canonical RLKs in *Helicosporidium sp.*, an interesting non-photosynthetic parasite of many organisms (Tartar et al., 2002). This organism may have dispensed with typical cell-surface sensors like RLKs due to its sheltered intracellular environment and thus could be an example of overadaptation to a host, although it is also possible that incomplete genome annotation contributed to this apparent absence. In contrast, we predicted 21 RLKs in the endolithic symbiotic alga *Ostreobium quekettii* (**Table2, Figure2**), which forms a symbiosis with caribbean coral reefs (Lukas, 1974) – yet interestingly we could not detect LysM-domain RLKs in this species. In plant–microbe interactions, LysM-RLKs play a central role in perceiving microbial cell wall signals (such as rhizobial Nod factors or fungal chitin fragments) and initiating symbiotic or immune responses (Antolín-Llovera et al., 2014; Buendia et al., 2018). The absence of LysM motifs in *O. quekettii* suggests that early symbioses might rely on different recognition mechanisms, or alternatively, that certain conserved domains in algae have diverged beyond recognition by current motif searches. These two case studies illustrate how ecological specialization can shape the RLK complement: a highly specialized parasitic lifestyle might reduce the need for a wide array of receptors, whereas a symbiotic lifestyle still demands a suite of RLKs but potentially with a shifted domain composition aligned to its particular host or environment.

Correlating RLK diversity with habitat must be done cautiously, but our work hints at intriguing patterns. Indeed, we observed a tendency for freshwater and soil-dwelling algae to encode more RLKs compared to marine algae. Freshwater/terrestrial environments typically subject organisms to greater fluctuations in factors such as water availability, salinity, and interactions with diverse communities of bacteria and fungi. These conditions may demand more robust or varied sensing systems, which could explain the higher RLK counts in freshwater species. In contrast, many marine algae (especially planktonic forms) experience relatively stable open-ocean conditions or

have alternative adaptive strategies (e.g. rapid life cycles or heavy reliance on chemical defenses), potentially reducing the selective pressure for large RLK families. That said, there are notable exceptions to the simple marine–freshwater hypothesis. For instance, the marine symbiont *O. quekettii* (discussed above) has a substantial number of RLKs, presumably due to the demands of a symbiotic lifestyle. Meanwhile, *Chlorella denticata*, a marine alga with one of the highest RLK counts, was recently reclassified as *Nannochloris denticata* (Sanders et al., 2022), placing it in a different lineage; this taxonomic change complicates straightforward correlations between environment and RLK number. A larger sampling of algal genomes from diverse habitats, combined with careful phylogenetic analysis, will be needed to untangle how environmental parameters have influenced RLK expansion or contraction in different algal clades.

Overall, our findings support an evolutionary scenario in which the need to perceive and respond to complex extracellular signals increased as plants transitioned to new environments, such as the shift from aquatic to terrestrial life. The *Chlorophyta*, which diverged before land plants, already exhibit a non-trivial diversity of RLKs, suggesting that the foundational elements of receptor-mediated signaling were in place early on. It is noteworthy that RLK-like proteins can even be observed in more ancient photosynthetic lineages like glaucophyte algae (Gong and Han, 2021), implying that the origin of cell-surface kinase signaling predates the split between major algal groups. As plants colonized land, they likely co-opted and greatly expanded this existing toolkit of RLKs to cope with new challenges – ranging from desiccation and mechanical stress to novel pathogens – resulting in the hundreds of RLKs found in modern terrestrial plants. Further comparative analyses of early-diverging algae (including charophyte algae, the closest relatives of land plants) and bryophytes could shed light on when specific RLK subfamilies arose and how their diversification coincided with key evolutionary transitions. By tracing RLK evolution across algae and plants, we can better understand how multicellular communication networks evolved in concert with the demands of new habitats and increasing organismal complexity.

4. Towards a comprehensive understanding of RLK–ligand pairs and wall–receptor coevolution

An important next step will be to directly test the interactions between the identified microalgal RLKs and their putative ligands. Biochemical and biophysical assays can validate whether the diverse ectodomains predicted *in silico* indeed bind specific cell wall-derived molecules or

peptides. In parallel, comprehensive profiling of cell wall composition in these algae is necessary to pinpoint the actual signals that RLKs may perceive. Many *Chlorophyta* cell wall components remain uncharacterized, therefore, employing a combination of glycomic and biochemical analyses will be invaluable. Glycome profiling — a high-throughput method using panels of glycan-directed monoclonal antibodies— can systematically detect and quantify major polysaccharide epitopes present in algal cell walls (Pattathil et al., 2012). This approach, already widely used in land plant cell wall studies (Alonso Baez and Bacete, 2023), would reveal which carbohydrate motifs (e.g. pectins, arabinogalactans, β -glucans) are present or enriched in different algal species. Complementary chemical analyses, such as sequential fractionation of cell wall material followed by monosaccharide composition analysis (using GC-MS or HPLC), can identify the building blocks and proportions of sugars in each wall (Moore et al., 2020). Recent work has demonstrated the feasibility of detailed cell wall characterization in green microalgae. For instance, advanced spectroscopic and glycoproteomic techniques were able to resolve the molecular architecture of *C. reinhardtii*'s glycoprotein-rich cell wall (Poulhazan et al., 2024), underscoring that even highly diverged wall matrices can be dissected. Applying these methods across diverse *Chlorophyta* will catalog the potential ligands (polysaccharides, glycoproteins, and glycopeptides) available for RLK binding in each lineage. This knowledge will guide which ligands to include in the binding assays mentioned above and will help interpret negative results (e.g. absence of binding may simply reflect absence of the ligand in that species' wall). To strengthen causal links, experimental perturbations *in vivo* can be considered: for example, altering cell wall composition (through enzymatic treatment or genetic modification in model algae like *Chlamydomonas*) and observing the expression or activation of specific RLKs would directly show a wall–receptor functional relationship. For example, in *Arabidopsis*, altered pectin structure due to pectin methylesterase inhibitor overexpression was critical to identify the CWI receptor RLP44, linking wall modification to receptor signaling (Wolf et al., 2014).

Ultimately, by combining our prediction and annotation pipeline with ligand-binding assays, cell wall profiling, and perturbation analyses, future studies can build a cohesive picture of how *Chlorophyta* RLKs have been tailored to their extracellular context. This multifaceted approach will not only validate predicted RLK functions but also shed light on how the dynamic evolution of cell walls in the green lineage has driven, and been guided by, an equally dynamic evolution of cell wall integrity sensing mechanisms.

Methods

Data sources

Reference proteomes belonging to *Chlorophyta* phylum (NCBI Taxonomy ID 3041) were retrieved from UniProt Proteomes (entries with taxonomy ID 3041 that were classified as reference proteomes; accessed 15/08/2024). Kinase domain containing proteins were identified through InterProScan, employing entry IPR011009 (*Kinase-like_dom_sf*) as filter (<https://www.ebi.ac.uk/interpro/entry/InterPro/IPR011009/>) (accessed 01/12/2024).

Identification of RLKs

Here, we defined Receptor-Like Kinases (RLKs) as proteins containing: (a) a signal peptide (SP); (b) a single transmembrane domain (TMD), and (c) a kinase domain. To identify the presence of signal peptide and/or transmembrane domains, kinase domain containing sequences were submitted to DeepTMHMM webserver (accessed 05/12/2024). This tool exploits a deep learning protein language algorithm for the accurate detection of signal peptides and the topology of both alpha helical and beta barrels (Hallgren et al., 2022).

Structural and sequence analysis on identified RLKs

Out of all putative RLKs sequences, those belonging to nine species were subjected to further characterization. For each protein, the 3D structure was predicted using AlphaFold webserver (accessed 15/12/2024) (Abramson et al., 2024), and the model choice was based on the highest ranking score provided by the tool. The amino acid region corresponding to the ectodomain, based on DeepTMHMM prediction, was isolated for downstream analyses. Foldseek webserver (accessed 20/12/2024) (van Kempen et al., 2024) was used to determine the structure similarity of putative RLKs ectodomains with known structures deposited in the RCSB PDB100 database (Berman et al., 2000). At the same time, distant homology detection was performed on predicted ectodomain sequences using pLM-BLAST (Kaminski et al., 2023) through the MPI Bioinformatics Toolkit webserver (accessed 07/01/2025) (Gabler et al., 2020).

Selection of annotated regions

Here, we define “annotated regions” as the amino acid sequences within RLK ectodomains that exhibit sequence or structural similarity to known domains associated with extracellular matrix perception or modification.

To obtain this, we first trimmed pLM-BLAST and Foldseek data based on the generated scores. For Foldseek, the two best matching structures, score-wise, were selected for further characterization. Then, the pLM-BLAST data was processed based on two criteria: annotation type and length. First, overlapping regions were merged based on annotation similarity (i.e. sequences belonging to related family and topological group). Next, the two annotations having the highest scores for a given sequence region were selected for further steps. Annotations longer than 20 amino acids were prioritized when the difference in pLM-BLAST score was lower than 0.05. Finally, the resulting annotated regions were manually inspected to assess their relevance in the context of CWI and grouped into different classes according to their biological role. Annotated regions with poor or misleading relevance were discarded.

Data availability.

All protein sequences analyzed in this study were retrieved from the UniProt database using their corresponding UniProt accession codes. All sequence data, domain annotations, and processed results generated in this study are provided as supplementary materials accompanying this manuscript. These include the full set of input sequences, domain prediction outputs, and final annotation tables used in the analyses. No additional datasets were generated or analyzed during the current study.

Funding

This work was supported by Kempestiftelserna, Sweden (JCSMK23-0228), the Knut and Alice Wallenberg Foundation, the Swedish Research Council VR (2024-05371 to LB and 2024-05463 to CF) and grants from the Sven and Lilly Lawski foundation, Sweden.

Author contributions

Marcianò D. and Basso F. designed and operated RLKs identification pipeline with contribution from Dauphin B.G. Data processing, statistical analysis and figures were made by Marcianò D., Dauphin B.G. and Basso F. Marcianò D. and Dauphin B.G. wrote the manuscript with

contributions from all authors. Bacete L. and Funk C. conceptualized, supervised and administrated this work and were responsible for funding acquisition.

Acknowledgments

Add if needed.

Declaration of interests

The authors declare no competing interests.

Bibliography

- Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630:493–500.
- Adair, W. S., Steinmetz, S. A., Mattson, D. M., Goodenough, U. W., and Heuser, J. E. (1987). Nucleated assembly of *Chlamydomonas* and *Volvox* cell walls. *J. Cell Biol.* 105:2373–2382.
- Alonso Baez, L., and Bacete, L. (2023). Cell wall dynamics: novel tools and research questions. *J. Exp. Bot.* Advance Access published August 4, 2023, doi:10.1093/jxb/erad310.
- Antolín-Llovera, M., Petutsching, E. K., Ried, M. K., Lipka, V., Nürnberger, T., Robatzek, S., and Parniske, M. (2014). Knowing your friends and foes – plant receptor-like kinases as initiators of symbiosis or defence. *New Phytol.* 204:791–802.
- Bacete, L., and Hamann, T. (2020). The Role of Mechanoperception in Plant Cell Wall Integrity Maintenance. *Plants Basel Switz.* 9:574.
- Bacete, L., Mérida, H., Miedes, E., and Molina, A. (2018). Plant cell wall-mediated immunity: cell wall changes trigger disease resistance responses. *Plant J. Cell Mol. Biol.* 93:614–636.
- Baez, L. A., Tichá, T., and Hamann, T. (2022). Cell wall integrity regulation across plant species. *Plant Mol. Biol.* 109:483–504.
- Batchelor, M., Prasannan, S., Daniell, S., Reece, S., Connerton, I., Bloomberg, G., Dougan, G., Frankel, G., and Matthews, S. (2000). Structural basis for recognition of the

- translocated intimin receptor (Tir) by intimin from enteropathogenic *Escherichia coli*. *EMBO J.* 19:2452–2464.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.
- Bhatla, S. C. (2018). Signal Perception and Transduction. In *Plant Physiology, Development and Metabolism* (ed. Bhatla, S. C.) and A. Lal, M.), pp. 729–765. Singapore: Springer Nature.
- Buendia, L., Girardin, A., Wang, T., Cottret, L., and Lefebvre, B. (2018). LysM Receptor-Like Kinase and LysM Receptor-Like Protein Families: An Update on Phylogeny and Functional Characterization. *Front. Plant Sci.* 9.
- Chae, K., Gonong, B. J., Kim, S.-C., Kieslich, C. A., Morikis, D., Balasubramanian, S., and Lord, E. M. (2010). A multifaceted study of stigma/style cysteine-rich adhesin (SCA)-like Arabidopsis lipid transfer proteins (LTPs) suggests diversified roles for these LTPs in plant growth and reproduction. *J. Exp. Bot.* 61:4277–4290.
- Chinchilla, D., Shan, L., He, P., de Vries, S., and Kemmerling, B. (2009). One for all: the receptor-associated kinase BAK1. *Trends Plant Sci* 14:535–41.
- Chiu, C. H., and Paszkowski, U. (2020). Receptor-Like Kinases Sustain Symbiotic Scrutiny. *Plant Physiol.* 182:1597–1612.
- De, K., Pal, D., Shanks, C. M., Yates, T. B., Feng, K., Jawdy, S. S., Hassan, M. M., Prabhakar, P. K., Yang, J.-Y., Chapla, D., et al. (2023). The Plasminogen-Apple-Nematode (PAN) domain suppresses JA/ET defense pathways in plants Advance Access published June 18, 2023, doi:10.1101/2023.06.15.545202.
- Del Hierro, I., Mérida, H., Broyart, C., Santiago, J., and Molina, A. (2021). Computational prediction method to decipher receptor–glycoligand interactions in plant immunity. *Plant J.* 105:1710–1726.
- Derelle, E., Ferraz, C., Rombauts, S., Rouzé, P., Worden, A. Z., Robbens, S., Partensky, F., Degroove, S., Echeynié, S., Cooke, R., et al. (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc. Natl. Acad. Sci. U. S. A.* 103:11647–11652.
- Dievart, A., Gottin, C., Périn, C., Ranwez, V., and Chantret, N. (2020). Origin and Diversity of Plant Receptor-Like Kinases. *Annu. Rev. Plant Biol.* 71:131–156.
- Domozych, D., Ciancia, M., Fangel, J. U., Mikkelsen, M. D., Ulvskov, P., and Willats, W. G. T. (2012). The Cell Walls of Green Algae: A Journey through Evolution and Diversity. *Front. Plant Sci.* 3.

- Fangel, J. U., Sørensen, K. M., Jacobsen, N., Mravec, J., Ahl, L. I., Bakshani, C., Mikkelsen, M. D., Engelsens, S. B., Willats, W., and Ulvskov, P. (2024). The legacy of terrestrial plant evolution on cell wall fine structure. *Plant Cell Environ.* 47:1238–1254.
- Furumizu, C., and Aalen, R. B. (2023). Peptide signaling through leucine-rich repeat receptor kinases: insight into land plant evolution. *New Phytol.* 238:977–982.
- Gabler, F., Nam, S.-Z., Till, S., Mirdita, M., Steinegger, M., Söding, J., Lupas, A. N., and Alva, V. (2020). Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr. Protoc. Bioinforma.* 72:e108.
- Gong, Z., and Han, G.-Z. (2021). Flourishing in water: the early evolution and diversification of plant receptor-like kinases. *Plant J.* 106:174–184.
- Hallgren, J., Tsigirigos, K. D., Pedersen, M. D., Armenteros, J. J. A., Marcatili, P., Nielsen, H., Krogh, A., and Winther, O. (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks Advance Access published April 10, 2022, doi:10.1101/2022.04.08.487609.
- Hallmann, A., Godl, K., Wenzl, S., and Sumper, M. (1998). The highly efficient sex-inducing pheromone system of *Volvox*. *Trends Microbiol.* 6:185–189.
- Helland, R., Larsen, A. N., Smalås, A. O., and Willassen, N. P. (2006). The 1.8 Å crystal structure of a proteinase K-like enzyme from a psychrotroph *Serratia* species. *FEBS J.* 273:61–71.
- Heuberger, D. M., and Schuepbach, R. A. (2019). Protease-activated receptors (PARs): mechanisms of action and potential therapeutic modulators in PAR-driven inflammatory diseases. *Thromb. J.* 17:4.
- Illergård, K., Ardell, D. H., and Elofsson, A. (2009). Structure is three to ten times more conserved than sequence--a study of structural response in protein cores. *Proteins* 77:499–508.
- Kaminski, K., Ludwiczak, J., Pawlicki, K., Alva, V., and Dunin-Horkawicz, S. (2023). pLM-BLAST: distant homology detection based on direct comparison of sequence representations from protein language models. *Bioinforma. Oxf. Engl.* 39:btad579.
- Lefort, C. T., Wojciechowski, K., and Hocking, D. C. (2011). N-cadherin Cell-Cell Adhesion Complexes Are Regulated by Fibronectin Matrix Assembly. *J. Biol. Chem.* 286:3149–3160.
- Lehti-Shiu, M. D., Zou, C., Hanada, K., and Shiu, S.-H. (2009). Evolutionary History and Stress Regulation of Plant Receptor-Like Kinase/Pelle Genes. *Plant Physiol.* 150:12–26.

- Lemmon, M. A., and Schlessinger, J. (2010). Cell signaling by receptor-tyrosine kinases. *Cell* 141:1117–1134.
- Liu, J., Li, W., Wu, G., and Ali, K. (2024). An update on evolutionary, structural, and functional studies of receptor-like kinases in plants. *Front. Plant Sci.* 15.
- Lukas, K. J. (1974). TWO SPECIES OF THE CHLOROPHYTE GENUS *OSTREOBIUM* FROM SKELETONS OF ATLANTIC AND CARIBBEAN REEF CORALS^{1,2}. *J. Phycol.* 10:331–335.
- Ma, Y., Shafee, T., Mudiyanse, A. M., Ratcliffe, J., MacMillan, C. P., Mansfield, S. D., Bacic, A., and Johnson, K. L. (2023). Distinct functions of FASCILIN-LIKE ARABINOGLYCAN PROTEINS relate to domain structure. *Plant Physiol.* 192:119–132.
- Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., Maréchal-Drouard, L., et al. (2007). The Chlamydomonas Genome Reveals the Evolution of Key Animal and Plant Functions. *Science* 318:245–250.
- Miller, S. (2010). Volvox, Chlamydomonas, Evolution of Multicellularity. *Nat. Educ.* 3:65.
- Moore, J. P., Gao, Y., Zietsman, A. J. J., Fangel, J. U., Trygg, J., Willats, W. G. T., and Vivier, M. A. (2020). Analysis of Plant Cell Walls Using High-Throughput Profiling Techniques with Multivariate Methods. In *The Plant Cell Wall* (ed. Popper, Z. A.), pp. 327–337. New York, NY: Springer New York.
- Morita, J., Kato, K., Nakane, T., Kondo, Y., Fukuda, H., Nishimasu, H., Ishitani, R., and Nureki, O. (2016). Crystal structure of the plant receptor-like kinase TDR in complex with the TDIF peptide. *Nat. Commun.* 7:12383.
- Oelmüller, R., Tseng, Y.-H., and Gandhi, A. (2023). Signals and Their Perception for Remodelling, Adjustment and Repair of the Plant Cell Wall. *Int. J. Mol. Sci.* 24:7417.
- Pal, D., De, K., Shanks, C. M., Feng, K., Yates, T. B., Morrell-Falvey, J., Davidson, R. B., Parks, J. M., and Muchero, W. (2022). Core cysteine residues in the Plasminogen-Apelin-Nematode (PAN) domain are critical for HGF/c-MET signaling. *Commun. Biol.* 5:1–12.
- Pattathil, S., Avci, U., Miller, J. S., and Hahn, M. G. (2012). Immunological Approaches to Plant Cell Wall and Biomass Characterization: Glycome Profiling. In *Biomass Conversion* (ed. Himmel, M. E.), pp. 61–72. Totowa, NJ: Humana Press.
- Popper, Z. A., and Tuohy, M. G. (2010). Beyond the Green: Understanding the Evolutionary Puzzle of Plant and Algal Cell Walls¹. *Plant Physiol.* 153:373–383.

- Poulhazan, A., Arnold, A. A., Mentink-Vigier, F., Muszyński, A., Azadi, P., Halim, A., Vakhrushev, S. Y., Joshi, H. J., Wang, T., Warschawski, D. E., et al. (2024). Molecular-level architecture of *Chlamydomonas reinhardtii*'s glycoprotein-rich cell wall. *Nat. Commun.* 15:986.
- Rashidi, B., and Trindade, L. M. (2018). Detailed biochemical and morphologic characteristics of the green microalga *Neochloris oleoabundans* cell wall. *Algal Res.* 35:152–159.
- Restrepo-Montoya, D., Brueggeman, R., McClean, P. E., and Osorno, J. M. (2020). Computational identification of receptor-like kinases “RLK” and receptor-like proteins “RLP” in legumes. *BMC Genomics* 21:459.
- Sanders, C. K., Hanschen, E. R., Biondi, T. C., Hovde, B. T., Kunde, Y. A., Eng, W. L., Kwon, T., and Dale, T. (2022). Phylogenetic analyses and reclassification of the oleaginous marine species *Nannochloris* sp. “desiccata” (Trebouxiophyceae, Chlorophyta), formerly *Chlorella desiccata*, supported by a high-quality genome assembly. *J. Phycol.* 58:436–448.
- Sathe, S., and Durand, P. M. (2016). Cellular aggregation in *Chlamydomonas* (Chlorophyceae) is chimaeric and depends on traits like cell size and motility. *Eur. J. Phycol.* 51:129–138.
- Shiu, S.-H., Karlowski, W. M., Pan, R., Tzeng, Y.-H., Mayer, K. F. X., and Li, W.-H. (2004). Comparative Analysis of the Receptor-Like Kinase Family in Arabidopsis and Rice[W]. *Plant Cell* 16:1220–1234.
- Song, W., Wang, B., Li, X., Wei, J., Chen, L., Zhang, D., Zhang, W., and Li, R. (2015). Identification of Immune Related LRR-Containing Genes in Maize (*Zea mays* L.) by Genome-Wide Sequence Analysis. *Int. J. Genomics* 2015:231358.
- Spain, O., and Funk, C. (2022). Detailed Characterization of the Cell Wall Structure and Composition of Nordic Green Microalgae. *J. Agric. Food Chem.* 70:9711–9721.
- Stegmann, M., Monaghan, J., Smakowska-Luzan, E., Rovenich, H., Lehner, A., Holton, N., Belkhadir, Y., and Zipfel, C. (2017). The receptor kinase FER is a RALF-regulated scaffold controlling plant immune signaling. *Science* 355:287–289.
- Tanaka, T., Matsuzawa, H., and Ohta, T. (1998). Substrate Specificity of Aqualysin I, a Bacterial Thermophilic Alkaline Serine Protease from *Thermus aquaticus* YT-1: Comparison with Proteinase K, Subtilisin BPN' and Subtilisin Carlsberg. *Biosci. Biotechnol. Biochem.* 62:2161–2165.
- Tartar, A., Boucias, D. G., Adams, B. J., and Becnel, J. J. (2002). Phylogenetic analysis identifies the invertebrate pathogen *Helicosporidium* sp. as a green alga (Chlorophyta). *Int. J. Syst. Evol. Microbiol.* 52:273–279.

- Tör, M., Lotze, M. T., and Holton, N. (2009). Receptor-mediated signalling in plants: molecular patterns and programmes. *J. Exp. Bot.* 60:3645–3654.
- van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C. L. M., Söding, J., and Steinegger, M. (2024). Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.* 42:243–246.
- Vasta, G. R., Ahmed, H., and Odom, E. W. (2004). Structural and functional diversity of lectin repertoires in invertebrates, protochordates and ectothermic vertebrates. *Curr. Opin. Struct. Biol.* 14:617–630.
- Vukašinović, N., Serif, M., and Bacete, L. (2023). Cracking the green wall code: insights into cell wall integrity across organisms. *Front. Plant Physiol.* 1.
- Wei, X., Wang, Y., Zhang, S., Gu, T., Steinmetz, G., Yu, H., Guo, G., Liu, X., Fan, S., Wang, F., et al. (2022). Structural analysis of receptor-like kinase SOBIR1 reveals mechanisms that regulate its phosphorylation-dependent activation. *Plant Commun.* 3:100301.
- Weidenbach, D., Esch, L., Möller, C., Hensel, G., Kumlehn, J., Höfle, C., Hückelhoven, R., and Schaffrath, U. (2016). Polarized Defense Against Fungal Pathogens Is Mediated by the Jacalin-Related Lectin Domain of Modular *Poaceae*-Specific Proteins. *Mol. Plant* 9:514–527.
- Wolf, S. (2022). Cell Wall Signaling in Plant Development and Defense. *Annu. Rev. Plant Biol.* 73:323–353.
- Wolf, S., van der Does, D., Ladwig, F., Sticht, C., Kolbeck, A., Schurholz, A. K., Augustin, S., Keinath, N., Rausch, T., Greiner, S., et al. (2014). A receptor-like protein mediates the response to pectin modification by activating brassinosteroid signaling. *Proc Natl Acad Sci U A* 111:15261–6.
- Yan, J., Su, P., Meng, X., and Liu, P. (2023). Phylogeny of the plant receptor-like kinase (RLK) gene family and expression analysis of wheat RLK genes in response to biotic and abiotic stresses. *BMC Genomics* 24:224.
- Yin, Z., Shen, D., Zhao, Y., Peng, H., Liu, J., and Dou, D. (2023). Cross-kingdom analyses of transmembrane protein kinases show their functional diversity and distinct origins in protists. *Comput. Struct. Biotechnol. J.* 21:4070–4078.
- Yin, Z., Liu, J., and Dou, D. (2024). RLKdb: A comprehensively curated database of plant receptor-like kinase families. *Mol. Plant* 17:513–515.

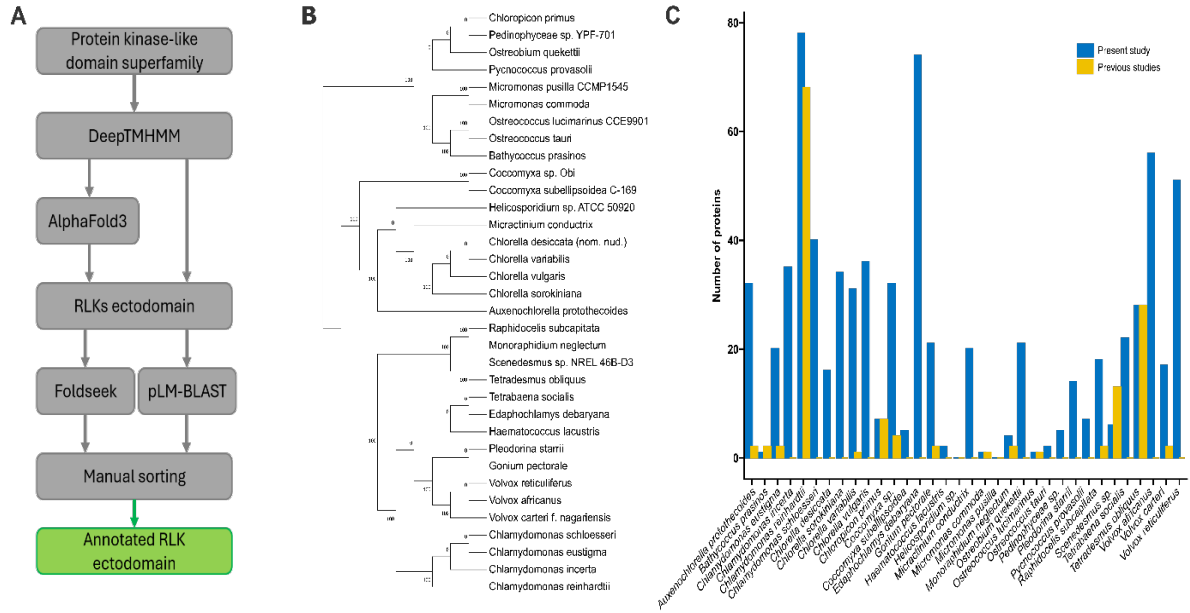


Figure 1: A pipeline for the identification of new RLK in selected *Chlorophyta* species. (A) Using available proteomes from *Chlorophyta* species, protein sequences containing the kinase-like domain were extracted and subsequently filtered through DeepTMHMM (Halgren et al 2022). Sequences with predicted signal peptide and at least one transmembrane domain were subjected to 3D structure prediction using AlphaFold3, and the predicted extracellular region (i.e. ectodomain) was selected for further analyses. The identification of functional domains present in the ectodomains was performed through Foldseek and pLM-BLAST tools to gain overall insight on protein function. **(B)** Phylogenetic tree for selected species, based on NCBI taxonomy browser. Note that *C. dessicata* has recently been reclassified as *Nannochloris dessicata* and its lineage is not as clear, sharing closer ancestry with *Auxenochlorellae* (Sanders et al., 2022). **(C)** Number of putative RLKs identified using the pipeline presented in this study (blue), against the number of RLKs identified in previous studies (in yellow) (Yin et al., 2024) (Liu et al., 2024) (Gong and Han, 2021).

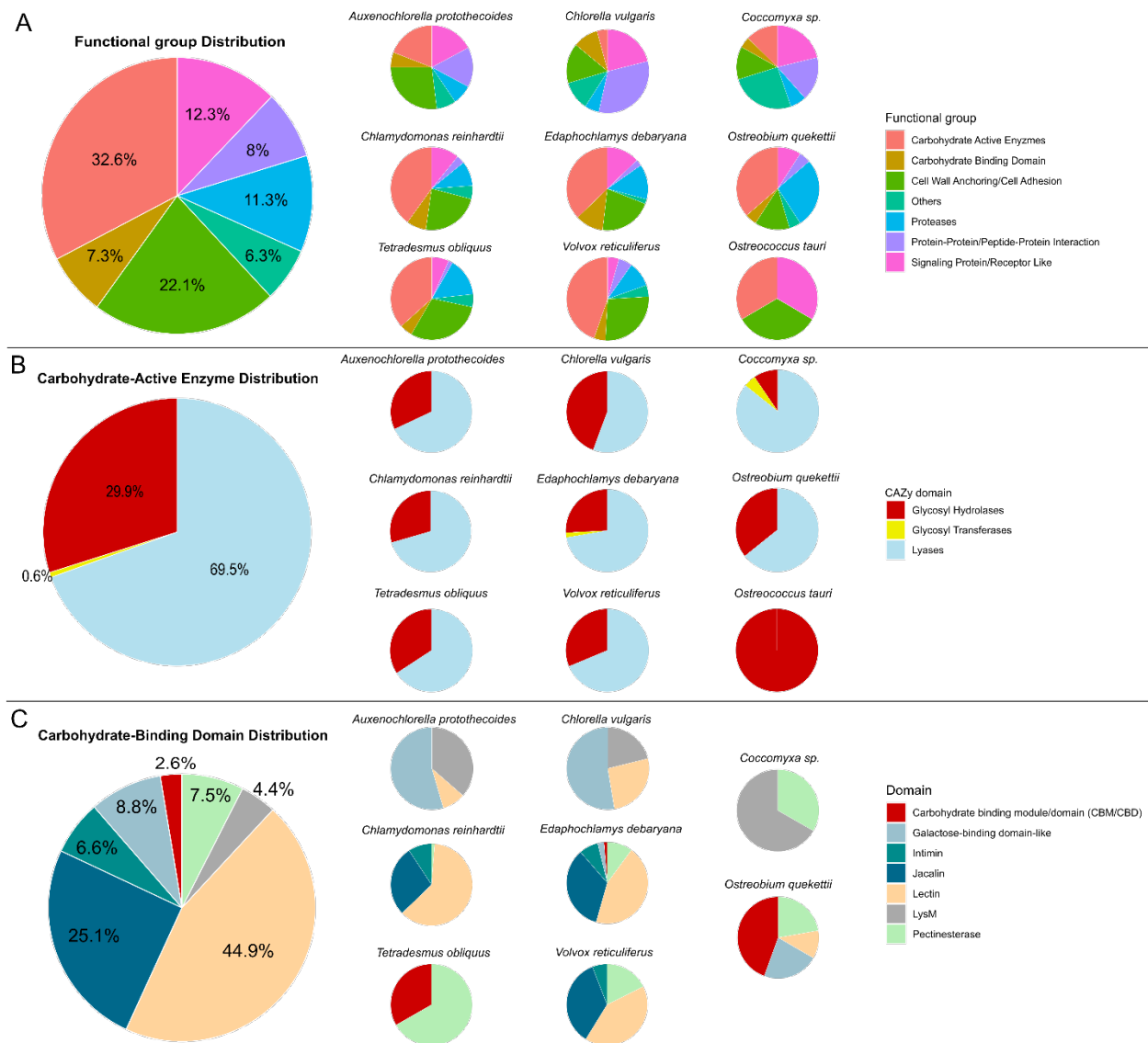


Figure 2: A comparative analysis reveals 7 distinct functional group in newly-identified RLK's ectodomains. (A) Predictions obtained through Foldseek and pLM-BLAST analyses on RLK's ectodomain from 9 *Chlorophyta* species were sorted in 7 functional groups according to their biological role or function. Pie charts show the percentage of annotated regions corresponding to a given entry both grouped together and mapped for each species (B) Annotated regions having Carbohydrate active enzyme moieties that contained glycosyl hydrolases (GH), transferases (GT) or lyases. (C) Annotated regions having Carbohydrate binding domain moieties (including different lectin families and pectinesterase-related binding domains). The overall distribution

shows distinct CBD features in a species dependent manner (Adair et al., 1987; Spain and Funk, 2022; Poulhazan et al., 2024). Others: Metal binding and oxidoreduction; Defence; Apoptosis; Viral-related proteins.

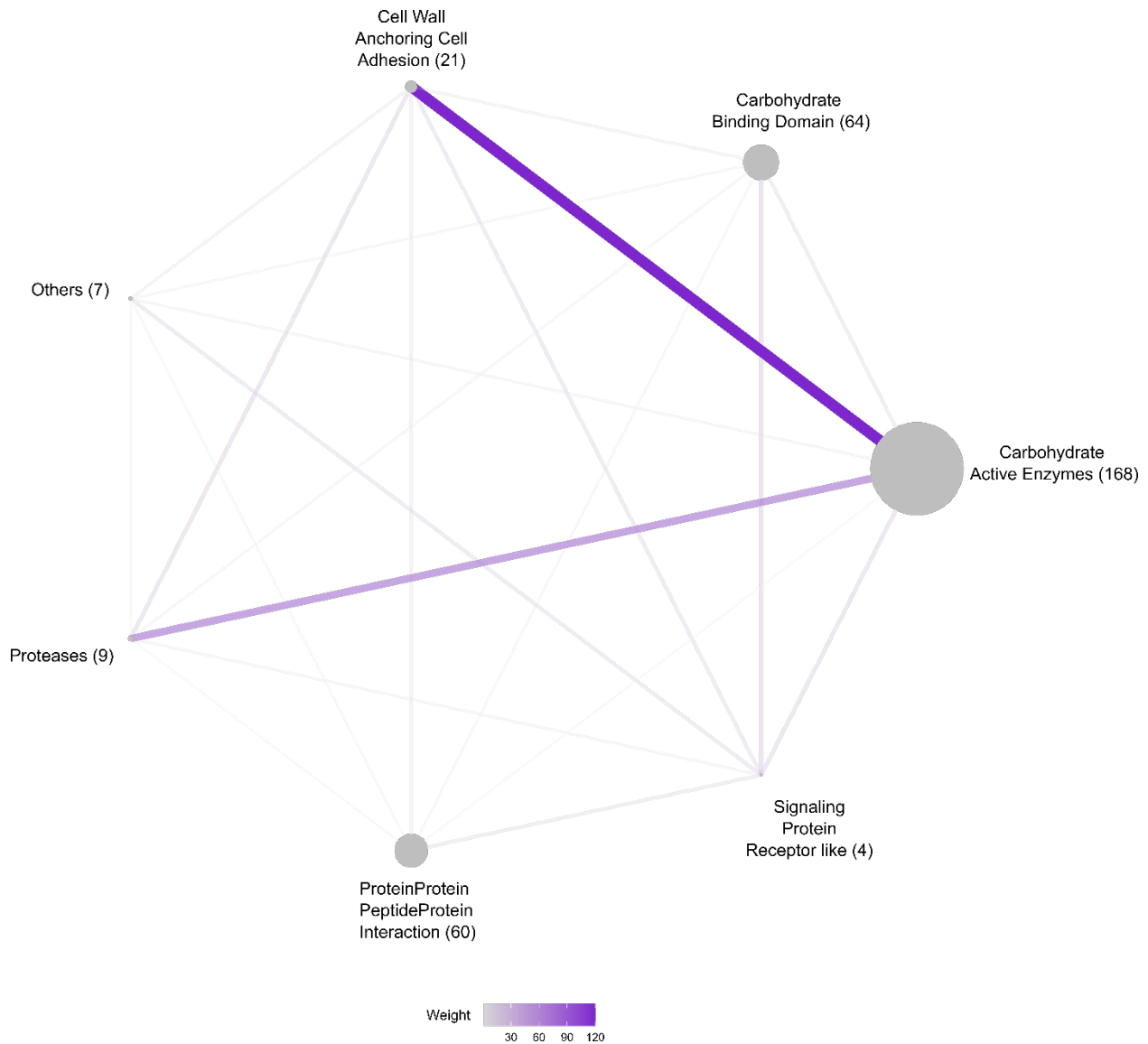
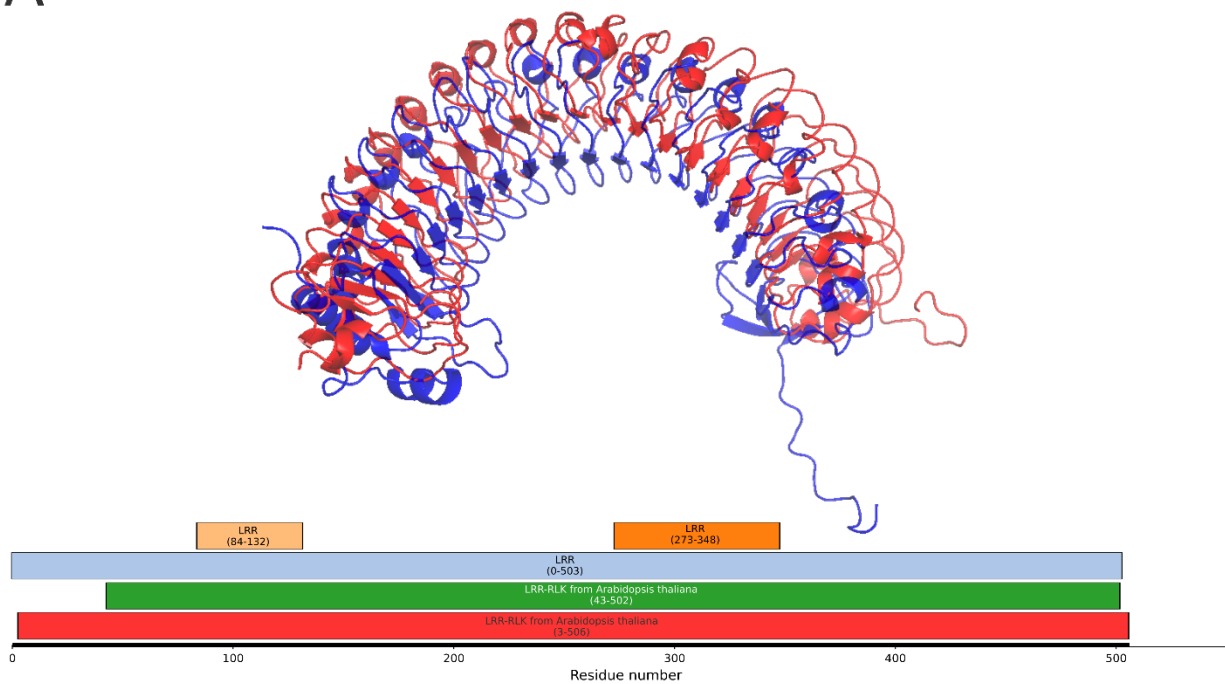


Figure 3: A weighted network analysis highlights connections existing among overlapping annotated regions. Node size is proportional to the number of annotated regions belonging to given functional group (i.e., how often does a domain appear without overlapping with others). Line thickness and color intensity represent the frequency of overlaps between groups.

A



B

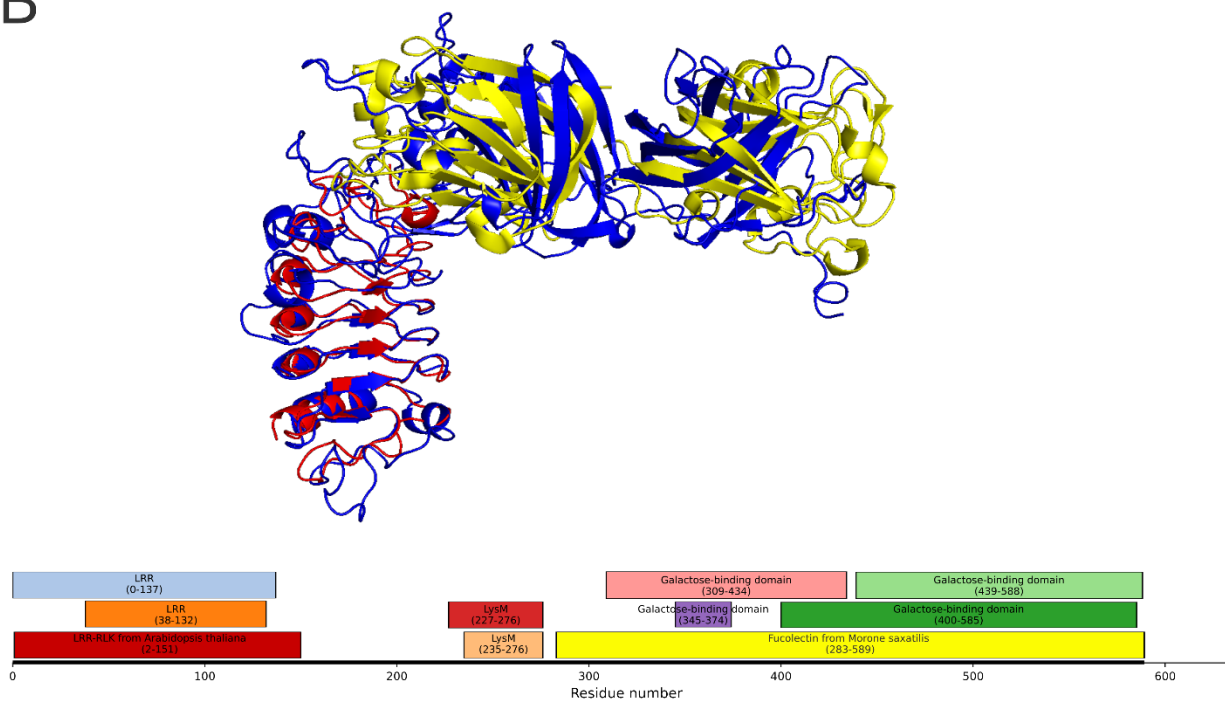


Figure 4: Putative RLKs identified in *Chlorella vulgaris* show similarities with previously described CWI sensors in *A. thaliana*.

RLK identification and annotation was performed using the bioinformatic pipeline described in this work. **(A)** The extracellular (ecto)domain of protein A0A9D4TQY7 predicted with AlphaFold3 (blue) overlapped against the crystal structure of the leucine rich repeat (LRR) *Arabidopsis thaliana* receptor TDR (red; PDB entry: 4m7e). **(B)** The bipartite organization of putative RLK A0A9D4TVN9 (blue) with corresponding region overlapping to the *Arabidopsis thaliana* LRR-receptor BRI1 (red; PDB entry: 4m7e) and the fucose-binding lectin from *Morone saxatilis* (yellow; PDB entry: 3cqq). The graphical view below displays annotated domains along the sequence and structural matches identified via Foldseek.

Species	Proteome ID	Taxonomy ID	Kinase domain containing proteins	Putative RLKs
<i>Chlamydomonas</i>				
<i>reinhardtii</i>	UP000006906	3055	828	78
<i>Volvox carteri</i>	UP000001058	3068	486	17
<i>Auxenochlorella</i>				
<i>protothecoides</i>	UP000028924	3075	504	32
<i>Chlorella sorokiniana</i>	UP000239899	3076	381	34
<i>Chlorella vulgaris</i>	UP001055712	3077	253	36
<i>Tetradismus obliquus</i>	UP000256970	3088	420	28
<i>Gonium pectorale</i>	UP000075714	33097	774	21
<i>Bathycoccus prasinus</i>	UP000198341	41875	134	1
<i>Pycnococcus</i>				
<i>provasolii</i>	UP000660262	41880	288	7
<i>Haematococcus</i>				
<i>lacustris</i>	UP000485058	44745	585	2
<i>Edaphochlamys</i>				
<i>debaryana</i>	UP000612055	47281	1000	73
<i>Tetrabaena socialis</i>	UP000236333	47790	587	22
<i>Chlamydomonas</i>				
<i>incerta</i>	UP000650467	51695	855	35
<i>Volvox africanus</i>	UP000747399	51714	632	56
<i>Ostreococcus tauri</i>	UP000009170	70448	290	2
<i>Ostreobium quekettii</i>	UP000708148	121088	453	21
<i>Monoraphidium</i>				
<i>neglectum</i>	UP000054498	145388	458	4
<i>Micromonas commoda</i>	UP000002009	296587	184	1
<i>Raphidocelis</i>				
<i>subcapitata</i>	UP000247498	307507	369	18
<i>Pleodorina starrii</i>	UP001165080	330485	491	14

<i>Ostreococcus</i>				
<i>lucimarinus</i>	UP000001568	436017	144	1
<i>Micractinium</i>				
<i>conductrix</i>	UP000239649	554055	295	20
<i>Chlorella variabilis</i>	UP000008141	554065	276	31
<i>Micromonas pusilla</i>	UP000001876	564608	175	0
<i>Coccomyxa</i>				
<i>subellipsoidea</i>	UP000007264	574566	242	5
<i>Pedinophyceae sp.</i>	UP000836011	765719	226	5
<i>Chlamydomonas</i>				
<i>eustigma</i>	UP000232323	1157962	456	20
<i>Helicosporidium sp.</i>	UP000026042	1291522	107	0
<i>Chlorella desiccata</i>	UP000693320	1650286	112	16
<i>Volvox reticuliferus</i>	UP000747110	1737510	1038	51
<i>Chloropicon primus</i>	UP000316726	1764295	244	7
<i>Chlamydomonas</i>				
<i>schloesseri</i>	UP000613740	2026947	756	40
<i>Coccomyxa sp.</i>	UP000827759	2315456	283	32
<i>Scenedesmus sp.</i>	UP000588193	2650976	339	6

Table1: Number of proteins containing at least one kinase domains identified in each species according to InterPro classification, together with the number of putative RLKs identified in the present study. For each species, UniProt proteome ID and NCBI Taxonomy ID are also reported. Species highlighted in bold have been selected for further characterization.

Supplemental table 1: Protein sequence details according to DeepTMHMM predictions. The table includes predicted positions and sequences of signal peptides, ectodomain and endodomain regions as well as transmembrane domains.

Supplemental table 2: Aminoacid sequences and descriptions of the associated function, classification or homology of the annotated regions from Foldseek structural analysis (on crystal structure from PDB100 database) and protein sequence alignment (e.g. remote homology from pLM-BLAST). Annotated region position numbers are referring to the annotated region on the ectodomain sequence (see **Supplementary material 1**).

Supplemental table 3: Functional classification of RLKs proteins across selected green algae species. The table summarizes the functional group distribution for the protein of interest. The table also includes the relative proportion of proteins having cell wall-related domains (carbohydrate-binding modules/domains and carbohydrate-active enzyme classes). Percentages indicate the relative abundance of each subcategory within its respective group ("n.a." = data not available).

Supplemental table 4: Number of putative RLKs containing at least one annotated region classified for its potential relevance in cell wall integrity. The table shows the number of putative receptor-like kinase (RLK) proteins having domains classified for their relevance in cell wall integrity (CWI). The classification includes 7 functional categories, including: Carbohydrate Active Enzymes (CAZymes), Carbohydrate Binding Domains (CBDs), Cell Wall Anchoring/Adhesion, Signaling Proteins/Receptors, Protein-Protein/Peptide-Protein Interactions, Proteases and other domains. Additionally, the table includes counts of proteins possessing regions with different classifications, represented by two-letter combinations (e.g., ac, ab, ad).

Supplemental table 5: The table summarizes all annotated regions excluded from further analyses together with their position in the ectodomain and a description of their function.

Supplemental material 1: For each protein, the figure illustrates the location of annotated regions within the amino acid sequence, as identified by two prediction tools: Foldseek and pLM-BLAST. Each colored bar represents a distinct domain, mapped to its corresponding position in the protein sequence. Annotations from Foldseek include the functional description of the crystal structures to which they are related. Note that a single protein may have multiple entries highlighting domains

in overlapping or adjacent regions. In cases where overlapping annotations shared similar functional descriptions, they were merged into a consensus domain, represented as a single annotated region (see Methods for details).