

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

Demetrio Marcianò<sup>1,2#</sup>, Bastien G. Dauphin<sup>1#</sup>, Fabian Basso<sup>3</sup>, Christiane Funk<sup>2\*</sup> and Laura Bacete<sup>1,4\*</sup>

<sup>1</sup>Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, SE-901 87 Umeå, Sweden

<sup>2</sup>Department of Chemistry, Umeå University, SE-901 87 Umeå, Sweden

<sup>3</sup>School of Computation, Information and Technology, Technische Universität München, D-80333 München, Germany

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

#These authors contributed equally to this article.

\*Corresponding authors.

## **Abstract**

Receptor-like kinases (RLKs) detect external and internal signals, triggering responses essential for growth and adaptation. Among internal cues, cell wall integrity (CWI) sensing plays a key role, as changes in cell wall structure activate responses critical for development and defense. While RLKs are well studied in vascular plants, their diversity and function remain largely unknown in green algae belonging to Chlorophyta phylum, a group relevant for global oxygen production and carbon cycling. Due to their varied cell wall structures, Chlorophyta offer a useful system to study the origins of CWI sensing. In this study, we used advanced bioinformatics and AI-based tools to analyze RLKs in 34 Chlorophyta species, mapping their distribution, structural features, and similarity to plant RLKs. We identified 736 putative RLKs, expanding the known repertoire in green algae. Structural analyses showed a wide range of extracellular domains, including motifs related to plant CWI sensors: domains mediating protein interactions (e.g. Leucine Rich Repeats - LRR, Plasminogen Apple Nematode-PAN, Armadillo repeat - ARM), cell wall remodeling (e.g. glycosyl hydrolases, lyases), and mechanosensing (e.g. Leucine-Proline-X-Threonine-Glycine motifs - LPXTG, Fibronectin). This diversity suggests that mechanisms for extracellular sensing and CWI monitoring emerged early in evolution. The results provide a basis for future studies into RLK function in algae and their evolutionary links to vascular plant signaling.

**Keywords:** Green algae signal transduction; Receptor functional divergence; Algae-plants evolutionary conservation; *Chlorella vulgaris*

## 1-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 & Schlessinger, 2010; Bhatla, 2018). To integrate this information, extracellular pattern-recognition receptors (PRRs) occur in large numbers (Lemmon & 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 domain (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 &

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 & Tuohy, 2010; Domozych *et al.*, 2012; Spain & 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, including *Chlorophyta* (five 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 bioinformatic pipeline to predict 736 RLKs –increasing

fivefold the previous counts and exposing a rich, previously hidden receptor landscape including lectin and other rare motifs. Whereas previous approaches (Yin et al., 2024) have surveyed RLKs across *Viridiplantae* with limited algal coverage, our work focuses on *Chlorophyta* species and emphasizes understanding RLK structure and function rather than simply cataloguing new sequences. Our domain-focused analysis reveals that composite ectodomains are substantially more frequent in *Chlorophyta* (61% of RLKs) than in land plants (~30%) and uncovers domain architectures, such as intimin-like or F-type lectin modules, not previously reported in land plants. By connecting these structural patterns to possible ecological and evolutionary drivers, our work provides insights into how RLK architecture relates to cell wall sensing in aquatic environments. This structural-functional connection establishes a foundation for understanding how CWI surveillance mechanisms diversified before terrestrialization. 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. Furthermore, 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.

## **2-Materials and 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). A sequence motif analysis was performed using custom Python script to detect canonical (LxxLxLxxN/C/T/S L), plant-type (LxxLxLxxN/C/T/SxLxxL), short (LxLxxN/T/S L), and cysteine-containing (LxxLxLxxC L) LRR motifs through regular expression pattern searches. Motif occurrences were counted on each predicted LRR-containing regions from RLK ectodomains and summarized at the protein level.

### **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.

### 3-Results

#### **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 the *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**). However, it must be noted that the dataset does not include members from four *Chlorophyta* classes (*Ulvophyceae*, *Nephroselmidophyceae*, *Picocystophyceae*, *Pseudoscourfieldiophyceae*, and *Chlorodendrophyceae*). For many of these taxa, only genome assemblies are currently available, and no complete annotated proteomes are deposited in public repositories, therefore precluding their inclusion in our proteome-based analysis.

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 (Yin *et al.*, 2024), up from 137 (**Figure 1C; Table S1**). Most notably, our data presents significant progress in some species for which past

predictions only highlighted a few RLKs (**Figure 1C**; See also **Table S1**). 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 amino acid sequence may suggest (Illergård *et al.*, 2009), a particularity already observed in *A. thaliana* RLK intraspecific variation (Wei *et al.*, 2022).

### **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 (**Table S2** and **Supplemental Material 1**). These were classified in seven functional groups according to their known biological role (i.e. (i) Carbohydrate active enzyme (CAZymes) (32.6 %), (ii) Cell wall anchoring/adhesion (22.1 %), (iii) Signaling protein/Receptor like (12.3 %), (iv) Proteases (11.3 %), (v) Protein-protein/Protein-peptide interaction (8 %), (vi) Carbohydrate binding domains (7.3 %), (vii) Other domains (6.3 %)) (**Figure 2A**; See also **Table S3** and **Table S4**). Most annotated regions matched plant motifs or structure, including Lectin, LRR, PAN, thaumatin and LysM domain, often in a composite fashion (**Table S5**). 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 relationships 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 (**Table S5**). 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 of lyases related motifs, particularly active on pectins (**Figure 2B, Table S2**). 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. These domains have role in the self-perception of cell wall modifications (e.g. Damage Associated Molecular Pattern -DAMP), as well as in the recognition of microbial-derived signals (Microbe Associated Molecular Pattern – MAMP). 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 RLKs we predicted composite ectodomains: a single protein often contained motifs from multiple functional groups (**Table S2, Table S4**). A network diagram of the annotated regions (**Figure S1**) highlights these cross-overs, with prominent links between adhesion/anchoring and CAZyme motifs, between CAZymes and proteases, and between carbohydrate-binding modules and signaling/receptor-like regions—connections that mirror established RLK pathways in land plants.

### **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 3A**). Moreover, a sequence motif analysis of the predicted LRR regions in the ectodomains of *C. vulgaris* RLKs identified known LRR motifs in 92% of the sequences. The number of repeats varied from 1 to 14, with short-type motifs being the most abundant (51%), followed by canonical (25%), plant-type (22%), and cysteine-containing motifs (2%). Overall, 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 (Tracheary element Differentiation inhibitory factor Receptor) (PDB entry: 5gij), an LRR-RLK able to perceive CLE (CLAVATA3/Embryo Surrounding Region-Related) 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 (**Figure 3A**) that was also endorsed by the pLM-BLAST analysis identifying LRR motifs approximately spanning the same region (**Figure 3A**). 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%), motif at N terminus with eight cysteines (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* (PDB entry: 4m7e) (**Figure 3B**). 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, (**Figure 3B**). 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*; PDB entry: 3cqo) (**Figure 4B**). Such bipartite designs could couple peptide and glycan perception within a single receptor, expanding signal-integration capacity.

## 4-Discussion

### 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 (**Figure 1A, B**). Besides, 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, **Figure 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, 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.

### **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 *Chlorophyta*. 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 from 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.

The composition and structure of the extracellular matrix in green algae are highly diverse. Some species possess cell walls resembling those of land plants, with cellulose fibrils and pectic polysaccharides, whereas other members have hydroxyproline-rich and glycine-rich glycoproteins (HRGPs) as main constituents but lack cellulose (Domozych *et al.*, 2012; Rashidi & Trindade, 2018). As sensors immersed in these matrices, algal RLK ectodomains would be expected to monitor both carbohydrate and protein components of the cell wall for signs of modification or damage. Consistently, we found an overrepresentation of certain motif types associated with cell walls, providing clues to their functional significance in algal perception and signaling. In this context, the high proportion of carbohydrate-related motifs as well as protease domains suggests that most of the ectodomains identified in this study could be involved in these processes.

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; Zhai *et al.*, 2024; Fuertes-Rabanal *et al.*, 2025). One striking result is the abundance of carbohydrate-binding domains in *Chlorophyta* RLK ectodomains (**Figure 2, Table S3**). 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 & 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 *Chlorophyta* 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 (**Figure 2, Table S3**). 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 *Chlorophyta* 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 (Hunsperger *et al.*, 2015; Chen *et al.*, 2021) 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, which have not been previously reported in land plants, 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 (**Figure 2, Table S3**). 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 & 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 (**Figure 2, Table S3**). 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 & Aalen, 2023). The consistent presence of numerous LRR-containing RLKs across diverse *Chlorophyta* – with *C. 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. In *Chlorophyta*, however, many of the identified RLKs carry combinations of distinct domain types, with composite ectodomains (receptors containing two or more extracellular modules) representing 61% of the RLKs identified across the nine species analyzed (Figure S1). This modularity may allow individual receptors to integrate multiple signal types or respond to compound ligands with both structural and biochemical features. Although domain combinations like these also occur in land plants—for example, the LRR–malectin receptor IOS1 in *A. thaliana* (Yeh *et al.*, 2016)—their overall frequency is substantially lower. Indeed, a large-scale analysis of plant RLKs reported that composite ectodomains account for roughly 30% of receptors when unaffiliated classes are excluded, and an even smaller fraction when these are considered (Dievart *et al.*, 2020). Moreover, we must consider the fact that in Arabidopsis, many PRRs have lost their independent signaling capabilities and act like scaffolds, such as FERONIA (FER) or BAK1 (Brassinosteroid insensitive Associated receptor Kinase 1) (Chinchilla *et al.*, 2009; Stegmann *et al.*, 2017). In contrast, *Chlorophyta*, which generally features fewer RLKs, appear to rely more heavily on multifunctional ectodomains rather than complex co-receptor systems. Although this may result from the fact that such systems have not yet been investigated in sufficient depth, this quantitative contrast may also indicate that *Chlorophyta* may have undergone extensive domain shuffling or retention of multifunctional architectures to combine recognition and anchoring within a single receptor. 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. In *Chlorella vulgaris*, for instance, several predicted RLKs show notable structural resemblance to land plant receptors involved in CWI surveillance

(**Figure 3**). One receptor, A0A9D4TQY7, displays strong structural overlap with the Arabidopsis CLE peptide receptor TDR (Li *et al.*, 2017), while another, A0A9D4TVN9, features a bipartite ectodomain combining LRR and fucose-binding lectin folds, reminiscent of signal-integration modules found in vascular plants. For example, the malectin-domain RLK FER functions as a scaffold in *Arabidopsis*, coordinating signals from extracellular modules such as leucine-rich repeat extensins (LRXs), rapid alkalization factor peptides (RALFs), and LORELEI-like GPI-anchored proteins to regulate processes including growth, mechanosensing and immunity (Feng *et al.*, 2018; Dunser *et al.*, 2019; Duan *et al.*, 2020; Herger *et al.*, 2020). Such domain organizations, coupling peptide and glycan perception within a single receptor, may have provided algae with versatile sensors capable of detecting a broad range of extracellular cues. 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.

### **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 ranges 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 (**Figure 2**, **Table S3**). 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 early diverging green algae (**Figure 4**). Interestingly, although we highlighted double digit number of RLKs in multiple proteomes, we could not detect similar amounts, or any, in several cases (**Table 1**). 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 queketti* (**Figure 2, Table 1**), 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 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 (average of 22 or 6 RLKs per species, respectively). 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 (21), 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. Future work could extend this approach to include charophyte green algae (Streptophyta), such as *Penium* or *Klebsormidium*, which have been examined more extensively for signalling evolution and terrestrial adaptation (Dievart *et al.*, 2020; Gong & Han, 2021). Such analyses would enable direct comparison with land plants and complement the Chlorophyta-centred perspective presented here.

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 & 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. Interestingly, examples from our detailed structural analysis also support the view that *Chlorophyta* RLKs began assembling complex architectures early. These findings reinforce the idea that foundational elements of multicellular coordination and environmental perception were already emerging in the *Chlorophyta*. 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.

### **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 & Bacete, 2023), would reveal which carbohydrate motifs (e.g. pectins, arabinogalactans,  $\beta$ -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.

### **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. The authors declare no competing interests.

### **Acknowledgment**

This work was supported by Kempestiftelsen, 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. We thank Manuel Serif for proofreading the manuscript. LB, DM, BGD, and FB also thank the members of the Plant Cell

Wall Dynamics Lab for their support. LB dedicates this work to the memory of Dr Lauri Vaahtera, whose friendship and support shaped her early postdoctoral career.

### **Data availability statement**

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.

### **Reference list**

**Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, *et al.* 2024.** Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**: 493–500.

**Adair WS, Steinmetz SA, Mattson DM, Goodenough UW, Heuser JE. 1987.** Nucleated assembly of *Chlamydomonas* and *Volvox* cell walls. *Journal of Cell Biology* **105**: 2373–2382.

**Alonso Baez L, Bacete L. 2023.** Cell wall dynamics: novel tools and research questions (A Geitmann, Ed.). *Journal of Experimental Botany*: erad310.

**Antolín-Llovera M, Petutsching EK, Ried MK, Lipka V, Nürnberger T, Robatzek S, Parniske M. 2014.** Knowing your friends and foes – plant receptor-like kinases as initiators of symbiosis or defence. *New Phytologist* **204**: 791–802.

**Bacete L, Hamann T. 2020.** The Role of Mechanoperception in Plant Cell Wall Integrity Maintenance. *Plants (Basel, Switzerland)* **9**: 574.

**Bacete L, Mélida H, Miedes E, Molina A. 2018.** Plant cell wall-mediated immunity: cell wall changes trigger disease resistance responses. *The Plant journal : for cell and molecular biology* **93**: 614–636.

**Baez LA, Tichá T, Hamann T. 2022.** Cell wall integrity regulation across plant species. *Plant Molecular Biology* **109**: 483–504.

**Batchelor M, Prasannan S, Daniell S, Reece S, Connerton I, Bloomberg G, Dougan G, Frankel G, Matthews S. 2000.** Structural basis for recognition of the translocated intimin receptor (Tir) by intimin from enteropathogenic *Escherichia coli*. *The EMBO Journal* **19**: 2452–2464.

- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000.** The Protein Data Bank. *Nucleic Acids Research* **28**: 235–242.
- Bhatla SC. 2018.** Signal Perception and Transduction. In: Bhatla SC, A. Lal M, eds. *Plant Physiology, Development and Metabolism*. Singapore: Springer Nature, 729–765.
- Buendia L, Girardin A, Wang T, Cottret L, Lefebvre B. 2018.** LysM Receptor-Like Kinase and LysM Receptor-Like Protein Families: An Update on Phylogeny and Functional Characterization. *Frontiers in Plant Science* **9**.
- Chae K, Gonong BJ, Kim S-C, Kieslich CA, Morikis D, Balasubramanian S, Lord EM. 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. *Journal of Experimental Botany* **61**: 4277–4290.
- Chen R, Huangfu L, Lu Y, Fang H, Xu Y, Li P, Zhou Y, Xu C, Huang J, Yang Z. 2021.** Adaptive innovation of green plants by horizontal gene transfer. *Biotechnology Advances* **46**: 107671.
- Chinchilla D, Shan L, He P, de Vries S, Kemmerling B. 2009.** One for all: the receptor-associated kinase BAK1. *Trends Plant Sci* **14**: 535–41.
- Chiu CH, Paszkowski U. 2020.** Receptor-Like Kinases Sustain Symbiotic Scrutiny. *Plant Physiology* **182**: 1597–1612.
- De K, Pal D, Shanks CM, Yates TB, Feng K, Jawdy SS, Hassan MM, Prabhakar PK, Yang J-Y, Chapla D, et al. 2023.** The Plasminogen-Apple-Nematode (PAN) domain suppresses JA/ET defense pathways in plants. : 2023.06.15.545202.
- Del Hierro I, Mélida H, Broyart C, Santiago J, Molina A. 2021.** Computational prediction method to decipher receptor–glycoligand interactions in plant immunity. *The Plant Journal* **105**: 1710–1726.
- Derelle E, Ferraz C, Rombauts S, Rouzé P, Worden AZ, Robbens S, Partensky F, Degroeve S, Echeynié S, Cooke R, et al. 2006.** Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 11647–11652.
- Dievart A, Gottin C, Périn C, Ranwez V, Chantret N. 2020.** Origin and Diversity of Plant Receptor-Like Kinases. *Annual Review of Plant Biology* **71**: 131–156.
- Domozych D, Ciancia M, Fangel JU, Mikkelsen MD, Ulvskov P, Willats WGT. 2012.** The Cell Walls of Green Algae: A Journey through Evolution and Diversity. *Frontiers in Plant Science* **3**.
- Duan Q, Liu MJ, Kita D, Jordan SS, Yeh FJ, Yvon R, Carpenter H, Federico AN, Garcia-Valencia LE, Eyles SJ, et al. 2020.** FERONIA controls pectin- and nitric oxide-mediated male-female interaction. *Nature* **579**: 561–566.

- Dunser K, Gupta S, Herger A, Feraru MI, Ringli C, Kleine-Vehn J. 2019.** Extracellular matrix sensing by FERONIA and Leucine-Rich Repeat Extensins controls vacuolar expansion during cellular elongation in *Arabidopsis thaliana*. *EMBO J* **38**.
- Fangel JU, Sørensen KM, Jacobsen N, Mravec J, Ahl LI, Bakshani C, Mikkelsen MD, Engelsen SB, Willats W, Ulvskov P. 2024.** The legacy of terrestrial plant evolution on cell wall fine structure. *Plant, Cell & Environment* **47**: 1238–1254.
- Feng W, Kita D, Peaucelle A, Cartwright HN, Doan V, Duan Q, Liu MC, Maman J, Steinhorst L, Schmitz-Thom I, *et al.* 2018.** The FERONIA Receptor Kinase Maintains Cell-Wall Integrity during Salt Stress through Ca(2+) Signaling. *Curr Biol* **28**: 666-675 e5.
- Fuertes-Rabanal M, Rebaque D, Largo-Gosens A, Encina A, Mélida H. 2025.** Cell walls: a comparative view of the composition of cell surfaces of plants, algae, and microorganisms. *Journal of Experimental Botany* **76**: 2614–2645.
- Furumizu C, Aalen RB. 2023.** Peptide signaling through leucine-rich repeat receptor kinases: insight into land plant evolution. *The New Phytologist* **238**: 977–982.
- Gabler F, Nam S-Z, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V. 2020.** Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Current Protocols in Bioinformatics* **72**: e108.
- Gong Z, Han G-Z. 2021.** Flourishing in water: the early evolution and diversification of plant receptor-like kinases. *The Plant Journal* **106**: 174–184.
- Hallgren J, Tsirigos KD, Pedersen MD, Armenteros JJA, Marcatili P, Nielsen H, Krogh A, Winther O. 2022.** DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. : 2022.04.08.487609.
- Hallmann A, Godl K, Wenzl S, Sumper M. 1998.** The highly efficient sex-inducing pheromone system of *Volvox*. *Trends in Microbiology* **6**: 185–189.
- Helland R, Larsen AN, Smalås AO, Willassen NP. 2006.** The 1.8 Å crystal structure of a proteinase K-like enzyme from a psychrotroph *Serratia* species. *The FEBS Journal* **273**: 61–71.
- Herger A, Gupta S, Kadler G, Franck CM, Boisson-Dernier A, Ringli C. 2020.** Overlapping functions and protein-protein interactions of LRR-extensins in *Arabidopsis* (GK Muday, Ed.). *PLOS Genetics* **16**: e1008847.
- Heuberger DM, Schuepbach RA. 2019.** Protease-activated receptors (PARs): mechanisms of action and potential therapeutic modulators in PAR-driven inflammatory diseases. *Thrombosis Journal* **17**: 4.
- Hunsperger HM, Randhawa T, Cattolico RA. 2015.** Extensive horizontal gene transfer, duplication, and loss of chlorophyll synthesis genes in the algae. *BMC Evolutionary Biology* **15**: 16.

- Illergård K, Ardell DH, 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, Dunin-Horkawicz S. 2023.** pLM-BLAST: distant homology detection based on direct comparison of sequence representations from protein language models. *Bioinformatics (Oxford, England)* **39**: btad579.
- van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J, Steinegger M. 2024.** Fast and accurate protein structure search with Foldseek. *Nature Biotechnology* **42**: 243–246.
- Lefort CT, Wojciechowski K, Hocking DC. 2011.** N-cadherin Cell-Cell Adhesion Complexes Are Regulated by Fibronectin Matrix Assembly. *Journal of Biological Chemistry* **286**: 3149–3160.
- Lehti-Shiu MD, Zou C, Hanada K, Shiu S-H. 2009.** Evolutionary History and Stress Regulation of Plant Receptor-Like Kinase/Pelle Genes. *Plant Physiology* **150**: 12–26.
- Lemmon MA, Schlessinger J. 2010.** Cell signaling by receptor-tyrosine kinases. *Cell* **141**: 1117–1134.
- Li Z, Chakraborty S, Xu G. 2017.** Differential CLE peptide perception by plant receptors implicated from structural and functional analyses of TDIF-TDR interactions (B Kobe, Ed.). *PLOS ONE* **12**: e0175317.
- Liu J, Li W, Wu G, Ali K. 2024.** An update on evolutionary, structural, and functional studies of receptor-like kinases in plants. *Frontiers in Plant Science* **15**.
- Lukas KJ. 1974.** Two species of the Chlorophyte genus *Ostreobium* from skeletons of atlantic and caribbean reef corals. *Journal of Phycology* **10**: 331–335.
- Ma Y, Shafee T, Mudiyansele AM, Ratcliffe J, MacMillan CP, Mansfield SD, Bacic A, Johnson KL. 2023.** Distinct functions of FASCILIN-LIKE ARABINOGALACTAN PROTEINS relate to domain structure. *Plant Physiology* **192**: 119–132.
- Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Maréchal-Drouard L, et al. 2007.** The Chlamydomonas Genome Reveals the Evolution of Key Animal and Plant Functions. *Science (New York, N.Y.)* **318**: 245–250.
- Miller S. 2010.** *Volvox*, *Chlamydomonas*, Evolution of Multicellularity. *Nature Education* **3**: 65.
- Moore JP, Gao Y, Zietsman AJJ, Fangel JU, Trygg J, Willats WGT, Vivier MA. 2020.** Analysis of Plant Cell Walls Using High-Throughput Profiling Techniques with Multivariate Methods. In: Popper ZA, ed. *Methods in Molecular Biology. The Plant Cell Wall*. New York, NY: Springer New York, 327–337.



- Morita J, Kato K, Nakane T, Kondo Y, Fukuda H, Nishimasu H, Ishitani R, Nureki O. 2016.** Crystal structure of the plant receptor-like kinase TDR in complex with the TDIF peptide. *Nature Communications* **7**: 12383.
- Oelmüller R, Tseng Y-H, Gandhi A. 2023.** Signals and Their Perception for Remodelling, Adjustment and Repair of the Plant Cell Wall. *International Journal of Molecular Sciences* **24**: 7417.
- Pal D, De K, Shanks CM, Feng K, Yates TB, Morrell-Falvey J, Davidson RB, Parks JM, Muchero W. 2022.** Core cysteine residues in the Plasminogen-Apple-Nematode (PAN) domain are critical for HGF/c-MET signaling. *Communications Biology* **5**: 1–12.
- Pattathil S, Avci U, Miller JS, Hahn MG. 2012.** Immunological Approaches to Plant Cell Wall and Biomass Characterization: Glycome Profiling. In: Himmel ME, ed. Biomass Conversion. Totowa, NJ: Humana Press, 61–72.
- Popper ZA, Tuohy MG. 2010.** Beyond the Green: Understanding the Evolutionary Puzzle of Plant and Algal Cell Walls<sup>1</sup>. *Plant Physiology* **153**: 373–383.
- Poulhazan A, Arnold AA, Mentink-Vigier F, Muszyński A, Azadi P, Halim A, Vakhrushev SY, Joshi HJ, Wang T, Warschawski DE, et al. 2024.** Molecular-level architecture of *Chlamydomonas reinhardtii*'s glycoprotein-rich cell wall. *Nature Communications* **15**: 986.
- Rashidi B, Trindade LM. 2018.** Detailed biochemical and morphologic characteristics of the green microalga *Nannochloris oleoabundans* cell wall. *Algal Research* **35**: 152–159.
- Restrepo-Montoya D, Brueggeman R, McClean PE, Osorno JM. 2020.** Computational identification of receptor-like kinases “RLK” and receptor-like proteins “RLP” in legumes. *BMC Genomics* **21**: 459.
- Sanders CK, Hanschen ER, Biondi TC, Hovde BT, Kunde YA, Eng WL, Kwon T, 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. *Journal of Phycology* **58**: 436–448.
- Sathe S, and Durand PM. 2016.** Cellular aggregation in *Chlamydomonas* (*Chlorophyceae*) is chimaeric and depends on traits like cell size and motility. *European Journal of Phycology* **51**: 129–138.
- Shiu S-H, Karlowski WM, Pan R, Tzeng Y-H, Mayer KFX, Li W-H. 2004.** Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *The Plant Cell* **16**: 1220–1234.
- Song W, Wang B, Li X, Wei J, Chen L, Zhang D, Zhang W, Li R. 2015.** Identification of Immune Related LRR-Containing Genes in Maize (*Zea mays* L.) by Genome-Wide Sequence Analysis. *International Journal of Genomics* **2015**: 231358.

- Spain O, Funk C. 2022.** Detailed Characterization of the Cell Wall Structure and Composition of Nordic Green Microalgae. *Journal of Agricultural and Food Chemistry* **70**: 9711–9721.
- Stegmann M, Monaghan J, Smakowska-Luzan E, Rovenich H, Lehner A, Holton N, Belkhadir Y, Zipfel C. 2017.** The receptor kinase FER is a RALF-regulated scaffold controlling plant immune signaling. *Science* **355**: 287–289.
- Tanaka T, Matsuzawa H, 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. *Bioscience, Biotechnology, and Biochemistry* **62**: 2161–2165.
- Tartar A, Boucias DG, Adams BJ, Becnel JJ. 2002.** Phylogenetic analysis identifies the invertebrate pathogen *Helicospiridium* sp. as a green alga (Chlorophyta). *International Journal of Systematic and Evolutionary Microbiology* **52**: 273–279.
- Tör M, Lotze MT, Holton N. 2009.** Receptor-mediated signalling in plants: molecular patterns and programmes. *Journal of Experimental Botany* **60**: 3645–3654.
- Vasta GR, Ahmed H, Odom EW. 2004.** Structural and functional diversity of lectin repertoires in invertebrates, protochordates and ectothermic vertebrates. *Current Opinion in Structural Biology* **14**: 617–630.
- Vukašinović N, Serif M, Bacete L. 2023.** Cracking the green wall code: insights into cell wall integrity across organisms. *Frontiers in Plant Physiology* **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 Communications* **3**: 100301.
- Weidenbach D, Esch L, Möller C, Hensel G, Kumlehn J, Höfle C, Hüchelhoven R, Schaffrath U. 2016.** Polarized Defense Against Fungal Pathogens Is Mediated by the Jacalin-Related Lectin Domain of Modular *Poaceae*-Specific Proteins. *Molecular Plant* **9**: 514–527.
- Wolf S. 2022.** Cell Wall Signaling in Plant Development and Defense. *Annual Review of Plant Biology* **73**: 323–353.
- Wolf S, van der Does D, Ladwig F, Sticht C, Kolbeck A, Schurholz AK, 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 S A* **111**: 15261–6.
- Yan J, Su P, Meng X, 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.
- Yeh Y-H, Panzeri D, Kadota Y, Huang Y-C, Huang P-Y, Tao C-N, Roux M, Chien H-C, Chin T-C, Chu P-W, et al. 2016.** The Arabidopsis Malectin-Like/LRR-RLK IOS1 Is Critical for

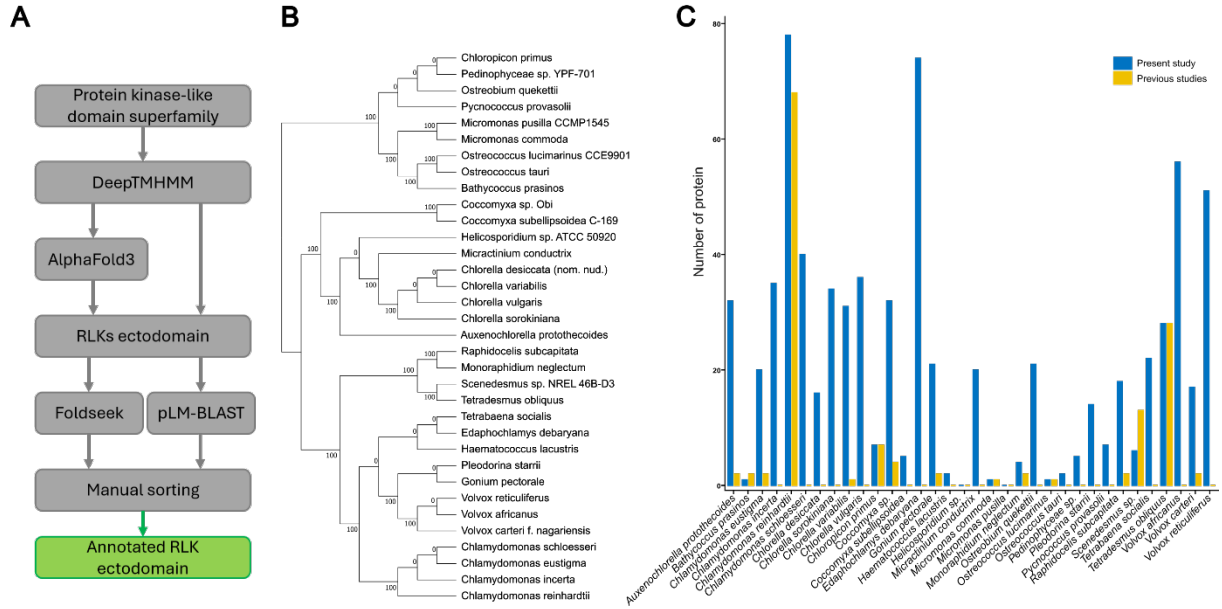
BAK1-Dependent and BAK1-Independent Pattern-Triggered Immunity. *The Plant Cell* **28**: 1701–1721.

**Yin Z, Liu J, Dou D. 2024.** RLKdb: A comprehensively curated database of plant receptor-like kinase families. *Molecular Plant* **17**: 513–515.

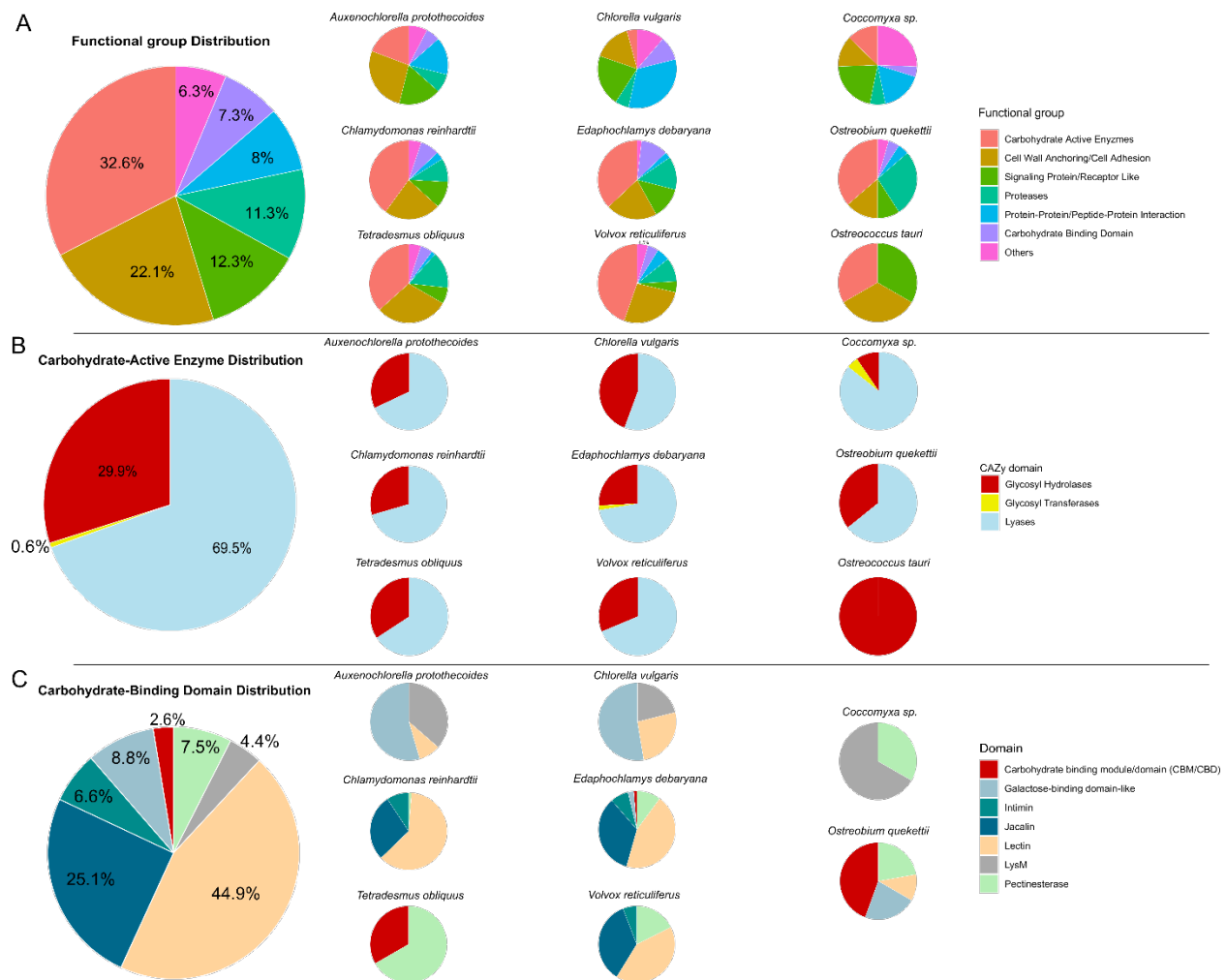
**Yin Z, Shen D, Zhao Y, Peng H, Liu J, Dou D. 2023.** Cross-kingdom analyses of transmembrane protein kinases show their functional diversity and distinct origins in protists. *Computational and Structural Biotechnology Journal* **21**: 4070–4078.

**Zhai K, Rhodes J, Zipfel C. 2024.** A peptide-receptor module links cell wall integrity sensing to pattern-triggered immunity. *Nature Plants* **10**: 2027–2037.

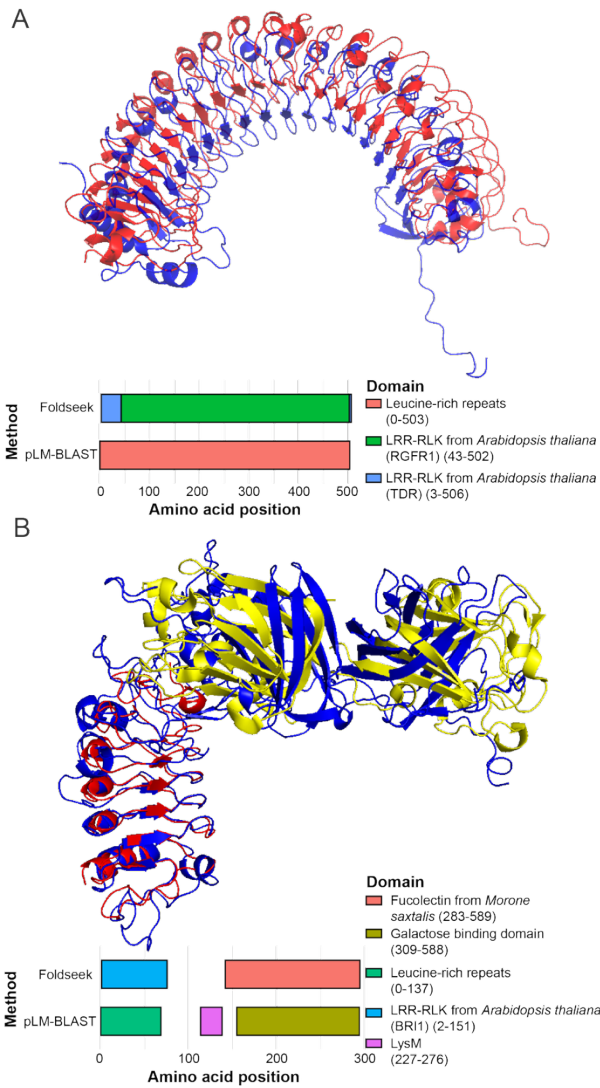
## Figures



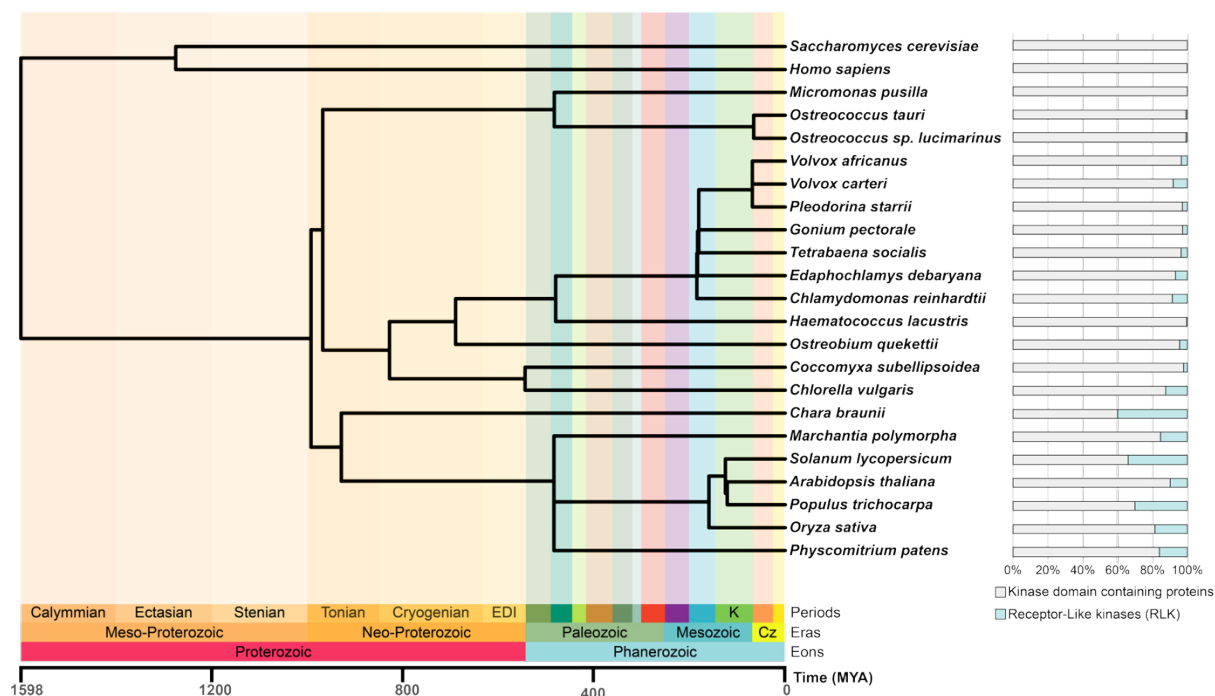
**Figure 1: A pipeline for the identification of new RLK in selected *Chlorophyta* species. (A)** Using available proteomes from *Chlorophyta* species, protein sequences containing the kinase-like domain were extracted and subsequently filtered through DeepTMHMM (Halgren et al 2022). Sequences with predicted 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 & Han, 2021).



**Figure 2: A comparative analysis reveals 7 distinct functional groups in newly-identified RLK's ectodomains.** (A) Predictions obtained through Foldseek and pLM-BLAST analyses on RLK's ectodomain from Nine *Chlorophyta* species were sorted into 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 & Funk, 2022; Poulhazan *et al.*, 2024). Others: Metal binding and oxidoreduction; Defense; Apoptosis; Viral-related proteins.



**Figure 3: 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: 3cqo). The graphical views below protein structures display annotated domains along the protein sequence, combining domain predictions from pLM-BLAST with structural matches identified via Foldseek; Residue ranges for each domain are indicated in brackets.



**Figure 4. Time-calibrated phylogeny and relative abundance of receptor-like kinases (RLKs) across representative green lineages.** The tree shows estimated divergence times in million years, aligned with major geological eons, eras, and periods. For each species, the bar on the right indicates the proportion of RLKs among all kinase-domain-containing proteins. Divergence times are based on published molecular clock estimates; taxa lacking reliable calibration data were excluded.

## Tables

**Table 1:** Number of proteins containing at least one kinase domain 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.

Clade	Class; Family	Genus species	Proteome ID	Taxonomy ID	Kinase domain containing proteins	RLKs
Chlorophyta (Green algae)	Chloropicophyceae;	<i>Chloropicon</i>	UP00031672	1764295	244	7
	Chloropicaceae	<i>primus</i>	6			
	Pedinophyceae;	<i>Pedinophyceae</i>	UP00083601	765719	226	5
	Pedinomonadaceae	<i>sp.</i>	1			
	<b>Ulvophyceae;</b>	<b><i>Ostreobium</i></b>	<b>UP00070814</b>	<b>121088</b>	<b>453</b>	<b>21</b>
	<b>Ostreobiaceae</b>	<b><i>quekettii</i></b>	<b>8</b>			
	Pyramimonadophyceae	<i>Pycnococcus</i>	UP00066026	41880	288	7
	; Pycnococcaceae	<i>provasolii</i>	2			
	Mamiellophyceae;	<i>Micromonas</i>	UP00000200	296587	184	1
	Mamiellaceae	<i>commoda</i>	9			
	Mamiellophyceae;	<i>Micromonas</i>	UP00000187	564608	175	0
	Mamiellaceae	<i>pusilla</i>	6			
	<b>Mamiellophyceae;</b>	<b><i>Ostreococcus</i></b>	<b>UP00000917</b>	<b>70448</b>	<b>290</b>	<b>2</b>
	<b>Bathycoccaceae</b>	<b><i>tauri</i></b>	<b>0</b>			
	Mamiellophyceae;	<i>Ostreococcus</i>	UP00000156	436017	144	1
	Bathycoccaceae	<i>lucimarinus</i>	8			
	Mamiellophyceae;	<i>Bathycoccus</i>	UP00019834	41875	134	1
	Bathycoccaceae	<i>prasinos</i>	1			
	Trebouxiophyceae;	<i>Coccomyxa</i>	UP00000726	574566	242	5
	Trebouxiaceae	<i>subellipsoidea</i>	4			
	<b>Trebouxiophyceae;</b>	<b><i>Coccomyxa sp.</i></b>	<b>UP00082775</b>	<b>2315456</b>	<b>283</b>	<b>32</b>
	<b>Trebouxiaceae</b>		<b>9</b>			
	Trebouxiophyceae;	<i>Helicosporidium</i>	UP00002604	1291522	107	0
	Helicosporidiaceae	<i>sp.</i>	2			
	Trebouxiophyceae;	<i>Micractinium</i>	UP00023964	554055	295	20
	Chlorellaceae	<i>conductrix</i>	9			
	Trebouxiophyceae;	<i>Chlorella</i>	UP00023989	3076	381	34
	Chlorellaceae	<i>sorokiniana</i>	9			
	<b>Trebouxiophyceae;</b>	<b><i>Chlorella</i></b>	<b>UP00105571</b>	<b>3077</b>	<b>253</b>	<b>36</b>
	<b>Chlorellaceae</b>	<b><i>vulgaris</i></b>	<b>2</b>			
	Trebouxiophyceae;	<i>Chlorella</i>	UP00000814	554065	276	31
	Chlorellaceae	<i>variabilis</i>	1			
	Trebouxiophyceae;	<i>Chlorella</i>	UP00069332	1650286	112	16
	Chlorellaceae	<i>desiccata</i>	0			
	Trebouxiophyceae;	<i>Auxenochlorella</i>	<b>UP00002892</b>	<b>3075</b>	<b>504</b>	<b>32</b>
	Chlorellaceae	<i>protothecoides</i>	<b>4</b>			
	Chlorophyceae;	<i>Raphidocelis</i>	UP00024749	307507	369	18
	Selenastraceae	<i>subcapitata</i>	8			
	Chlorophyceae;	<i>Monoraphidium</i>	UP00005449	145388	458	4
	Selenastraceae	<i>neglectum</i>	8			



	Chlorophyceae; Scenedesmaceae	<i>Scenedesmus sp.</i>	UP00058819 3	2650976	339	6
	<b>Chlorophyceae; Scenedesmaceae</b>	<b><i>Tetrademus obliquus</i></b>	<b>UP00025697 0</b>	<b>3088</b>	<b>420</b>	<b>28</b>
	Chlorophyceae; Tetrabaenaceae	<i>Tetrabaena socialis</i>	UP00023633 3	47790	587	22
	<b>Chlorophyceae; Chlamydomonadales incertae sedis</b>	<b><i>Edaphochlamys debaryana</i></b>	<b>UP00061205 5</b>	<b>47281</b>	<b>1000</b>	<b>73</b>
	Chlorophyceae; Haematococcaceae	<i>Haematococcus lacustris</i>	UP00048505 8	44745	585	2
	Chlorophyceae; Goniaceae	<i>Gonium pectorale</i>	UP00007571 4	33097	774	21
	Chlorophyceae; Volvocaceae	<i>Pleodorina starrii</i>	UP00116508 0	330485	491	14
	Chlorophyceae; Volvocaceae	<i>Volvox carteri</i> *	UP00000105 8	3068	486	17
	Chlorophyceae; Volvocaceae	<i>Volvox africanus</i> *	UP00074739 9	51714	632	56
	<b>Chlorophyceae; Volvocaceae</b>	<b><i>Volvox reticuliferus</i>*</b>	<b>UP00074711 0</b>	<b>1737510</b>	<b>1038</b>	<b>51</b>
	<b>Chlorophyceae; Chlamydomonadaceae</b>	<b><i>Chlamydomonas reinhardtii</i></b>	<b>UP00000690 6</b>	<b>3055</b>	<b>828</b>	<b>78</b>
	Chlorophyceae; Chlamydomonadaceae	<i>Chlamydomonas incerta</i>	UP00065046 7	51695	855	35
	Chlorophyceae; Chlamydomonadaceae	<i>Chlamydomonas eustigma</i>	UP00023232 3	1157962	456	20
	Chlorophyceae; Chlamydomonadaceae	<i>Chlamydomonas schloesseri</i>	UP00061374 0	2026947	756	40
Streptophyta (Charophyte algae)	Charophyceae; Characeae	<i>Chara braunii</i>	UP00026551 5	69332	652	435
Embryophyt a (Land plants)	Marchantiopsida; Marchantiaceae	<i>Marchantia polymorpha</i>	UP00024400 5	3197	1284	234
	Bryopsida; Funariaceae	<i>Physcomitrella patens</i>	UP00000672 7	3218	1723	329
	Magnoliopsida; Salicaceae	<i>Populus trichocarpa</i>	UP00000672 9	3694	2785	1192
Tracheophyt a (Vascular plants)	Liliopsida; Poaceae	<i>Oryza sativa</i>	UP00005968 0	39947	4981	1132
	Magnoliopsida; Brassicaceae	<i>Arabidopsis thaliana</i>	UP00000654 8	3702	5564	610

**\*Note:** The *Volvox* genus is polyphyletic, which could explain the observed disparities in the number of putative RLKs identified across species within the same genus.

### **Supplementary material:**

**Table S1:** 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.

**Table S2:** Amino acid 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**).

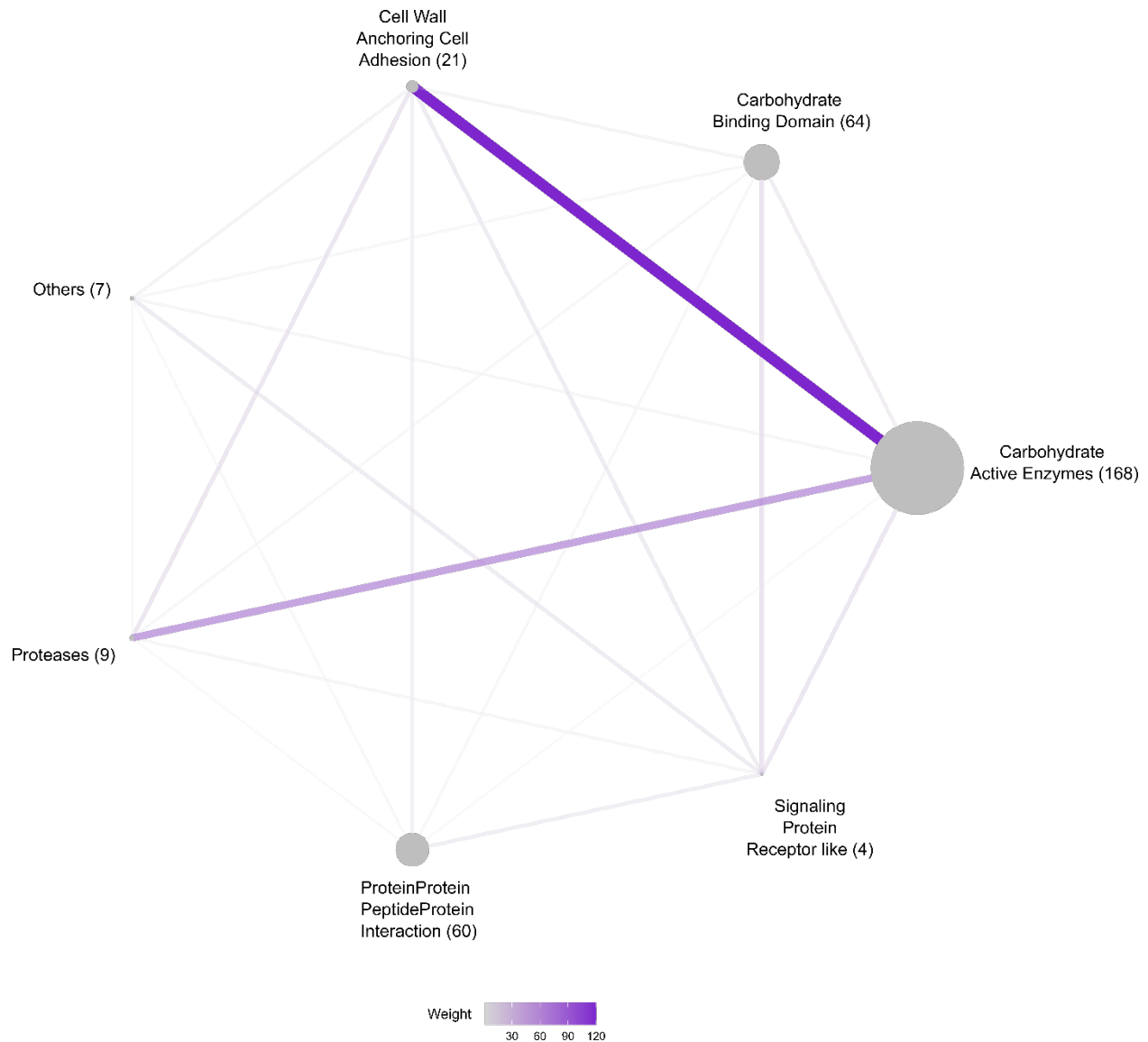
**Table S3: Functional classification of RLKs proteins across selected green algae species.** The table summarizes the functional group distribution for the protein of interest. The table also includes the relative proportion of proteins having cell wall-related domains (carbohydrate-binding modules/domains and carbohydrate-active enzyme classes). Percentages indicate the relative abundance of each subcategory within its respective group ("n.a." = data not available).

**Table S4:** 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).

**Table S5:** 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).



**Figure S1: 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.